An Efficient Privacy Preserving Encryption of One to Many Orders Based on Cloud Data Search

Gurram Sai Suneetha[1], K.Jagdeeshwara Rao[2]

Final M.Sc. Student[1], Lecturer[2]

[1, 2] M. Sc Computer Science, Chaitanya Women's PG College, Old Gajuwaka, Visakhapatnam Andhra Pradesh

Abstract:

Now a day's almost everything can be shared through the cloud and also provide more security. Before sharing everything we can stored into cloud servers and the servers are manager or process their data. They can easily maintaining and storing data into cloud servers, to provide an efficient service without paying too much money and energy, as one of the most attractive feature of cloud computing. So that large number of people worry about this technology for provide security of data. If cloud server gets direct access to all these users' data, it may try to analyses the documents to get private information. The initial purpose of this action may be kind. The server wants to provide better service by digging into these data and then displaying customer-oriented advertisement, which could be convenient but also annoying. Besides, when we consider sensitive data such as personal health records and secret chemical ingredients, the situation becomes even more serious. Theoretically, the server is not supposed to have access to sensitive data at all; therefore we should ensure the server has no access to leaking these data to an untrusted third party. Thus, sensitive data have to be encrypted before being outsourced to a commercial public cloud. However to provide security of sensitive data we can encrypt and processing of the data. We can retrieve the encrypted data is difficult process because the amount outsourced files can be large and traditional search patterns cannot be deployed to cipher text retrieval directly. To overcome this, we are proposed an encrypted searchable query domain while preserving user's privacy. In this paper we are proposed an efficient one to many orders preserving encryption technique for searching query related documents in the cloud. By implementing this process we can retrieve out sourced large files easily and download required documents.

Keywords: *Privacy Preserving, One to Many Order, Searching, Cryptography, Encryption and Decryption.*

I. INTRODUCTION

Now a day almost everything is moving to cloud. Cloud has been the most scalable and cost effective way to store our data. No extra work is required to store the data on cloud as almost everything on cloud is automatic. But the real concern of cloud is that cloud is managed by the cloud service providers and everyone is now thing of store the data on the cloud making cloud too much available for public. Even if a person is having data securely in the laptop a person prefers to take the back up of data on the cloud for the reason that if data is lost or get crash in his own laptop or system. Another reason for preferring cloud is that data is increasing so fast that there is problem of storage on personal system and organization. Thus if sensitive data such as chemical formulae, combinations, patents, medical history of data, bank statements, password etc. are store in cloud it may prone to attack by the attackers so it is very necessary to address the security of the system. In this paper we will discuss different techniques to search the encrypted data over the cloud so that others can search over the data and get the result in a relevant manner. This will eliminate the need to download the data and then decrypt it Data or file sent to cloud is in encrypted form and downloaded in the encrypted form only and then it is decrypted by the owner. But what if user or a particular system wants to search something on cloud? Manually or traditionally we can say download all the files that are relevant for the search and then decrypt it. If the item that is to be searched in not found then again the same process will be repeated. But this eliminates the security and privacy of the data and there is more irrelevancy of document. Thus we need a technique where user can search an item in the cloud when the file is in encrypted form only i.e. no need to download each and every file and then decrypt it. Also searching should be in such way that it should return the user the most relevant search first same as google does.

Traditional encryption search techniques such as Searchable Encryption, OPE are good and preserves security of the document but have limitation with ranked search. For this reason we proposed a system that will effectively return the search result according to the most relevant document. For ranked search in encrypted cloud data, order preserving encryption (OPE) is an efficient tool to encrypt relevance scores of the inverted index. When using deterministic OPE, the cipher texts will reveal the distribution of relevance scores. We will implement an advanced form of OPE to eliminate above limitation. Also in the paper it is analysed that the security over one-to-many order preserving is more prone over the differential attack so we will propose a system where it will ensure the security over the differential attacks.

II. RELATED WORK

Clustering is a fundamental form of data analysis that is applied in a wide variety of domains, from astronomy to zoology. With the radical increase in the amount of data collected in recent years, the use of clustering has expanded even further, to applications such as personalization and targeted advertising. Clustering is now a core component of interactive systems that collect information on millions of users on a daily basis . The ultimate aim of the clustering is to provide a grouping of similar records. Clustering is a technique in which, the information that is logically similar is physically stored together. In order to increase the efficiency of search and the retrieval in database management, the number of disk accesses is to be minimized. In clustering, since the objects of similar properties are placed in one class of objects, a single access to the disk can retrieve the entire class. However, with the never ending data in today's era it is becoming impractical to store all relevant information in memory at the same time, often necessitating the transition to incremental methods called Incremental Clustering.

Clustering problems arise in many different applications, such as data mining and knowledge discovery, data compression and vector quantization, and pattern recognition and pattern classification. The notion of what constitutes a good cluster depends on the application and there are many methods for finding clusters subject to various criteria, both ad hoc and systematic. These include approaches based on splitting and merging such as ISODATA, randomized approaches such as CLARA , CLARANS , methods based on neural nets , and methods designed to scale to large databases, including DBSCAN BIRCH, and ScaleKM . For further information on clustering and clustering algorithms, see. Among clustering formulations that are based on minimizing a formal objective function, perhaps the most widely used and studied is k-means clustering. Given a set of n data points in real d-dimensional space, Rd, and an integer k, the problem is to determine a set of k points in Rd, called centers, so as to minimize the mean squared distance from each data point to its nearest center. This measure is often called the squared-error distortion and this type of clustering falls into the general category of variancebased clustering. Clustering based on k-means is closely related to a number of other clustering and location problems. These include the Euclidean k-medians (or the multisource Weber problem) in which the objective is to minimize the sum of distances to the nearest center and the geometric k-center problem [1] in which the objective is to minimize the maximum distance from every point to its closest center. There are no efficient solutions known to any of these problems and some formulations are NP-hard. An asymptotically efficient approximation for the k-means clustering problem has been presented by Matousek , but the large constant factors suggest that it is not a good candidate for practical implementation. One of the most popular heuristics for solving the k-means problem is based on a simple iterative scheme for finding a locally minimal solution. This algorithm is often called the k-means algorithm. There are a number of variants to this algorithm, so, to clarify which version we are using, we will refer to it as Lloyd's algorithm. (More accurately, it should be called the generalized Lloyd's algorithm since Lloyd's original result was for scalar data.

Fast and robust clustering algorithms play an important role in extracting useful information in large databases. The aim of cluster analysis is to partition a set of N object into C clusters such that objects within cluster should be similar to each other and objects in different clusters are should be dissimilar with each other[1]. Clustering can be used to quantize the available data, to extract a set of cluster prototypes for the compact representation of the dataset, into homogeneous subsets. Clustering is a mathematical tool that attempts to discover structures or certain patterns in a dataset, where the objects inside each cluster show a certain degree of similarity. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Cluster analysis is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization. It will often necessary to modify pre-processing and parameter until the result achieves the desired properties. In Clustering, one of the most widely used algorithms is fuzzy clustering algorithms. Fuzzy set theory was first proposed by Zadeh in 1965 & it gave an idea of uncertainty of belonging which was described by a

membership function. The use of fuzzy set provides imprecise class membership function. Applications of fuzzy set theory in cluster analysis were early proposed in the work of Bellman, Zadeh, and Ruspini This paper opens door step of fuzzy clustering . Integration of fuzzy logic with data mining techniques has become one of the key constituents of soft computing in handling challenges posed by massive collections of natural data. The central idea in fuzzy clustering is the no unique partitioning of the data into a collection of clusters.

III.PROPOSED SYSTEM

In the proposed system we are implementing an efficient one to many orders preserving encryption over the cloud data. Before performing the searching process each user or group member will verify by the data owner. The data owner will generate secret key and send that key to all verified by members in a group. After sharing the secret key the data owner will encrypt the data by using cipher format key and secret key. By using those two keys the data owner will encrypt the data and stored into cloud. After completion of encryption process each user or group member will retrieve secret key and enter the search query for performing searching operation. Before retrieve the query related documents the search engine will decrypt the file or data and download the required file. By implementing those concepts we can improve the privacy of data or file and also provide an efficient search process.

Users Registration and Verification process:

In this module each user will enter the personal information for the registration process. After completion of registration process each user will get his usernames and passwords. By using those usernames and passwords we can provide login credential to each user in the group. The data owner will see the group member details and generate universal key (U_i) for each group members. The data owner will randomly generate universal key and send that key to individual group members. Each group member will login by using username and password, getting universal key. Each group member or user will choose his own two prime numbers P_i ,G_i and also choose one private key (a_i). By using those values each group member will generate public key by using following formula.

$$pub_i = G_i^{\ ai} \bmod P_i$$

After generating the public keys of each member it will send that public key to data owner. The data owner will retrieve the public keys of each user and generate another public of individual user. Before generating public key the data owner will generate private keys ($private_i$) randomly. Using those private keys the data owner will generate public keys of individual users by following formula.

$$pubkey_i = pubi^{\ privatei} \bmod P_i$$

After completion of public keys the data owner will send those keys to individual users. Each user will retrieve public key and generate shared key. The generation of shared key is as follows.

$$sharedkey_i = pubkey_i^{\ ai} \bmod P_i$$

After generation of shared key, each group member will generate signature by using following formula.

$$sig_i = hash\ (ID \mid U_i \mid pub_i \mid sharedkey_i)$$

The completion of generating signature each group member will send that value to data owner. The data owner will retrieve signature and verify the each user will again generate signature. The generation signature can be done by using above formula and take that signature and verify. After completion of verification process the verification status will send to the each group member.

Group Key Generation Process:

In this module the data owner will generate secret key for all users in a group. Before generating group key the data owner will get all shared keys of each user and perform the xor operation. The generation of group key is as follows.

$K = sharedkey_1 \ ® \ sharedkey2 \ ® \ \ldots\ldots \ ® \ sharedkey_n$

After generating the group key the data owner will send that key in the form the secret points. The generation of secret points is as follows.

$$X = K / P_i$$
$$Y = K \% P_i$$

Take those X, Y values and send those secret points to individual users.

Data Encryption and Stored into Cloud:

In this module the data owner will perform the encryption process and stored the encrypted file format data into cloud server. Before performing the encryption process the data owner will use two keys and the keys are name of file, group key. By using those keys the data owner will encrypt the file or data and stored into cloud server. So that the encryption process can be done by two times and using those two keys. Before performing the

encryption process the data owner will use first key is name of file. Here the name of file will not use the directly for the encryption process. So that the data owner will retrieve name of file and file key encryption process, it will get cipher format first level key. The key encryption process is as follows.

1. Get each character from the file name and convert into decimal format.

2. Take the each decimal values and convert into 8 bit binary format. [Binary value should be 8 digits (no matter how much the length of it, we should represent it in 8 digits. (28=256). e.g. for decimal 32 binary number should be 00100000 (underlined zeros are required)]

3.Take the eight bit binary format data and perform the reverse process.

4. After completion of reverse process, split the 8 bit binary data into two equal parts.

5. Take each part again performs the reverse process and performs the xor operation with first part and second part.

6. The first part and xor result will be combining to get 8 bit binary data.

7. Take those 8 bit binary data and convert into decimal value.

8. Take that decimal value and get character and used that cipher format data into encryption process.

The data owner will take those cipher formats key and perform the encryption process. By performing encryption process we are using data encryption standard algorithm (DES). After completion of encryption process the data owner will take group key and again perform the encryption process using des algorithm. After that the data owner will stored that data into cloud server.

Group key Generation, Decryption and Searching process:

In this module the group members will retrieve the verification status and the status if true it will retrieve the secret points. If the status is false it will not get secret points and do not perform the any operation regarding searching process. Suppose to status if true get the secret point (X_i, Yi) and generate the group key. The generation of group key is as follows.

$$K= X_i * P_i + Y_i$$

After getting the group key the group member will enter search related query. The search engine will take that query and perform the searching process. Before performing searching process the search engine will perform the decryption process. By applying decryption process the search engine will contains plain format data and perform the searching process. The search engine will take each file name and generate first level cipher format key by using above encryption process. After generating first level cipher format the search also take group key and perform the decryption process first by using group key. By applying first level encryption using data encryption standard algorithm decryption process. After completion of first level encryption the search will take cipher format second level key and perform the decryption process. Take that plain format file and perform the searching process by using following process.

```
CharactershuffleMatch (T, P, Σ)
 L ← lastOccurenceFunction (P, Σ )
 i ← m − 1 j ← m − 1
repeat
if T[i] = P[j]

if j = 0
 return i {match at i }

  else i ← i − 1 j ← j − 1

else

{Character-jump}

l ← L[T[i]]

 i ← i + m – min(j, 1 + l)
j ← m – 1
 until i > n − 1
return −1 {no match}
```

After finding the each query related file or documents those files will be displayed and download the required file. By applying those processes we can provide more security of stored data and also provide efficient searching operation for respected query related documents

IV. CONCLUSIONS

In this paper we are proposed an efficient one to many orders preserving encryption schema for provide more security of data and also perform an efficient searching operation. Before performing searching process each group member will verified by the data owner and the data owner will provide group key for all users. In this paper we are proposed many order encryption process for provide more security of stored data into the cloud server.

Before perform the encryption process the data owner will choose the upload file and also generate the first level encryption key based on file name. The data owner will take the uploaded file name and perform the encryption process. By applying encryption process we can get cipher format first level encryption key. Take that key perform the first level encryption of data and take another group key for second level encryption process using data encryption standard algorithm. After completion encryption process the data owner will stored that cipher format file into cloud server. If the user enter search query and the search engine will take that query and perform two time decryption process using des algorithm. By completion decryption process the search will search each document and get query related document. After completion of searching process the group member will choose required document and download it. By proposing those concepts we can improve efficiency in the searching process and also provide more security.

REFERENCES

[1]. N Cao, C. Wang, M. Li,K.Ren, and W. Lou, "Privacy preserving keyword ranked search over encrypted cloud data," IEEE Trans. Parallel Distrib. Syst., vol, 25, no. 1, pp. 222-223,
Jan. 2014
.

[2] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou,"secure ranked keyword search over encrypted cloud data," in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst. (ICDCS), jun.2010, pp. 253-262.

[3] B. Wang, S. Yu, W.Lou, and Y.T.Hou, "privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in Proc. IEEE INFOCOM, Apr./May 2014, pp. 2112- 2120.

[4] C.Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE Trans.Parallel Distrib. Syst., vol.23, no.8, pp. 1467- 1479, Aug. 2012.

[5]. K. Srinivasa Reddy and S. Ramachandram "A New Randomized Order Preserving Encryption Scheme" In
International Journal of Computer Applications (0975 – 8887) Volume 108 – No 12, December 2014.

[6] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," In Proc. ACM SIGMOD Int. Conf. Manage. Data, 2004, pp. 563–574.

[7]. A. Boldyreva, N. Chenette, Y. Lee, and A. O" Neill, "Order preserving symmetric encryption," In Advances in
Cryptology. Berlin, Germany: Springer-Verlag, 2009, pp. 224– 241.

[8] Raluca Ada Popa , Frank H. Li , Nickolai Zeldovich "An Ideal-Security Protocol for Order Preserving Encoding" In Proc. of the 34th IEEE Symposium on Security and Privacy.

[9] K. Srinivasa Reddy and S. Ramachandram "A novel Dynamic Order-Preserving Encryption Scheme" In Networks & Soft Computing (ICNSC), 2014 First International Conference on 19- 0 Aug. 2014.

[10]. R. Agrawal, J. Kiernan and R. Srikant, "Order preserving encryption for numeric data," *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*. ACM, pp. 563-574, 2004.

BIOGRAPHIES:

Gurram Sai Suneetha is student in M.Sc. (Computer Science) in Chaitanya Women's pg college, Old Gajuwaka, Visakhapatnam. She has received his Degree B. Sc (MSCS) from Chaitanya Degree College for women's, old Gajuwaka, Visakhapatnam. Her interesting areas are data mining, network security and cloud computing

K.Jagdeeshwara Rao is working as a Lecturer in Chaitanya Women's pg college, old Gajuwaka, Visakhapatnam Andhra Pradesh. He received his M.Sc. (Computer Science) from Gayatri Vidya Parishad, Andhra Pradesh. His research areas include Network Security and Computer Networks