Open access Journal International Journal of Emerging Trends in Science and Technology

DOI: http://dx.doi.org/10.18535/ijetst/v2i9.08

Brand New Query Processing Using Seed Block Algorithm on Cloud Storage

Author
V. Mahesh Babu
M.Tech Student, Dept of Computer Science and Engineering,
Sri Venkateswara College of Engineering and Technology, Chittoor (D), Andhra Pradesh, India
Email: *mahesh.valameti@gmail.com*

Abstract

Now-a-days incrementing users of public cloud computing infrastructures, utilizing clouds to store data with query accommodations are solution which gives more scalability and cost-preserving. Hence, most of the data is sensitive that the data owner does not optate to move their data to the cloud without utilizer get the data confidentiality and query privacy are ascertained. In cloud computing, data engendered in electronic form are sizably voluminous in amount. To maintain this data efficiently, there is a desideratum of data instauration accommodations. To cater this, in this paper we propose a Seed Block Algorithm. The objective of proposed algorithm is twofold; first it avail the users to amass information from any remote location in the absence of network connectivity and second to recuperate the files in case of the file expunction or if the cloud gets eradicated due to any reason. The time cognate issues are additionally being solved by proposed SBA such that it will take minimum time for the recuperation process. Proposed SBA withal fixates on the security concept for the back-up files stored at remote server, without utilizing any of the subsisting encryption techniques.

Keywords: *Target distribution, LBS, bucketization, query privacy, data confidentiality.*

1. Introduction

Query accommodations in the cloud are increasingly popular because of the unique advantages in quality and cost saving. With the cloud infrastructures, the accommodation owners can conveniently scale up or down the accommodation and only pay for the hours of utilizing the servers. This is a captivating feature because the workloads of query accommodations are highly dynamic, and it will be extravagant and inefficient to accommodate such dynamical workloads with in-house substructures. However, because the accommodation providers lose the control over the information in the cloud, data privacy and query privacy have become the major concerns. Summarization of these requisites for constructing a practical query accommodation in the cloud as the CPEL criteria: data privacy, query Privacy, effective query working, and Low in-house working cost. The rudimentary conception is to arbitrarily transform the multidimensional datasets with a merger of order preserving encryption, dimensionality expansion, desultory noise injection, and arbitrary project, so that the utility for processing range queries is preserved. Driven by lower cost, higher reliability, better performance, and more expeditious deployment, data and computing accommodations have been increasingly outsourced to clouds such as Amazon EC2 and S3, Microsoft Azure, and Google App Engine. However, privacy has been the key road auction block to cloud computing. On one hand, to leverage the computing and storage capability offered by clouds. While storing data on cloud users have to face the quandary of some delay in the retrieving data from cloud storage. Data privacy and efficiency utilizing file retrieval from Ostrovosky.

In this scheme utilizer can retrieves files from an untrusted server. Data Perturbation is to balance privacy aegis and data utility. The kNN-R algorithm is designed to function with the RASP range query algorithm to work the kNN queries.

2. Related Work

2.1 Existing System

Actually we require the query processing in the cloud to satisfy privacy and efficiency at Low processing cost. But considerably processing complications will be raised. Recently few mechanics were proposed to satisfy these requirements. But they have not reached the actual needs. The crypto index and Order-Preserving-Encryption are weak in prevention of attacks, increases the computational complexity to improve the privacy. The Casper-method uses cloaking boxes to secure information and queries, this reduces query efficiency.

Disadvantages of Existing System:

- Malicious users may attack on the data or queries.
- Weak in prevention of attacks.
- Increases the computational complexity to improve the privacy

2.2 Proposed System

We propose a secured and efficient Query processing Service in Cloud with Arbitrary Space Decomposition technique for range queries and nearest neighbour queries. The proposed technique satisfies the query processing in the cloud with privacy and efficiency at Low processing cost, without increasing processing complexities Seed Block Algorithm supports protection on well-organized range-queries, nearest neighbor search queries. The fundamental inspiration is to arbitrarily change the multi-dimensional information, order preserving encryption, volume extension, adding noise, and protect queries processing. The main mechanisms used are the discomposure, creating secure range queries and creating secured nearest neighbour queries.

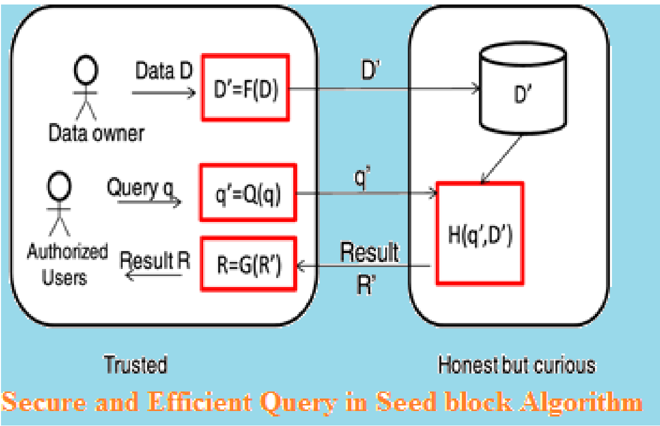


Fig 1: System Architecture.

The System architecture shown in Fig.1.It contains sundry clients, repository (web accommodation), main database, users and architecture is expounded as follows. The client application can be ported to any other machine like laptop or handheld contrivances. The stored data is platform independent that are sent to a central repository. When connected to network, the client application is authenticate into a central repository utilizing a web accommodation and submit all amassed information [10]. And if the central repository lost its data under any circumstances either of any natural calamity (for ex - earthquake, flood, fire etc.) or by human attack or effacement that has been done mistakenly and then it utilizes the information from the remote repository. The main objective of the remote backup facility is to avail utilizer to accumulate information from any remote location even if network connectivity is not available or if data not found on main cloud [12]. As shown in Fig-1 clients are sanctioned to access the files from remote repository if the data is not found on central repository (i.e. indirectly).

Advantages of Proposed System:

- The Discomposure integrates order conserving encryption, volume extension, and arbitrary noise additions.
- Provides strong confidentiality guarantee.
- computational complexity is less

- High precision on results

2.3 Algorithm Definition

Seed Block Algorithm

Simple but power full cryptography algorithm called the Seed Block Algorithm is proposed to reduce the computational overhead. Any transaction between the data owner, user, data Server is encrypted using the Seed Block algorithm.

SBA Algorithm

Algorithm 1:

Initialization: Main Cloud: Mc;
Remote Server: Rs;
Customers of Main Cloud: Ci;
Files: a1 and a1';
Seed block : Si; Random Number: r';
Customer's ID : Client_Idi
Input :a1 created by ci; is generated at; Mc
Output :Recovered file a1 after deletion at ci
Given : Authenticated clients could allow uploading,Downloading and do modification on its own the files only.
Step 1 :Generate a random number.
Int r=rand ( );
Int r=rand ( );
Step 2 :Create a seed Block Si for eachCi and Store
Si at Rs
Si=r xorClient_Idi / / (Repeat step 2 for all clients)
Step 3 : If Ci /admin creates/modifies an a1and stores atMc, then create as
a1'= a1 xor Si
Step 4 :Store a 'at Rs.
Step 5 : If server crashes a1 deleted from Mc,

In cloud computing, data is generated in that form of an electronic are large in amount. In cloud we have to maintain any data confidentiality or efficiency necessary of a data recovery services. Here seed block algorithm can also be mainly used to develop a smart remote data backup algorithm; this algorithm is also called as a seed block algorithm. The objective of seed block algorithm is double in the first place is clients to gather any data from any remote area without system integration or disappointments just and second to recuperate the documents in the event of the record erasure or if the cloud gets crushed because of any reason The SBA is mainly focused on the storing of the backup files in the remote server only, without using the existing encryption techniques.

Advantages of this algorithm is

- Recover same size data
- Low host
- Privacy
- Data security
- Remote Data Backup Server

Here data security can Giving full security to the client's data and this data also present in the utmost priority for the remote backup server either intentionally or unintentionally, here third party user or any un authorized users cannot access the data in the remote backup server only. The main advantages of remote backup server are we can access any data from remote backup server, this server is not connected to the network connectivity also, we can get the data from the remote backup server only.

The work flow of the above algorithm is:

Here we have to maintain one of the remote back-up server this is mainly used to storing of the user data, when the user login in to the cloud it can automatically generates the random id Int r=rand ( );and user id also, based on this user id and random id it has automatically generated by the seed block key

Create a seed Block Si for eachCi and Store
Si at Rs
Si=r xorClient_Idi

After the completion of this thing whenever user upload the data in to the remote back-up server with the help of user generated seed block key only the uploaded can be decrypted. For assumption we have to hack a remote back-up server that is not user understandable, here index format also encrypted using that of the md5 algorithm.

3. Implementation

Utilize is the time of the task when the theoretical setup is transmuted out into a working framework. Therefore it can be thought to be the most principal stage in accomplishing a cogent incipient structure and in giving the client, sureness that the incipient framework will work and be persuading.

The execution stage incorporates watchful organizing, examination of the current structure and its prerequisites on use, arranging of systems to achieve changeover and evaluation of changeover frameworks

The system is proposed to have the following modules along with functional requirements,

1) Information Transfer

The data owner uploads the special data into the Location Server. While the data is transfered the consummate data is encrypted utilizing his Seed Block. The server engenders index for this information. We have to upload any data at the time we will give the index keywords. These keywords are mainly used to probing of the data from database. These index keywords are automatically encrypted with the avail of md5 algorithm. The whole data will be stored in the cloud server in the format of encrypted.

2) Key Procreation

When an incipient utilizer joins into the accommodation, a unique secrete key (Seed block key) is engendered for confidential communications. This module engenders the Seed Block key utilizing the hash value of Utilizer ID, a desultorily culled Integer. The main utilization of the seed block key is whenever utilizer wants to upload data in to the cloud server with the avail of arbitrary key and hash key the whole data will be encrypted. With the avail of seed block algorithm we can reduce the time intricacy to the encrypted information, because in subsisting system we have to encrypt any data character to character.

3) Query Procreation

Whenever utilizer wants to find any information he engenders query (q), includes his query information and his goal point. The utilizer query is encrypted by the utilizer's secrete key and then the query is submitted the data server. The Location Server decrypts the query utilizing the same secrete key. Extent inquiry is the question used to recuperate the information from the database. It willinstaurate the information esteem that is between the upper bound and lower bound. The extent question is not mundane in light of the fact that client won't ken ahead of time about the result for the inquiry, the amount of passages will come as result for the inquiry.

4) Query Transform

After Receiving a Range Query from the client , the data Server extracts the obligatory inputs required for query transform , finds the more proximate goal points as outer-rage (grid) and the encrypted replication is send back to the client. Inquiry is chiefly used to look. Inquiries are developed by utilizing organized inquiry dialect. It is for the most part used to recuperating the required data from the database. Question administrations are the system for administrations that are unearthed through an execution of administration supplier. Here by utilizing Grate, reach question and kNN inquiry in cloud give secure, expeditious putting away and recuperating procedure of encryption and unscrambling of an information from database.

5) Cryptography

Simple but power full cryptography algorithm called the Seed Block Algorithm is proposed to reduce the computational overhead. Any transaction between the data owner, utilizer, Location Server is encrypted utilizing the Seed Block algorithm’s-Tree. The target of proposed calculation is twofold, first it avail the clients to accumulate data from any remote area without system integration and second to recuperate the records in the event of the document erasure or if the cloud gets crushed because of any reason. The time connected quandaries are likewise being light by orchestrated SBA designated it'll put aside least time for the rejuvenating method. Orchestrated SBA in integration concentrates on the safety plan for the move down documents place away at remote server, while not utilizing any of this coding methods.

3.1RANGE QUERY

Range query is the query used to retrieve the information from the database. It will retrieve the information value that is between the upper bound and lower bound. The range query is not customary because utilizer won't ken in advance about the result for the query, how much ingressions will come as answer for the query. For example

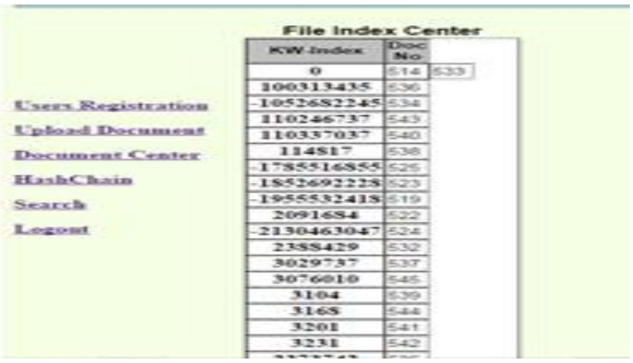
SELECT id
FROM table name
WHERE id (
SELECT top 10*
FROM United States
WHERE age >50
);

The above example expresses the sample query for range query. Here the example query is to retrieve the ingressions from Amalgamated States it will retrieve the persons who are above 50 years in the top 10 list from the record of Amalgamated States. The range search is mainly used to return the values that are present between the two designated values given in the query. For example database designation is AAAworkers2012 then

Go
SELECT product id
FROM AAAworkers2012.production
WHERE price BETWEEN 40 and 60

The above example will show another example of range query probe it will provide the ingressions of what are product id that are present in engenderment database with price above 40 and within 60. So by utilizing range query utilizer can facilely retrieve the data's from records and this query treat will be done in secure mode plus the accelerate of the query process will withal increase.

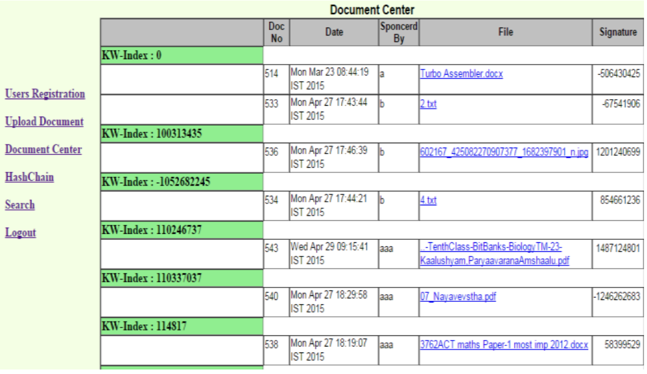
4. Experimental Work



File Index Center

KW-Index	Doc No	
0	514	533
100313435	536	
-1052682245	534	
110246737	543	
110337037	540	
114817	538	
-1785516855	525	
-1852692228	523	
-1955532418	519	
2091684	522	
-2130463047	524	
2388429	532	
3029737	537	
3076010	545	
3104	539	
3168	544	
3201	541	
3231	542	

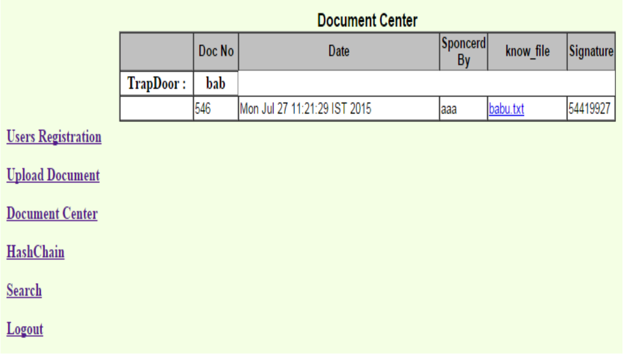
Fig 2: shows the index hash key value.



Document Center

	Doc No	Date	Sponcerd By	File	Signature
KW-Index : 0					
	514	Mon Mar 23 08:44:19 IST 2015	a	Turbo Assembler.docx	-506430425
	533	Mon Apr 27 17:43:44 IST 2015	b	2.txt	-67541906
KW-Index : 100313435					
	536	Mon Apr 27 17:46:39 IST 2015	b	602167_425082270907377_1682397901_n.jpg	1201240699
KW-Index : -1052682245					
	534	Mon Apr 27 17:44:21 IST 2015	b	4.txt	854661236
KW-Index : 110246737					
	543	Wed Apr 29 09:15:41 IST 2015	aaa	...-TenthClass-BitBanks-BiologyTM-23-Kaalushyam.ParyaavaranaAmshaalu.pdf	1487124801
KW-Index : 110337037					
	540	Mon Apr 27 18:29:58 IST 2015	aaa	07_Nayavevstha.pdf	-1246262683
KW-Index : 114817					
	538	Mon Apr 27 18:19:07 IST 2015	aaa	3762ACT maths Paper-1 most imp 2012.docx	58399529

Fig 4: list of uploaded documents with his encrypted key.



Document Center

	Doc No	Date	Sponcerd By	know_file	Signature
TrapDoor :	bab				
	546	Mon Jul 27 11:21:29 IST 2015	aaa	babu.txt	54419927

Fig 5: Based on the index The documents will be open.

The Fig-3 shows the CPU utilization at Main Cloud and Remote Server. As shown in Fig-3 the Main Cloud's CPU utilization starts with 0% and as per the client uploads the file onto it then utilization increases; such that it has to check whether the client is authenticated or not, at the same the time it send request to Remote Server for the corresponding Seed Block. When request reached to Remote Server it started collecting the details as well as the seed Block and gives response in form of the seed Block and during this period, load at Main Cloud decreases which in return cause for gradual decreases in CPU utilization at main cloud. After receiving the requested data, CPU utilization at main cloud increases as it has to perform the EXORed operation. Again the Final EXORed file sends to Remote Server. As compared to Table-IV the processing time given can be compare with the time showing in Fig-3.

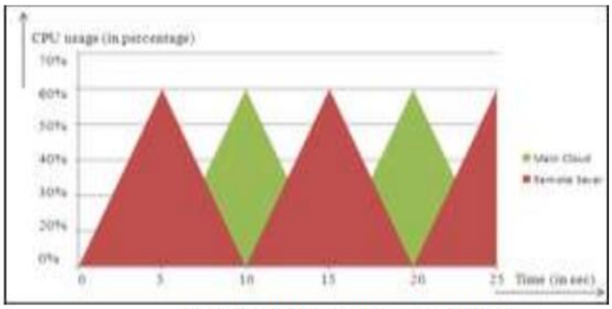


Fig.3 Graph Showing Processor Utilization

5. Conclusion

In this Paper, we presented detail design of proposed SBA algorithm. Proposed SBA is robust in availing the users to accumulate information from any remote location in the absence of network connectivity and additionally to recuperate the files in case of the file effacement or if the cloud gets eradicated due to any reason. Experimentation and result analysis shows that proposed SBA additionally fixates on the security concept for the back-up files stored at remote server, without utilizing any of the subsisting encryption techniques. The time cognate issues are withal being solved by proposed SBA such that it will take minimum time for the recuperation process.

References

1. HuiqiXu, Shumin Geo, Keke Chen, ”Building confidential and Efficient Query services in the Cloud with RASPData Parturbation” IEEE Transaction on knowledge and data Eng vol:26 no:2,2014.
2. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. K. andAndyKonwinski, G. Lee, D. Patterson, A. Rabkin, I.Stoica, and M. Zaharia, “Above the clouds: A berkeley view of cloud computing,” Technical Report, University ofBerkerley, 2009.
3. H. Hacigumus, B. Iyer, C. Li, and S. Mehrotra, “Executing sql over encrypted data in the database-service-provider model, ”in Proceedings of ACM SIGMOD Conference, 2002.
4. R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, “Order preserving encryption for numeric data,” in Proceedings ofACM SIGMOD Conference, 2004.
5. B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, “Private information retrieval,” ACM Computer Survey, vol.45, no. 6,pp. 965–981, 1998.
6. B. Hore, S. Mehrotra, and G. Tsudik, “A privacy-preserving index for range queries,” in Proceedings of Very Large Databases Conference (VLDB), 2004.
7. J. Bau and J. C. Mitchell, “Security modeling and analysis,” IEEE Security and Privacy, vol. 9, no. 3, pp. 18–25,2011.
8. K. Chen, L. Liu, and G. Sun, “Towards attack resilient geometric data perturbat-ion,” in SIAM Data Mining Conf, 2007.
9. M. F. Mokbel, C. yin Chow, and W. G. Aref, “The new casper: Query processing for location services without comprom-ising privacy,” in Proceedings of Very Large Databases Conference (VLDB), 2006, pp. 763–774.
10. B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, “Private information retrieval,” ACM Computer Survey, vol.45, no. 6, pp. 965–981,199