OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

PROFICIENT SENSIBLE SEARCH OVER ENCRYPTED DATA IN CLOUD

Sayali S. Wagh, Sneha S. Kochale, Smita S. Bhosale, Aarti D. Landge, S. N. Autade

Savitribai Phule Pune University, Pune, Maharashtra, india

Abstract: **Information retrieval on Encrypted Data Facilitates data confidentiality with sensing concept of data store into cloud. Now a days, current encryption search process support only monogram or bigram keyword search. Expressive keyword search process is computationally inefficient due to bilinear pairing process. Now, the proposed research implements proficient search over encrypted data with Boolean expression is supported. Then algorithmic process implements pattern matching concept. Information retrieval from this process assist user to search in cloud without decrypting data.**

Keywords: *Searchable encryption, Cloud computing, expressiveness, Attribute-based encryption*

I INTRODUCTION

The number of internet users across the globe increasing exponentially. So that the data by the users need to be store at huge storage spaces and the cloud is the best solution for this. Due to complex computational structure of cloud and its data handling techniques it is unable to provide the security for all the stored data in the cloud.

As cloud computing is nothing but sensitive, secret information are stored centralized into the cloud such as personal records, Emails etc. then the user can be access and stored this type of information onto the cloud to enjoy the on-demand high quality data storage services. However, the data owner an cloud server are not in trusted domain so the data source at risk, as the cloud server may no longer be full trusted.

Then this sensitive data have been encrypted to unreadable format for the data privacy and secured from unauthorized person or hackers.

So the cloud service provides manage to apply strong cryptographic algorithms to encrypt the data before storage process. And they provide original data to the users by decrypting the same on their request. Solution doesn't end here only as cloud system allocates huge storage space for the users, so users are free to take advantage of this and store huge number of documents. This again creates the problem of searching the document in the cloud as all are present in the encrypted state.

The common solution to this is after getting the user keyword for searching, every document needs to decrypt first and then the keywords need to match in every word of the document to retrieve the desired one. But this process takes much more time to search the documents, so a need of proper and fast searching technique arises which can search the documents in the cloud without decrypting the data to save the cost of cloud service provider and time of the end users.

However, data encryption is very efficient data usage a invoke tasks because there is large amount of information or data is presented. So, In the cloud computing data legatee may exchange or share their data with a maximum amount of users. The single user can access appropriate data files from cloud for interested in during.

Then Most of users search the records by using the keywords instead of accessing all the encrypted files back to their original formats.

Such keywords search method allows user to access the information which they needed mostly. The example of this method is called as "Google search".

Unfortunately, data encryptions do not allow user's ability to perform keyword search and thus makes the plaintext search methods does not suitable for cloud computing. Now a days, to securely search encrypted data, searching encryption technique have been developed. The information is searched by using keywords scheme and this keyword scheme usually build index for keyword of their interest and support index file that contains the keywords.

II MOTIVATION

In today's scenario the data stored over cloud is in the form of encrypted data and while accessing it there is need of decryption every time. Whenever the query is fired for searching the data, the related files are decrypted for matching purpose, which leads into increasing time to search the data.

Instead of that if the fired query get encrypted by the same encryption algorithm then the accessing files become more faster.

The n-gram pattern matching technique is used which searches the data more sensitively.

For enhancing the security factor the trapdoor system is used.

III SYSTEM ARCITECTURE

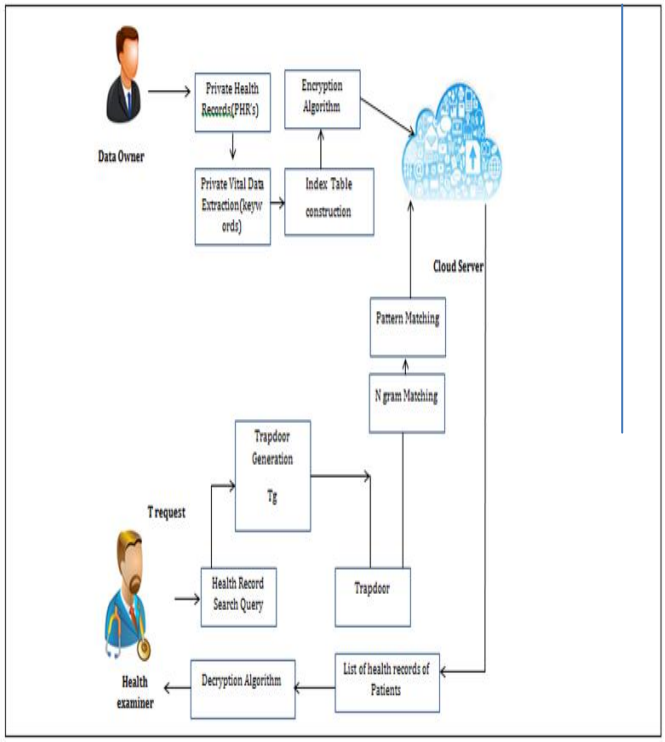

Figure 1:System Architecture

IV ALGORITHM

Algorithm 1 : Pre-processing
Step 0: Start
Step 1: Read string
Step 2: divide string into words on space and store in a vector V
Step 3: Remove Special Symbols
Step 4: Identify Stopwords
Step 5: Remove Stopwords
Step 6: Identify Stemming Substring
Step 7: Replace Substring to desire String
Step 8: Concatenate Strings
Step 9: stop

Algorithm 2 gives the steps for searching the encrypted file over cloud. Here, by the index build on the file while uploading, the documents are searched over the cloud. While searching for any file, the keywords are compared with the contents of file using the Pearson co-relation technique and the most matching relevant files are given as output to user.

Algorithm 2 :Searching
Step 0: Start
Step 1: Read Query
Step 2: Create bucket B of query where minimum word of bucket should have three characters.
Step 3: Apply AES algorithm to encrypt the bucket.
Step 4: Read encrypted file from cloud
Step 5: divide content of file into words on space and store in a vector V
Step 6: for i=0 to length of B
Step 7: for j=0 to length of V
Step 8: if (V_j is equals to B_i)
Send length of $B_i$and Length of vector V to the Pearson correlation
Step 9: Set 0.5 as a threshold to get more precise files and store them in vector F
Step 10: end inner for
Step 11: End outer for.
Step 12 : return File vector F

V MATHEMATICAL MODEL

(A) Set Theory

1. S= { } be as system for Effective and Expressive Search on cloud Data
2. Identify Input as D={ HD_1, HD_2, HD_3…..HD_n}
   Where HD_n =Health Records
3. Identify as Output M as i.e. Matched Health Records
   S= {HD_n, M}
4. Identify Process P
   S= {HD_n, P, M}
   P= { Q_C ,E, T_g, NP_M }

Where,

Q_C = Query processing
T_g= Trapdoor Generation
E_p= Encryption Process
NP_M= N Gram Generation and Pattern Match
S = { HD_n , Q_C ,T_g, E_p, NP_M }

(B) SET DESCRIPTION:
1: Query processing
Set Q_C:
Q_{C0} = Data cleansing
Q_{C1}= Keyword Identification
Q_{C2}= NLP protocols
Q_{C3}= Expressive Rule Addition
2. Trapdoor Generation
Set T_g:
T_{g0}=SHA512 initialization

T_{g1}=Apply Reducer for 32 bit key
T_{g2}=Generated Hash
T_{g3}=Store Hash
3. Encryption Process:
Set E_p:
E_{p0}=Initialize Algorithm AES
E_{p1}=Initiate Key

E_{p2}=Apply Encryption

4. N Gram Generation and Pattern Match:
Set $NP_{M:}$
$NP_{M\,0}$= Initiate N-gram generation value
$NP_{M\,1}$= Generated N Grams
NP_{M2}= Pattern generation and support calculation
NP_{M3}= Matched Patterns

Resultant Output of System is List of Documents matching Input Query

(C) Representation of Sets and its operation:-

1. Union Representation:-

A. Set $Q_C = \{ Q_{C0}, Q_{C1}, Q_{C2}, Q_{C3} \}$

Set $T_g = \{ T_{g0}, T_{g1}, T_{g2} T_{g3} \}$

Set $(Q_C \cup T_g) = \{ Q_{C0}, Q_{C1}, Q_{C2}, Q_{C3}, T_{g0}, T_{g1}, T_{g2} T_{g3} \}$

B. Set $E_p = \{ E_{p0}, E_{p1}, E_{p2}, E_{p3} \}$

Set $(Q_C \cup T_g UE_p) = \{ Q_{C0}, Q_{C1}, Q_{C2}, Q_{C3}, T_{g0}, T_{g1}, T_{g2} T_{g3}, E_{p0}, E_{p1}, E_{p2}, E_{p3} \}$

C. Set $NP_M = \{ NP_{M0}, NP_{M1}, NP_{M2}, NP_{M3} \}$

Set $(Q_C \cup T_g UE_p \cup NP_M) = Q_{C0}, Q_{C1}, Q_{C2}, Q_{C3}, T_{g0}, T_{g1}, T_{g2} T_{g3}, E_{p0}, E_{p1}, E_{p2}, E_{p3}, NP_{M0}, NP_{M1}, NP_{M2}, NP_{M3} \}$

VI RESULTS

Some experimental evaluations are performed to show the effectiveness of the system. And these experiments are conducted on the windows based java machine with universally used IDE NetBeans. Also, the numbers of retrieved documents are used to set the benchmark for performance evaluation. Numbers of relevant documents retrieved from the private cloud for the set of keywords are used to show the effectiveness of the system.

Numbers of scenarios present where one measuring parameter dominates the other. By taking such parameters into consideration, two measuring parameters such as precision and recall are used.

Below are the definition of the used measuring techniques i.e. precision and recall.

Precision: It is a ratio of numbers of proper documents retrieved to the sum of total numbers of relevant and irrelevant documents retrieved. Relative effectiveness of the system is well formulated by using precision parameters.

Recall: it is a ratio of total numbers of relevant documents retrieved to the total numbers of relevant and relevant documents not retrieved. Absolute accuracy of the system is well described by using recall parameter.

Therefore,

$$\text{Precision} = \frac{Documents\ Retrived}{Relevant\ documents + Irrelevant\ documents}$$

$= (X / (X+ Z))$

And

$$\text{Recall} = \frac{Relevant}{Relevant + Relevant\ not\ identified}$$

$=(X / (X+ Y))$

Following table 1shows the Precision and Recall Chart for the values of T, X, Y and Z.For the above values of X,Y and Z, on calculating the precision and recall values using the formula, we get the following graphs

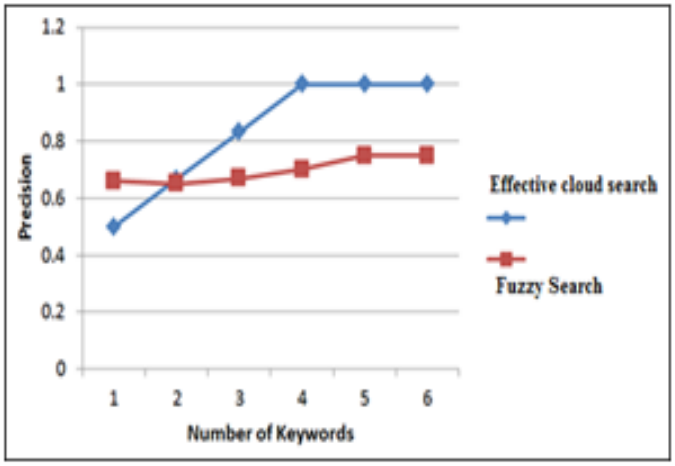

Figure 1:Average Precision comparison by Co-related Search and Fuzzy Search method depending on values of Table 1

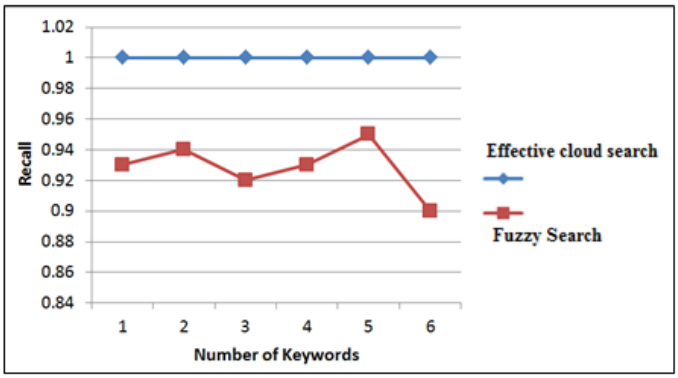

Figure2: Average Recall by Co-related Search and Fuzzy Search method depending on values of Table 1

Table No.1

T=Number of document extracted by the system	X= Relevant Documents in T	Y=No. of relevant documents not extracted	Z= Number of irrelevant Documents extracted	Precision = (X/ (X+ Z))		Recall = (X/ (X+ Y))	
				Effective cloud search	Fuzzy Search	Effective cloud search	Fuzzy Search
2	1	0	1	0.5	0.66	1	0.93
3	2	0	1	0.666666667	0.65	1	0.94
6	5	0	1	0.833333333	0.67	1	0.92
4	4	0	0	1	0.7	1	0.93
6	6	0	0	1	0.75	1	0.95
2	2	0	0	1	0.75	1	0.9

REFERENCES

[1]Efficient and Expressive Keyword Search OverEncrypted Data in Cloud
Hui Cui, Zhiguo Wan, Robert H. Deng, Guilin Wang, and Yingjiu Li

[2] "Efficient Similarity Search over Encrypted Data", Mehmet Kuzu, Mohammad Saiful Islam, Murat Kantarcioglu, ACM publications, 2012

[3] " Ginix: Generalized Inverted Index for Keyword Search ",Hao Wu , Guoliang Li, and Lizhu Zhou, IEEE Transaction, Volume 18, Number 1, February 2013 .

[4] "K-Gram Based Fuzzy Keyword Search over Encrypted Cloud Computing" Cong Wang1, Qian Wang1, Kui Ren1, and Wenjing Lou2

[5] "Above the clouds: A Berkley view of cloud computing".

[6] "Fuzzy Keyword Search over Encrypted Data in Cloud Computing" Jin Li, Qian Wang, Cong Wang, Ning Cao, KuiRen, and Wenjing Lou.

[7] "Secured Multiple-keyword Search over Encrypted Cloud Data." Prof. C. R. Barde1, Pooja Katkade2, Deepali Shewale3, Rohit Khatale4. www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 2, February 2014).

[8] "Privacy- Preserving Keyword-based Semantic Search over Encrypted Cloud Data" Xingming Sun, Yanling Zhu, Zhihua Xia and Lihong Chen International Journal of Security and Its Applications Vol.8, No.3 (2014), pp.9-20.

[9] "Authorized Private Keyword Search Encrypted Data into Cloud Computing" Ming Li∗, Shucheng Yu†, Ning Cao∗ and Wenjing Lou∗.