ISSN:0975-9646

Encryption based Cloud Data Search Technique for Privacy Preserving

Sandhya Pradip Mohite,
Department Of Computer Engineering
MIT Academy Of Engineering,Alandi(D),
Savitribai Phule Pune University.

Dr. Sunita S. Barve
Department Of Computer Engineering,
MIT Academy of Engineering, Alandi(D),
Savitribai Phule Pune University.

***Abstract*—Cloud have large storage space for storing huge amount of data.Data owner outsource their data contents on cloud server.Cloud server can have huge storage space.In this paper,data user request for data to the cloud server.Data owner produce encryption key for every data user which requested for data files.Due to this, data remains secure and accurate data may searched by semantic search.Also the information leakage is minimized due to the cloud storage system. RSA algorithm can be used for encryption technique.Key generation is performed by data owner for privacy preserving.**

Keywords*—Encryption technique, Key generation, Security, Encrypted documents.

I. INTRODUCTION

Information leakage is problem in big data environment. Encryption of data is common method to reduce information leakage searching encrypted documents on the server side is big challenging task[1]. Many cryptographic techniques are developed in past, but these techniques are much complex and time consuming[2].To preserve relationship between original and encrypted documents over cloud environment to improve search efficiency MRSE-HCI technique is discussed.MRSE-HCI is multi-keyword ranked search over encrypted databased on hierarchical clustering index[3].

In MRSE-HCI search time is increases linearly as aggressive growing size of data collection[6].In this project aim is to increase document searching speed by calculating relevance score between user query and documents. Due to relevance score evaluation user gets the document related with user query[7].Therefore, irrelevant fields get ignored which tend to increase the searching speed.Main aim of maintaining relation between different plain documents and encrypted document can achieved using clustering method.

Relevance score metric is used to calculate relationship between different documents.Problem in this technique is constraint on the cluster may break if any document added to the cluster. Cluster center is created dynamically and then number of cluster is further decided by attributes and property of dataset[8].

Hierarchical method is utilized to get better clustering result within larger amount of data collection.In hierarchical clustering methods number of clusters and minimum relevance score increases as increase in the level of maximum size of cluster reduced.If cluster exceeded its size level,it will be further divided into several sub-clusters[9].So ranked privacy preservation strategy followed. Searching user query document is an iterative processes in which system evaluates the relevance score between query and document included into the small cluster[7].

If the document in small cluster does not satisfies the user query document then the system again search back to its parent cluster.After whole searching procedure there is one more classification required for the most frequent document extraction hence the user query documents are ranked by system to make searching efficient and flexible[6].

Finally they were contributing themselves to make investigation to maintain relationship between plain documents over encrypted documents by processing clustering method[9].They utilized MRSE-HCI mechanism to speed up the searching operation.In this backtracking algorithm introduced to improve searching strategy with ranked privacy. Vector space model is used, every document is represented by vector. Relationship between different documents are classified into several categories[8]. Due to desired document categories, document search time is reduced.

Due to the small number of documents, cluster can be categorized into sub-categories[5]. Cloud server first search document in cluster. Cloud server will select the desired k-document. The value of k is previously decided by user and send to server[10]. If document cannot find in nearest cluster, it goes for sub-clusters. Further kdocument is not satisfied then, cloud server will trace back to the parent node and select the desired document. This process repeated recursively until respected k-document get satisfied[3].

II. REVIEW OF LITERATURE SURVEY

W.K.Wong,introduces kNN queries.kNN queries are used for encryption technique[1].The encryption technique can be developed to security support kNN application under the SCONEDB model(Secure Computation ON an Encrypted DataBase). A kNN query searches for k points in the database which are nearest to given query point q.Each database tuple can be modeled as multi-dimensional

point[2].To security support kNN,one approach is used such that Distance Preserving Transformation(DPT) to encrypt points E(DB) is same as that between corresponding original points in DB.kNN can be computed on encrypted database[4]. Unfortunately, such transformation is not secure in practice.

If attacker can access DPT encrypt database E(DB) and knows few points in plain database DB, he can get or recover DB entry[8]. A.Swaminathan,introduces framework for confidentiality preserving ranked-ordered search and retrieval large document collections. This not only protect documents or query confidentiality from intruder, but also protects an untrusted data centers from learning information about query.In this paper,cryptographic technique and relevance scoring is introduced. This technique used for preserving encryption,to protect data and indices.Also provide efficiency and accurate search to secure rank-order documents for requested query[6]. But critical issue is that to protect data collection and indices through encryption while providing efficient and effective search capabilities to authorize users.

Cryptographic encryption used to protect data from intrusion.For example,if information storage is outsourced to the third party data centers,then other users and system administrators involved may not be trusted to access the data contents[7]. Although traditional searchable encryption schemes to list a few) allow a user to securely search over encrypted data through keywords without first decrypting it,these techniques support only conventional Boolean keyword search1,without capturing any relevance of the files in the search result.When directly applied in large collaborative data outsourcing cloud environment,they may suffer from the following two main drawbacks[9].

On the one hand, for each search request, users without pre-knowledge of the encrypted cloud data have to go through every retrieved file in order to find ones most matching their interest, which demands possibly large amount of post processing overhead;On the other hand,invariably sending back all files solely based on presence/absence of the keyword further incurs large unnecessary network traffic[4].In this paper, data owner outsource its data to server but data owner introduce to allow clients to search the database such that client learn information in data owner[8].

Sun et al. introduces privacy constraints of the system.In this paper PKC and SKC search algorithms are introduced. In PKC, keys are generated but user can access any documents from cloud server without authenticated by data owner.

Cao.et al. produce SKC based encryption searching algorithm support for multi-keyword ranked search index[11].

Further more,Extending this model and clients queried value need to be hidden from data owner.Cloud computing brings users with many benefits such as the relief of the storage load and flexible data access,which motivate users to store their local data into the cloud[1]. As the cloud services become prevalent,more and more sensitive information, such as personal photos,government records and finance data, are outsourced into the cloud.

Boneh et al. introduces public-key encryption technique.PKC keyword search algorithm which is single keyword algorithm,but any user having public key can write or access data on the server[3].
Curtmola et al. produces searching scheme of single keyword queries.Karmara et al. support for dynamic addition and deletion of data files.

To protect the privacy of the sensitive data in the cloud, the data has to be encrypted by the data owner before outsourcing to the cloud.However,data encryption makes effective data utilization a challenging task when a large amount of data files are present users may have to download the whole data set from the cloud and then decrypt it to conduct keyword search over the data, which is very inefficient when the number of data files is large[9].Thus, effective keyword searching over encrypted data is of paramount importance,especially need to provide efficient ranked multiple keyword search, which supports a set of input keywords and achieves high efficiency simultaneously in users search behaviors. enabling the keyword search over encrypted data is not an easy task. Some techniques allow the user to search over encrypted data securely through single keyword to retrieve documents of interest[10].

This is insufficient as many users may tend to provide multiple keywords instead of one as their search interest.Recently, methods have been proposed for multiple keyword search in cloud computing we perform multi-keyword search over encrypted data in clouds leveraging polynomial functions.Specifically, we exploit the number of query keywords appearing in the document index to evaluate the similarity between the query and the document[9].

III. SOFTWARE AND HARDWARE REQUIRMENT SPECIFICATION

SOFTWARE :

Operating System : Windows

Technology : Java

Web Technology :HTML,Javascript

Database : My SQL

Java Version : JDK1.7

IDE : Netbeans 8.0

HARDWARE :

System : i3 processor

Hard Disk : 1TB

Ram : 4GB

Mouse : Optical Mouse

Keyboard :Standard 104 keys

Monitor : Dell colour

IV. MATHEMATICAL MODEL AND DESIGN

Input: Query from data user.

Output: Document matches with the query output.

D= Set of documents.

put set D which contains all documents stored in cloud server i.e.

D= d1,d2,d3...dn

Then put set of queries i.e.

Q= Set of queries.

Q= q1,q2,q3,...qn

After that set of keys are there i.e.

K= Set of keys.

K= k1,k2,k3,...kn

Then,

E(D)= Encrypted Documents.

Now take set of encrypted documents i.e.

E(D)= E(d1),E(d2),...E(dn)

When data user sends query for data to cloud server,then data owner check for valid user login. If user is valid and query i.e. q matches with document i.e.d then data owner provide dynamic key to the valid user and encrypted document i.e. E(d) is get as output to the user.

Hence,

E(D)= Q+k

V. PROPOSED SYSTEM

The three entities of a system model consists as shown in fig.1 the first containt is data user,second is the data owner and most important entity is the cloud server.The documents are collected by the data owner and which are responsible for the collecting documents from the user and server. For the accessing entire data the data user needs to get authorization from data owner before the accessing data from server.

A large space or storage area provides to the cloud server and ciphertext search required for the computation resources.By receiving the legal request or legal query from the data user the large storage cloud server which is most important entity in the system searches the encrypted index and instant send relevant documents which matches with user query.The main purpose of our whole system is to protect all documents and data sharing while improvin the system efficiency of ciphertext search in the system.

A. System flow process

Fig. 1. Module 1

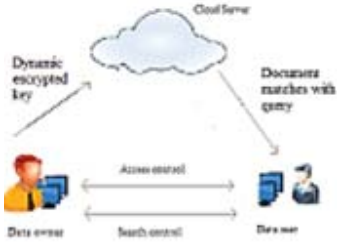


Fig. 2. Module 2

In this system,RSA(Ron Rivest Adi Shamir Adleman) algorithm is used.RSA is the key encryption algorithm.The data user must be registered in cloud server and the data owner having registration for every data user in the system.Every data user used document or information from the cloud server without any data lose.

Every time when the data user try to request for any document or information from the cloud server,data owner check for the valid data user into the cloud server.In this system the data owner send the data to the cloud server.The data owner uses the RSA Algorithm to encrypt the requesting data from user by using public key.When user request for any data from cloud server the data owner dynamically generate private key to decrypt the data.By using this system user gets its requested data without any loss of data.

This system is most useful or helpful for searching and sharing the data within data owner, cloud server and data user.Maximum security can be provide to the data by this method.Data can be confidentially store or share by this technique. This technique is more reliable to the data user for searching any query from the cloud server by getting encrypted key from the data owner.Because of key generation technique i.e. encryption, unauthorized user cannot access or get the data from cloud server.

VI. SYSTEM ARCHITECTURE

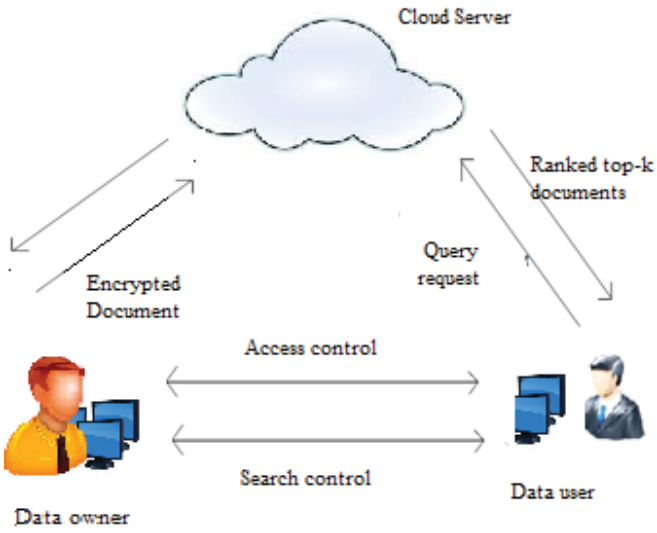


Fig. 3. System Architecture

System model consists of three entities i.e. data user, data owner and the cloud server. i] Data owner : It can collect documents, outsource them into encrypted format to cloud server. ii] Data user: It cannot access document from cloud Server without authorization from data owner.

Data user should authenticate from data owner. Then data user can send request for encrypted document to cloud server. iii] Cloud Server:If authenticated data user is requesting for document, then cloud server searches requested document in dataset and sends top k-document which matches with query. This helps to protect the information leakage.

VII. ADVANTAGES

i] Maximum security provide to thesystem.

ii] Due to encryption technique data stored in cloud become more secure and reliable.

iii] Unauthorized user cannot get data due to the key generation technique i.e encryption

iv] Data remains more confidential.

v] There is minimum data loss.

vi] Key generated by dynamically.

vii] Data storage capacity is more due to cloud server.

VIII. APPLICATIONS

i] Useful for education system.

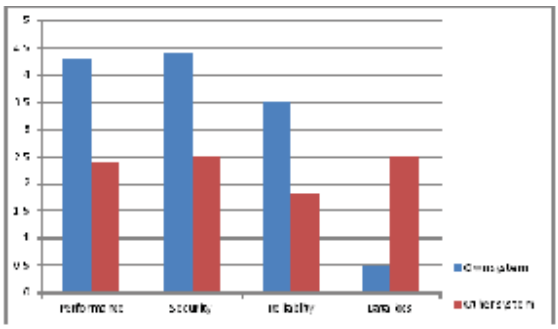
ii] For hospital management system.

iii] For governmental purpose.

iv] Useful for military system.

v] For shopping malls.

IX. EXPERIMENTAL EVALUATION



Due to using RSA encryption algorithm performance of system should increased.It gives best performance results.Security of system is increased due to dynamic encryption key generation by data owner to every user.Key generated dynamically,so that another user cannot use same key which is used previously and also same user can get another key even request for same document,so security is highly maintained than another system.

Reliability is more than another system because it become more trusted and authenticated due to encryption technique.Data must be provide to validate user by checking validation of user.

Data loss is minimum in this system.Cloud server is used for storing data,so that leakage of information is not possible.

X. SYSTEM ANALYSIS

TABLE I. TABLE NAME (PROPOSED SYSTEM VS.EXISTING SYSTEM)

System	Performance	Security	Reliability	Data loss
Proposed System	4.25	4.75	3.50	0.5
Existing System	2.25	2.50	2.75	2.50

XI. CONCLUSION

Some of system are poor in the privacy preserving constraints.To improve more privacy level here privacy preserving based RSA key encryption algorithm is used. Privacy of system is enhanced due to RSA key encryption algorithm.Dynamic key generation can improves security of the system.Semantic search can enhance the accuracy of the searching documents.Data loss is minimized due to cloud server storage.

REFERENCES

[1] C. Chen, X. J. Zhu, P. S. Shen, and J. K. Hu, A hierarchical clustering method For big data oriented ciphertext search, in Proc. IEEE INFOCOM, Workshop on Security and Privacy.

[2] D. X. D. Song, D. Wagner, and A. Perrig, Practical techniques for searches on encrypted data, in Proc. IEEE Symp. Security Priv., BERKELEY, CA, 2000, pp.4455.

[3] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, Publi key encryption with keyword search, in Proc. EUROCRYPT, Interlaken, SWITZERLAND,2004, pp. 506522.

[4] A. Swaminathan, Y. Mao, G. M. Su, H. Gou, A. Varna, S. He,M. Wu, and D. Oard, Condentiality-preserving rank-ordered search, in Proc. ACM ACM Workshop Storage Security Survivability, Alexandria, VA, 2007, pp.7-12.

[5] C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, Secure ranked keyword search over encrypted cloud data, in Proc. IEEE 30th Int.Conf. Distrib. Comput. Syst., Genova, ITALY, 2010, pp. 2532.

[6] C. Wang, N. Cao, K. Ren, and W. J. Lou, Enabling secure and efficient ranked keyword search over outsourced cloud data, IEEE cTrans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 14671479,Aug.2012.

[7] Pang, J. Shen, and R. Krishnan, Privacy-preserving similarity-based text retrieval, ACM Trans. Internet Technol., vol. 10, no. 1,pp. 39, Feb. 2010.

[8] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li,Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking, in Proc. 8th ACM SIGSAC Symp. Inform., Comput. Commun. Security, Hangzhou, China, 2013,pp. 7182.

[9] Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data, IEEE Transactions on parallel and distribute system,vol. 27, NO. 2, february 2016.

[10] Wei Zhang, Student Member, IEEE, Yaping Lin,Member, IEEE, Sheng Xiao, Member,Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing,journal of latex class files, VOL. 6, NO. 1, January 2015.

11]Cao N, Wang C, Li M, Ren K, Lou W. Privacy-preserving multi-keyword ranked search over encrypted cloud data. IEEE Trans Parallel Distrib Syst 2014;25 (1):22233.