World Applied Sciences Journal 32 (4): 601-604, 2014
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2014.32.04.373

Secure Multi keyword Top-K Retrieval over the Compressed Encrypted Cloud Data

Amit Kumar Dimha and K. John Singh

School of Information Technology and Engineering,
VIT University, Vellore, Tamil Nadu, India

Abstract: With the great advantages of the cloud computing the cloud customer moving from personal site to the commercial site where the data owner can store and share his data with the other cloud customer. Cloud computing is responsible for the best security of the data which is stored over the cloud by the customer of cloud for the better security of data should be stored in encrypted form that reduce the privacy risk and leakage of the data. Cloud provides the flexibility for the customer. Always the encryption performs on binary data. Any person or organization who is the customer of cloud and remotely stores his data over the cloud server want secure the data by the best privacy policies different type of data customer stored on the cloud server that may be the video recording of someone, e-mails, official file, movie, future plan of organization, employee Information. Encryption of the data provides the security over the cloud and other users can retrieve the data using various method. This paper solves the problem of data management issue by the cloud server, because server always have burden related to data management, data take more time for encryption and decryption according to its size, if the data would be used in compressed form then it takes less time for encryption and decryption and reduce its size. Hence the server will not have more burdens and user can retrieve the compressed data by using multi keyword search.

Key words: Cloud · Cloud Customer · Compression Machine · Data Encryption · Privacy

INTRODUCTION

Cloud computing [1] provides the flexibility for the cloud customer. The cloud customer remotely stores the data and information on the cloud server. That may be video, photograph, e-mail, files or some other information. It required an application that should be trustful. This client application is responsible for all the operation. It provides the security for the data owner. This user should have to log in the application for the security.

Usually, problem occurs on that data that is related with data management, data loss and its privacy. A good cloud server should provide data security, index security, keyword security; manage the data and others security related issues [2]. Cloud works between the cloud customer and the cloud server and when the cloud customer stores the data over the cloud server that time cloud works for data manage and make its secure on server, for the better security of the data it should be encrypted before the outsourcing [3] on the cloud server. The data owner provide the authentication of the data access for the limited users or the limited number of organization that is already the customer of the cloud who want search and retrieve data as per the interest over the cloud that they would access so all users required the list of search item according of to the search.

By compressing [4] the data users can reduce the size of data. The main goal of compression is represent the data at least number of bits. Many method are already existing for the data compression now some software is also available that can reduce the size of file by compressing it. When user want share any video, movie or any other large file so it will take time and if data owner compress that file and reduce its size, low size data can share easily. User can create the password on data at the time of compression that will required when anyone want access that data. If the data owner want access that compressed data then password is mandatory.

Corresponding Author: Amit Kumar Dimha, School of Information Technology and Engineering,
VIT University, Vellore, Tamil Nadu, India.

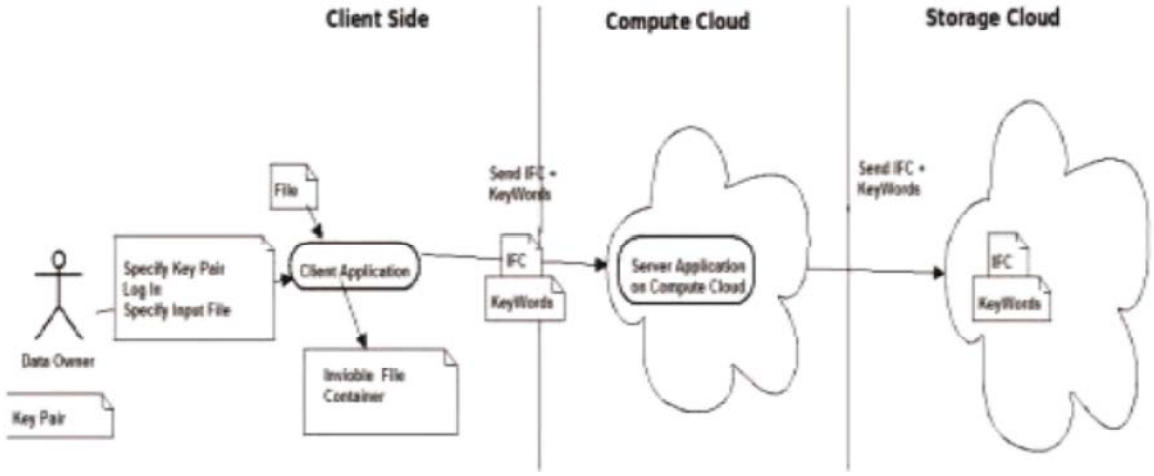

Fig. 1: Data owner-cloud server communication

Fig. 2: Data compression

This paper proposed a method that reduces the overhead of the data management [5] over the cloud server. Where first compress the data that user want outsource over the cloud by data owner, compression will reduce the size of our data and after the compression perform encryption [6]. On compressed data so it will take less time for encryption this will save the time. After the encryption outsource the data is outsource over the cloud server that time an index has created in the list form that is mapped with the keywords of the data [7]. Data and index both can access by the other cloud customer as index makes help for the data retrieval. These data always works with the privacy option that already created by the data owner related to the data access and users restriction and authentication. Data can retrieve by using multi keyword search method that is better than other existing method like that single keyword search, coordination search and others. In multi keyword search user input the set of characters that is divide and stores in fixe size of block that is matched on the data and sort the result.

Related Work: Many methods already exist for the retrieve of cloud data over the cloud server that helps for the other user who wants access different type of data as per the interest over the cloud.

Support similarity search used for the similar information retrieving over the cloud server [8]. It is based on fuzzy search formulation, if the user enter the n encrypted data files for search D= (d1,d2,d3,d4…..dn) and

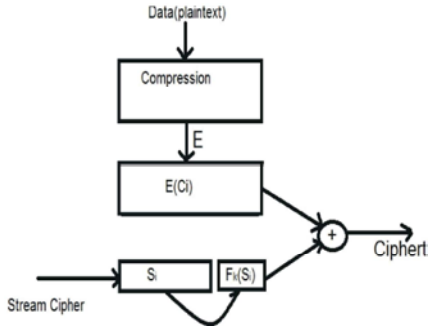

Fig. 3: Encryption of compressed data

the predefined set of that keyword on the server is S=(s1,s2,s3,s4……sn), a specific word s is searching then it will return the set of file.

Single Key-word search [9], in this search method store the hash value at place of the keyword at the time of index creation, the main drawback of this search is that it is not comfortable enough to express complex information need. Wang *et al.* [10] use hash chain to construct a single keyword search result verification scheme.

Boolean search, this method is used for the presence or absence of the searched (queried) data, this method can control the number of document that is matched for retrieving.

Multi keyword [11] search it search the each and every word that is inserted for search in every encrypted document over the cloud. Co-coordinating matching, when the data has encrypted and outsource the data on cloud server that time Index has created that don't describe that where is a particular word stored in data file that time users can not use coordinating and multi keyword search.

TOP-K Retrieval: The main goal of compression is reduce the size of data and represent it at least number of bits. Before outsourcing the data on the cloud server data owner want encrypt that data item for the better security and reduce the risk of data leakage. Once a data owner outsources the data item on the cloud server then other authorized user can retrieve that data as per the interest. For the data selection that is searched by the other user is a difficult job because server contains data in large amount that is outsource by the different data owners.

Reduce the overhead of data management and data retrieval by the other cloud. Customer use a method in that have 3 different entities that control the process, i.e, owner of the data (DO) who outsource his data over the cloud server, other user (OU) who retrieve the data from the cloud server and cloud server (CS) that manage and store the data. Data owner compressed the data after that encrypt that after the encryption create the index for the other users that is authorized as per the condition and remotely store on the cloud server, Cloud customer enter the keyword as per the interested (more than one keyword) that is divided by the fix size of the block and after that it access the index of the data item and matched it block by block and collect the most similar result that provided for the users

Keyword privacy, the most important thing is protecting the data history and the keyword that is searched by the other users and not shareable over the cloud customer. These all the things are managed by the cloud server, it provides the privacy for all the cloud customers.

In this proposed method, data owner should have to compress his data before the encryption and create an index for compressed data at the time of encryption. Other user who wants to retrieve that data over the cloud can search by entering the multi keyword, which words divided according to its length and stored in fixed length size block. Then after finds the compressed encrypted data from the index that is mapped with the keyword of the compressed encrypted data and perform the data matching block by block in each and every data file. When the authorized user retrieves the data by multi keyword searched over the compressed encrypted cloud data. It gives the most relevant result that is retrieving by the users in the descending orders that provide accuracy and flexibility for the other users who is retrieving the data over the cloud server.

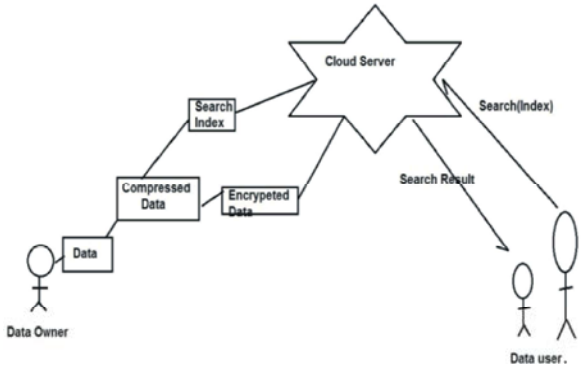

Fig. 4: Architecture of data outsourcing and searching

It provides top-k results in that have minor difference. Other user cannot judge data without retrieving it. If data has large size than user spend more time. To improve this type of problem this method helps users by saving his retrieving time. Because data user retrieves that data will in compressed form take less time for retrieving by the authorized user.

The given algorithm can search the top-k result of the compressed data that is searched by the cloud customer.

TOP-K (compressed data) selection

Input

Source list
Number kc;

Initialization

Set topcompk_0;
Set topcomkid_0;

Iteration:
For all data item_source do

Insert(TopcompK,(dataitem,dataitemindex));

End for
For all data tuple _ topcompk do
Topcompkid.append(data tuple[1])
End for

Analysis

Reduced Management Overhead: Compressed data requires a minimum space over the cloud server.

Outsourcing the compressed encrypted data at place of normal encrypted data reduce the overhead of the server. This is related with data management because cloud server provides the place for the user where they can outsource his data.

Time Analysis: Encryption and decryption take time according to the size of the data so if once the users reduce the size by the compression of data then encryption and decryption take less time and our system will work more effectively.

Commercial: If data owner required minimum space over the cloud server then it will give the financially and commercial benefit for the data owner.

Space: Cloud server provides the space for the customer who wants outsources and shares his data with the other cloud customer. If any customer want outsource video then it take more space on cloud and also take more time for data transfer. If any customer work with the compressed data then it take less space on cloud and also take less time for data transfer.

Data Transfer Rate: When users transfer data from one place to another place data take time according to the size of data so if user transfer the long data it will take more time in comparison of same data that is compressed.

CONCLUSION

This Method of outsourcing and retrieving the data from the cloud server reduce the burden of long size data storage on the server and save the time of encryption of data. This is remotely stored on the server by the data owner and also save the space on server that is provided by the server for the cloud customer. Other user can retrieve the data as per the interest easily by the keyword search, which will give all top result of compressed data that would have any relation with the given keyword.

REFERENCES

1. Rao Mikkilineni and Vijay Sarathy Kawa Objects, Inc. Los Altos, CA.Cloud Computing and the Lessons from the Past., 18th IEEE International Workshops 2009.
2. AHN, "Romney Hits Obama for Security Information Leakage," http://gantdaily.com/2012/07/25/romney-hits-obama-forsecurity-information-leakage/, 2012.
3. Cao, N., C. Wang, M. Li, K. Ren and W. Lou, "Privacy-Preserving Multikeyword Ranked Search over Encrypted Cloud Data,"Proc. IEEE INFOCOM, 2011.
4. Amir, A., 1996. "Let sleeping files lie: Pattern matching in zcompressed files,"Journal of Computer and System Sciences, 52: 299-307.
5. Cong Wang, Qian Wang and Kui Ren, 2011, "Towards Secure and Effective Utilization over Encrypted Cloud data", 31st International Conference on Distributed Computing System Workshop.
6. Song, D., D. Wagner and A. Perrig, 2000, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy.
7. Zerr, S., D. Olmedilla, W. Nejdl and W. Siberski, 2009. "Zerber+r: Top-k Retrieval from a Confidential Index," Proc. 12th Int'l Conf. Extending Database Technology: Advances in Database Technology (EDBT),
8. Zhang, Z., *et al*., 2010. "Bedtree: an All-Purpose Index Structure for String Similarity Search based on Edit Distance," Proc. SIGMOD, pp: 915-926.
9. Lu, Y., 2012. "Privacy-preserving logarithmic-time search on encrypted data in cloud," inProc. of NDSS.
10. Wang, C., N. Cao, K. Ren and W. Lou, 2012. "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE TPDS, 23(8): 1467-1479.
11. Sun, W., B. Wang, N. Cao, M. Li, W. Lou, Y.T. Hou and H. Li, 2012. "Privacy-preserving multi-keywordd text search in the cloud supporting similarity-based ranking" in proc. Of Acm Asiaccs, pp: 71-82.