VOL. 11, NO. 9, MAY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

AN EFFECTIVE RETRIEVAL OF SECURED AND RANKED DATA IN CLOUD ENVIRONMENT

R. Shanmugaapriyaa and A. Safiya Parvin
Department of Computer Science, Sathyabama University, Chennai, Tamil Nadu, India
E-Mail: saishanmu@gmail.com

ABSTRACT

Cloud computing is a successful environment due to many recent technologies. The main advantage of using cloud computing is that it allows users to store a copious amount of data. The mobile data which has been redistributed to the third party cloud requires protection from the trapdoor, intruders to do so the data were encrypted using asymmetric encryption into blocks and then distributed to the various cloud storages. But there exists some difficulties in search of those data in cloud, which may not give an accurate search results. The main aim of this paper is to overcome this drawback through gateway encryption and blind storage, through dynamic block splitting in order to preserve the redistributed data in the cloud storage. Natural Language Processing technique is used to extract keywords by secured multi-keyword search over the cloud data which pre computes the resulting data that the user needs to search and gives accurate and relevant results by without downloading the files contents .Indexing and the privilege preferences were also done prior to enhance the security and decryption of the entire group members file content is also done.

Keywords: cloud computing, searchable encryption, multi- keyword rank search, blind storage, access pattern.

1. INTRODUCTION

Cloud Computing provides communication and transport by using various network, hardware and software services to the client. Cloud computing overcomes the drawback [12]-[3] of hardware limitation of devices by researching the scalability of cloud storage and resources that are required for computing and provides very powerful and scalable services to the users. The data that were redistributed by the users may contain information that requires unwarranted disclose such as, personal information, photos, emails, card details, customer records ,medical records, banking information, this may lead to privacy violation and intruders attack [5] if there is no sufficient protection. To ensure the privacy the outsourced data must be encrypted before it is redistribution over the cloud for storage. Even this encryption may lead to many crises during the search by other users for their interesting data over the cloud. These frustrations has paved the way for more related exploration of searchable encryption techniques [6],[7], [10].

The search could be carried out over the cloud data [12], [10] may require symmetric encryption of the data in server side is done which is really a tough task and in secure, while searching the user may require to decrypt and download each and every document one by one for content based search and where the multi-key word search is cumbersome in this stage. Cao *et al.* [11] suggested multi-keyword search scheme which supports the ranking of results using the K-nearest neighbour (KNN) technique. Naveed *et al* [13] recommended dynamic searchable encryption technique over the blind storage to wrap the access pattern of the user's search. These hindrances were swamped in this paper by making use of the consecutives:

- The data are stored in the form of blocks in the memory by using the dynamic block splitting technique, which increases the efficiency and reduces the space occupation.
- Asymmetric encryption technique which is otherwise termed as public key encryption, using RSA algorithm is being used which enhances the security and reduces the time consumption.
- Public key encryption uses two keys that are linked mathematically they are: public key and private key .Public key is used by everyone for encryption of the data and the private key used by the receiver for the decryption of the text.
- Accessing privileges is being designed, for both the private and public access of the users as per their interest. And accessing capability of the user is also restricted in order to enhance the confidentiality.
- Natural Language Processing (NLP) and word net tool is being used here for mining the text from the cloud storage.

The paper is organized as follows,

In section II, the system design and design goals are illustrated.

In section III, Methods and Algorithms are used. In section IV, Various stages.

In section V, Related work and finally the conclusion of this paper in section VI.

2. RELATED WORK

Cash. D *et al.*, [3] put forward that in cloud computing every data is stored in the third party cloud server. Users may store some confidential data in the server; hence to search those sensitive data in the cloud server they made use of SSE (Symmetric Secured Encryption) technique. Boolean data is also handled here. The limitation faced here is that this paper did not support the relevance scoring and the results are also not ranked properly which has paved the way for utilizing huge amount of time in search of the data.

Cao. N *et al.*, [8] Multi-key word ranked search is adopted for searching up the encrypted data from the cloud

server. Similarity based search is being adopted for searching the data by making use of the similar keywords is done here. They use the concept of inner product similarity for performing the quantitative evaluation. Though they used up the technique of relevance scoring, it's quite cumbersome since the server requires it to calculate it twice for each and every inner product which requires much time and the computational trouble gets increased. This scheme is impractical for large databases.

Sun. W, *et al.*, [10] suggest dynamic collection of encrypted document is stored on the cloud server by making use of searchable symmetric encryption technique. The cloud server where the data resists is even not aware of the content within it. The server is just used for the data uploading and downloading. The drawback encountered here is the relevance scoring and search based on the rank is not supported.

Boneh. D, *et al*., [5] portrays that the public key system constructed a structure for searching securely the encrypted data over the cloud environment and also made use of the comparison queries and more general queries. In order to support queries the key made use of tokens, which allows the user to test the query over the cipher text by without disclosing the content within it. The only drawback here is unordered search is being followed which turns out very tough and quite confusing; this is over come in our system.

Cash. D, *et al.*, [2] *Dynamic symmetric searchable schemes* were adopted and used here, which is efficient and the privacy is too good .They used up huge data sets .They used singled keyword for searching purpose. The trouble here is no ranking is used, which we use in proposed.

Li. H, *et al.*, [1] suggested multi-key word search techniques along with the relevance scoring and the process of ranking is used for taking up the data from the cloud server and by retrieving it as per the users requirement the only demerit here is they used symmetric cryptography where single key is used for both encryption and decryption which leads to greatest vulnerability for data in the server.

Yang. Y *et al.*, [10] fostered the usage of *Multi-keyword search schemes* with similarity based ranking search is being used. *Indexed based searching* is done to pick out the redundantly occurring terms. Multi-dimensional algorithms and the cipher text models are used to improve the privacy of the data. These technologies were used as the basis for our system.

Zhen. Q, *et al.*, [9] proposed a unique architecture for verifiable attribute based keyword search which handovers the complete control of the encrypted static data to the owner. This affords the genuine users to access the outsourced data in the third party server. Searchable public Encryption which supports only the asymmetric cryptographic operations is being used.

Wang. B, *et al*., [1] suggested *Multi- dimensional public key range* search to encrypt the cloud data. This reduces the problem of utilization of space. *Indexed search* is used to retrieve data using R-Tree search. Ranking and scoring is not done here which turned out a great drawback for this system.

3. SYSTEM DESIGN AND GOALS

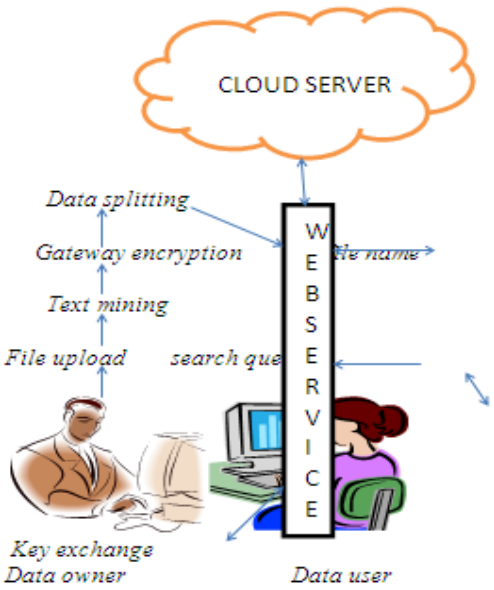
A. System design

As in Figure-1, the system consists of three entities

1. The data owner (*DO*)
2. Blind storage (*B*)
3. Data User (*DU*).

The data owner is one who is the authority of the data. The blind storage is the space in the cloud which is the third party server, holds up the data from various sources. The data user is one who is in need of the data that is present in the cloud server.

The data owner preservers a massive collection of encrypted documents *C* in the form of blocks in the blind storage *B*(cloud server).The data owner uploads the file by pre computing the encryption key through gateway encryption for each and every document before uploading it into the blind storage (*B*) as indexed file. POS tagger is used for extracting the keywords from the file stored in the blind storage (*B*). The data User (*DU*) searches the file on the cloud server (*B*). The cloud server (*B*) maps the keyword and searches the associated file and gives to the user .To access the file the data user (DU) must sent the file name to the data owner (*DO*). Since data owner is aware of every user's public key he will encrypt that particular file using the requested data user's (*DU*) public key and now the data user can decrypt the contents of the file.



B. System goals

To empower the efficient, accurate and secure search over the blind storage on the cloud these are the some goals followed:

Multi-key word search: To support and ease the work of the search user multi-keyword search is being used along with the relevancies, which helps user to reduce their search timings by easily grab their required search not only multi -key word search is used but also the search based on the relevance is also done.

Retaining privacy and confidentiality: In order to secure the document from the accessing the information within it by the third party cloud where it is being stored, many ways were followed like restricting the accessibility of the each and every requesting user (read only, write only, both read and write).

4. MODULES AND DESCRIPTION

This project is classified into four main modules like group creation, text mining, blind storage and query search. Various techniques were used here in easy module to improve the efficiency.

a. Group creation

The data owner must get registered to the cloud environment so that they can upload their encrypted documents .Not only data owners but also data users must get registered in the environment by selecting the group owner under whom the user wishes to join, only then they can request their needed document from the owner .If the user is being accepted by the owner, data owner will add the user under him only then he can sent request and communicate between the owner regarding his request. There exists numerous group owners and group in the cloud environment, data user's decision is selecting their own group and getting register into it.

b. Text mining

The data owner uploads the file to the cloud server, and the content of the file that is uploaded can be extracted using the technique called "Natural Language Processing". The synonym of the extracted words is obtained using a tool called the word net tool which is the large lexical database of English. In this tool the adjectives, adverbs, nouns and verbs are grouped together into synsets which is nothing but the set of synonyms.

These synsets are interlinked using the conceptual semantics and lexical relations.

POS tagger is used to extract the keyword from the content of the file that is extracted using the NLP technique. Text mining is a three step process,

Steps of text mining:

a) POS tagger for retrieving the keyword from the file.

b) NLP technique is used to provide the literal meaning for those keywords extracted in the previous step.

c) Analyzing of the words are done using the word net API.

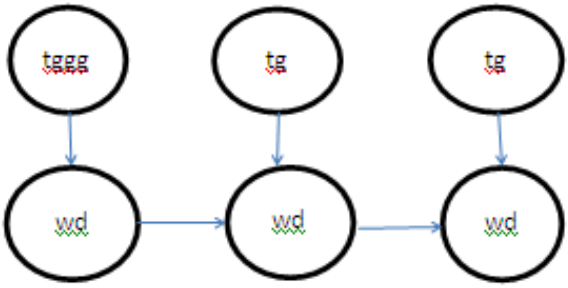
B.1 Algorithm

B.1.1. POS Chunnker

Step 1: Input *f*.// *f* is a file .
Step 2: Read *c*.// *c* is the content of the file
Step 3: Divide *c*, based on the words *w* into chunks.

B.1.2. POS Tagger

Step 1: Input*f*. // *f* is a file.
Step 2: Read *c*.//content of the file.
Step 3: Find the grammar for each chunk in the *c*.
Step 4: Enumerate which is verb, noun, adverb and adjective.



Where
tg-Tagger.
wd-Word.

$P(tg^g|wd^h)=\mu_1(wd^h)\ b(tg^g,wd^h)/b(wd^h)+\ \mu_1(wd^h)\ b_m(tg^g)/\ b_m()$.

Where
$\mu_1(wd^h)$ will be 1 if $b(wd^h)\geq 1$ else 0 otherwise.

Where
$B\ (tg^g,wd^h)$-Number of times the word h appears with the tag g.
$b\ (wd^h)$-Number of times the word h appears.
$b_m\ (tg^g)$-Number of times the untagged word g gets the tag g.
$b_m()$-Total occurrences.

C. Blind storage

The files that are uploaded by the data owner are encrypted using the Natural language processing and the file is stored after indexing. The data owner provides the access restrictions and certain privilege constraint for the users in order to enhance the security of their files.

Access restrictions are that shows the permission for the users to access the file in the cloud server. Privilege constraint denotes, whether the particular user has the right to only read or only write or both read and write. The files uploaded in the server is stored in the form of blocks using the asymmetric encryption technique -RSA (Ron Rivest, Adi Shamir and Leonard Adleman) it uses both the private and public keys .This RSA algorithm fascinates by the product of two prime numbers and then by means of supplementary operations of two set where one set encompasses public key for encryption and the other set involves the private key decryption.

C.1 Algorithm

Step 1: Choose two definite primes a and b.//a and b of same bit length.
Step 2: Calculate m=ab//m is the module for public and private key.
Step 3: Enumerateϕ(ab)=(a-1)*(b-1)//Euler's function.
Step 4: gcd(ϕ(m),u)=1// u is a public key and u and ϕ(m) are co- prime.1<u<(ϕ(m)).
Step 5: r*u mod (ϕ(m))=1
r=u^{-1} mod (ϕ(m))// r is a private key and multiplicative inverse of u^{-1} mod (ϕ(m)) .

d. Query search

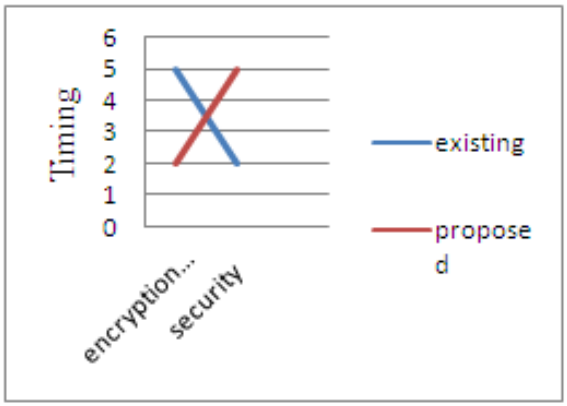
The data user tries to access the cloud server for the required data by searching through relevant keywords. The cloud server will send name of the file to the data user. At this stage the data user clicks the file name which was sent by the server and the data owner send the request for access the file to the server. Now the server forwards the request of the user, along with the file name to the respective data owner. Since the data owner consists of every data user's public key (P_u), he (data owner) will encrypt his private key (P_r) using the public key (P_u) of the user ((P_r)P_u) and send the key details to the server. Now the server will sent the encrypted key details to the user, user can decrypt it, using his public key (P_u)and gets the owner private key(P_r) and can now access the data through the blind storage.

5. ANALYSIS AND PERFORMANCE COMPARISON

a. Security and feasibility

Public key cryptosystem makes use of more than one key for both encryption and decryption and it complements the symmetric encryption techniques. In symmetric encryption there exists some disclose of information during public .

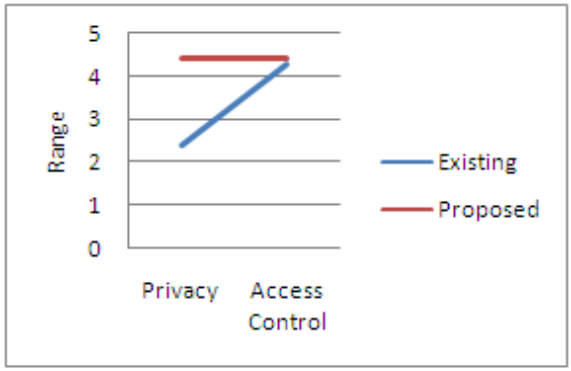
Communication [1] and when compared with the existing system the time consumed for encryption is more.



(a) Security and feasibility

b. Privacy and access control

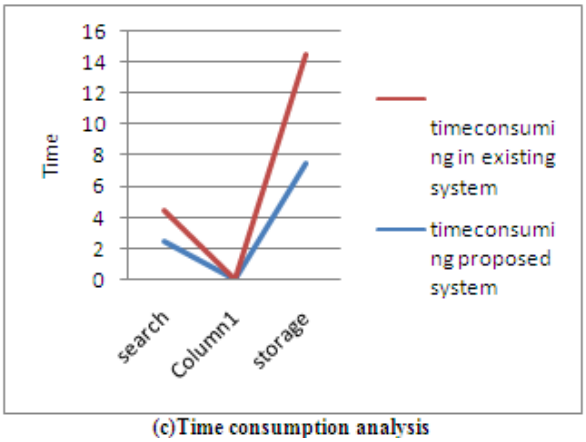
Multiple-keyword is handled better when compared to symmetric encryption. Ranking result is also very good. Confidentiality, Trapdoor un-link ability and concealing access pattern of search user is better when compared to symmetric encryption.



(b) Privacy and Access Control

c. Time consumption analysis

The time taken for searching information from the server is reduced when compared from the existing system (symmetric).The storage method used in the asymmetric process is also far different from existing system, where the information is stored in the form of block here in proposed system.



(c)Time consumption analysis

6. CONCLUSIONS

Here we have proposed the Multi key word search efficiently through the third party blind storage which facilitates the accurate, efficient and secured search over the encrypted data over the cloud third party blind storage. Privacy is preserved in for the data in the blind storage and the access control is also achieved through this schemes used. Trapdoor privacy is also increased. For future work investigations over authentications and the access control issues could be carried out.

REFERENCES

[1] Wang. B, Yu. S, Lou. W and Hou. Y.T. 2014. Privacy preserving multi-keyword fuzzy search over encrypted data. Proc. IEEE INFOCOM. pp. 2112-2120.

[2] Cash. D, Jaeger. J, Jutla. C, Krawczyk. H, Rosu. M. C and Steiner. M. 2014. Dynamic searchable encryption in a very large data structures and implementation. in Proc. NDSS.

[3] Cash. D. Jaeger. J, Jutla. C. Krawczyk. H, Rosu. M.C and Steiner. M. 2013. Highly Scalable Searchable symmetric encryption with support for Boolean queries. in Proc. CRIPTO. pp. 353-373.

[4] Boneh. D and Water. B. 2007. Conjunctive subset of range queries on encrypted data. in Proc. TCC. pp. 506-522.

[5] Boneh. D, Crescenzo. G.D, Ostrovsky. R, Persian. G. 2004. Public key encryption with key word search. Pro. EUROCRYPT. pp. 506-522.

[6] Liang. H, Cai. X.L, Huang. D, Shen. X and Peng. D. 2012. An SMDP-based service model for inter domin allocation in mobile cloud networks. in proc. IEEE Trans. Veh Technol. 61(5): 2222-2232.

[7] Naveed. M, Prabhakaran. M and Gunter. A.C. 2014. Dynamic searchable encryption via Blind storage. In: Pro. IEEE Symp. Secur. Privacy. pp. 639-654.

[8] Cao. N, Wang. C, Li. M and Lou. W. Privacy preserving multi-keyword search over encrypted cloud data. IEEE Trans. Parallel Distrib. Syst. 25(1): 222-2333.

[9] Zhen. Q, Xu.S and Ateniese. G. 2014. VABKS: Verifiable attributed based keyword search over outsourced encrypted data. In: Proc. IEEE INFOCOM. pp. 522-530.

[10] Sun. W, Yu. S, Lou. W, Hou. Y.T and Li. H. 2014. Protecting your right: Attribute based keyword search with fine grained owner-enforced search authorization in the cloud. In: Proc. IEEE INFOCOM. pp. 226-234.

[11] Yang. Y, Li. H, Liu. W, Yang. H and Wen. M. 2014. Secured dynamic searchable Symmetric encryption with constant document update cost. In: Proc. GLOBECOM. Anaheium, CA, USA.

[12] Li. H, Liu. D, Dai. Y, Luan. H.T and Xuemin, Shen. 2015. Enabling Efficient Multi-keyword ranked search Over Encrypted Mobile Cloud data Through Blind Storage. 3(1).