MULTI KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA

S. Manjula devi[1] and P.Sridevi[2]

[1]Research Scholar, Department of Computer Science, Vellalar College for Women Tamilnadu, India

[2]Assistent Professor, Department of Computer Science, Vellalar College for Women Tamilnadu, India

ABSTRACT

Cloud computing environment provides on-demand access to shared resources that can be managed with minimal interaction of cloud service provider and validated service to the user. Cloud storage can be either public or private. Data in the public storage can be viewed by all cloud users. The private data can be viewed by the authorized user only. This paper enhance the security of the cloud data using Advanced Encryption Standard (AES) encryption algorithm. Data owners are motivated to outsource their data in cloud servers for great convenience. Private data should be encrypted before outsourcing by using keys. Encryption is an important concept in cloud computing to maintain the database. Existing system maintained the database by providing password for files and documents. The proposed system provides keys to access the file and keys are maintained as private and keys are provided by the data owner. The paper focused ostrovsky scheme (private information retrieval) that allows a user to retrieve file without any information leakage. Experimental result are presented to test the security of AES algorithm and information leakage.

Keywords: Cloud computing, Encryption, Public key, Private key, AES, Data leakage.

I. INTRODUCTION

Cloud computing use resources like software and hardware that are deliver service over the network like internet. The name comes from the use of a cloud-shaped symbol as a thought for the complex infrastructure it contains in system diagrams. Cloud computing assign remote service with user’s data, computation and software. The cloud provider manages platforms and infrastructure on which the applications run. Cloud users can access cloud-based applications through a web browser or mobile application. The user data are stored on servers at a remote location. Cloud computing allows institution to get their applications and running faster with enhanced manageability and less maintenance. It enables IT professional to adjust resources to meet fluctuating and unpredictable business demand.Cloud security architecture recognizes issues with security control and management. Security control help to recover weakness in the system and reduce the effect of an attack. There are some techniques are described below to enhance security from data loss.

A. ENCRYPTION

The plaintext taken to encryption and encodes into unreadable form at using mathematical algorithms, effectively rendering data in unreadable form unless a cryptographic key is applied to convert it. Encryption gives data security and integrity.The encryption techniques are of two categories, they are Symmetric and

Asymmetric encryption techniques.Symmetric encryption is the simplest encryption where the same key is used for encryption and decryption. Insymmetric encryption sender and receiver use a shared key to encrypt or decrypt the data. The only problem withthis technique is that if the key is known to others the entire system is collapsed. In Asymmetric encryptiontechnique both sender and receiver use a separate key to encrypt and decrypt the data. Asymmetric encryptionuses two keys to encrypt a plain text. One of the key is known as the private key and the other is known as the public key. The private key is kept secret by the owner and the public key is either shared among authorizedrecipients or made available to the public at large [8].

B. DATA LEAKAGE

Data leakage is an unauthorized transfer of information from inside an organization to an external destination or recipient that may be electronic or physical. Unauthorized does not automatically mean the data leakage by the user was intentional or malicious. The data leakage includes data loss or destruction of information due to hardware failure or destruction. Recent incidents have shown that data leakage caused by internal users is least as much threat caused by external attacks. The data leakage by internal user has more potential to cause greater financial losses than external attacks. Reduce or mitigating data leakage incidents for financial and business reasons, organizations may be obliged to adhere to various regulatory requirements enforcing the prevention of data leakage.

The data loss prevention (DLP) schemes that help end users, they do not send sensitive or critical information outside the corporate network. The Critical Security Controls data loss prevention (DLP) refers such as covering people, processes, and systems that identify, monitor, and protect data in use, data in motion, and data at rest through deep content inspection and with a centralized management framework. Data governance, risk assessment, regulatory and privacy compliance, data or information classification, policies, standards, procedures, data discovery, remediation processes and training and awareness are all elements are implemented by organization to effective implement of data leakage. The ultimate goal to prevent data leakage is to stop sensitive information from leaving to unauthorized organization. Email, instant messaging, social media, file transfer, web pages, mobile storage devices and hard copies are common vector for data leakage. To effectively prevent the data from leaving the organization to enforce their policies, it needs to deploy DLP solutions. DLP solutions installed on a user workstation can for block users from transferring files out of the control of the organization.

The rest of this paper is organized as follows. Section II explains the related researches briefly. Section III provides the details of entire system architecture is carried out. Section IV provides the experimental results and their discussions. Section V concludes the research work.

II. RELATED WORK

Chi Chen et al., [2015] [1] discussed about hierarchical clustering method is proposed to support more search semantics and also to meet the demand for fast cipher-text search within a big data environment. The proposed hierarchical approach clusters the documents based on the minimum relevance threshold, and then partitions the resulting clusters into sub-clusters until the constraint on the maximum size of cluster is reached. The experiment result proves that the proposed architecture not only properly solves the multi-keyword ranked search problem, but also brings an improvement in search efficiency, rank security, and the relevance between retrieved documents.

Z.Xia et al., [2015] [9] constructs a special keyword balanced binary tree as the index, and propose a "Greedy Depth-first Search" algorithm to obtain better efficiency than linear search. In addition, the parallel search process can be carried out to further reduce the time cost. Experimental results demonstrate the efficiency of our proposed scheme. In the proposed scheme, the data owner is responsible for generate, update information and send to cloud server. Thus, the data owner needs to store the unencrypted index tree and the information that are necessary to recalculate the IDF values. Such an active data owner may not be very suitable for the cloud computing model. Most of works about searchable encryption, our scheme considers the challenge from the cloud server. Actually, there are many secure challenges in a multi-user scheme.

Zhihua Xia et al., [2015] [10] presented a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Construct a special tree-based index structure and propose a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search. The kNN algorithm is expended the use to encrypt the index and query vectors, and relevance score calculation among encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. The use of tree based index structure proposed which supports not only the multi-keyword ranked search but also dynamic deletion and insertion of documents.

Ning Cao et al., [2014][3] focus the challenging problem of privacy preserving multi-keyword ranked search over encrypted cloud data (MRSE).We establish a set of strict privacy requirements for such a secure cloud data utilization system. To protect data privacy and combat unsolicited accesses in the cloud and beyond, sensitive data, e.g., emails, personal health records, photo albums, tax documents, financial transactions, etc., are encrypted by data owners before outsourcing to the commercial public cloud. The trivial solution of downloading all the data and decrypting locally is clearly impractical, due to the huge amount of bandwidth cost in cloud scale systems.

Ruixuan Li Zhiyong Xu et al., [2014] [6] said that the cloud computing infrastructure is a promising new technology and greatly accelerates the development of large scale data storage, processing and distribution. The security and privacy become major concerns when data owners outsource their private data onto public cloud servers that are not within their trusted management domains. For the query matching result which contains a large number of documents, the out-of-order ranking problem may occur. It makes it hard for the data consumer to discover the subset of satisfying requirements. MKQE greatly reduces the maintenance overhead during the keyword dictionary expansion. It takes keyword weights and user access history into consideration when generating the query result. The documents have higher access frequencies and that match closer to the user's access history and get higher rankings of matching result set.

III. SYSTEM ARCHITECTURE

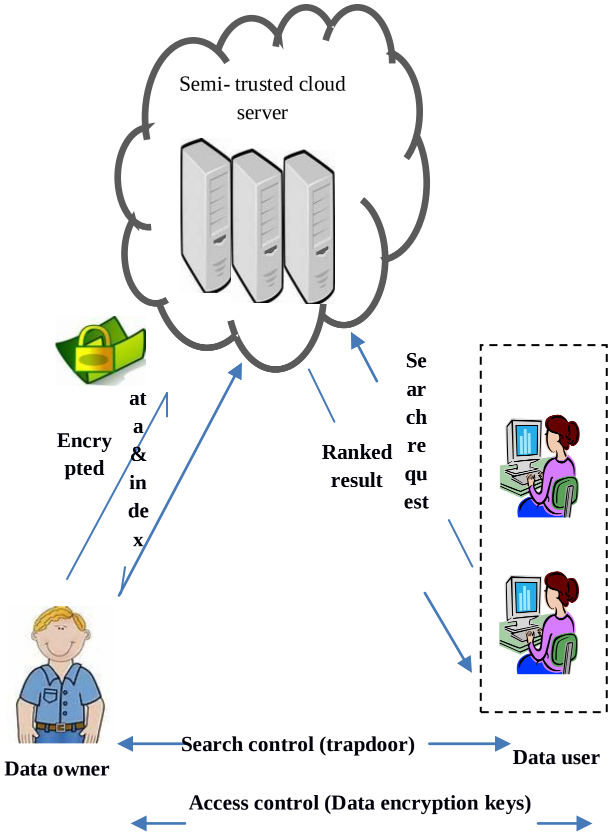
Semi- trusted cloud server
Encry pted
at a & in de x
Ranked result
Se ar ch re qu est
Data owner
Search control (trapdoor)
Data user
Access control (Data encryption keys)

Fig 3.1 System Architecture

The data owner has collection of documents and outsources the data to the cloud server in encrypted form. In proposed scheme, data owners build the search tree index for document collection and encrypt the private document. The data owner securely sends the encrypted document to the cloud server, and then securely distributes the key to the trapdoor generation and then decrypt the data by authorized data users.

The data users are authorized one to access the data owner document. The query keywords, authorized user can generate from trapdoor according to search control (SC) mechanisms and fetch encrypted document from cloud server. The data user can decrypt the data with the secret key send by data owner.

The cloud server stores the collection of encrypt document and search tree index for data owner. Receive trapdoor from the data user, the cloud server execute search over the index tree and return the collection of top-k ranked encrypted document. The cloud server is employ by lot of works on secure cloud data search. The cloud server executes the query based on instruction of the designed protocol. Meanwhile, cloud server analysis received data, which help to acquire additional information.

A. AES ALGORITHM

The**Advanced Encryption Standard** (**AES**), also known by its original name **Rijndael** algorithm. AES is a block cipher intended to replace DES for commercial applications. It uses a 128-bit block size and a key size of 128, 192, or 256 bits. Compared to public key cipher such as RSA, the structure of AES, and most symmetric ciphers, is very complained easily as RSA and similar algorithms.There are two main way to do encryption. First kind symmetric encryption has only public key (shared secret key). The recipients need the shared secret key to unlock the data. Asymmetric encryption splits the key into two keys. One key is made for public and one key is kept as private. The recipient need private key to decrypt the message. AES encryption is the process of converting the plain text into a format which is not easily readable and is called as cipher text. Series of mathematical operations are applied iteratively to get cipertext. The decryption of the data which is done by inverting all the encryption operations with the same key under AES symmetric encryption standard. In the decryption process the sequence of the transformations differs from that of the encryption but the key expansion for encryption and decryption are same.

B. PRIVATE INFORMATION RETRIEVAL

A private information retrieval (PIR) is also called as ostrovsky scheme, it allows a user to retrieve an item from a server database without revealing which item is retrieved. One trivial, but very inefficient way to achieve PIR is for the server to send an entire copy of the database to the user. PIR can be achieved with constant communication and k-database. Ostrovsky scheme allows a user to retrieve files of interest from an untrusted server without leaking any information.

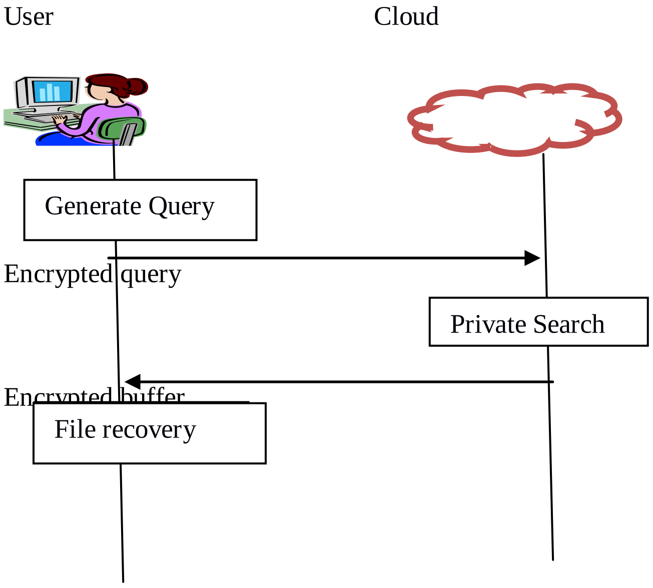

Fig 3.2 Working Process of Ostrovsky Scheme

C. ASSOCIATION RULE

Association rule mining is primarily focused on finding frequent co-occurring associations among a collection of items. It is sometimes referred to as "Market Basket Analysis", since that was the original application area of association mining. The goal is to find associations of items that occur together more often than you would expect from a random sampling of all possibilities. Association rules analysis is a technique to uncover how items are associated to each other. Association rules are if/then statements that help to uncover relationships between unrelated data in a database, relational database or other information repository. Association rules are used to find the relationships between the objects which are frequently used together. Applications of association rules are basket data analysis, classification, cross-marketing, clustering and loss-leader analysis etc.

Effective Evaluation of Query Algorithm

```
float score[N]=0
 for each d
 do initialize length[d] to the length doc d
 for each query term t
do calculate $w_{t,q}$and fetch posting list for t
                for each pair(d,$tf_{t,d}$)in posting list
do add wft,d to scores[d]
read the array length[d]
for each d
do divide scores[d]by length[d]
Return top k components of score[]
```

IV. RESULT S AND DISCUSSION

A. ENCRYPTION PERFORMANCE

The more popular and widely adopted symmetric encryption algorithm used nowadays is the Advanced Encryption Standard (AES). It is found six times faster than triple DES.A replacement for DES was needed as its key size was too small. With increasing computing power, it was considered vulnerable against exhaustive key search attack.

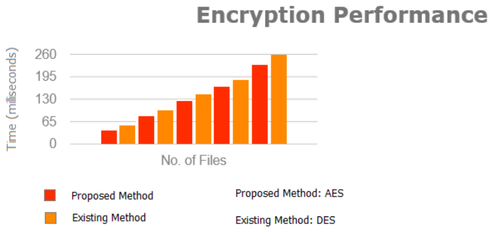

Fig 4.1 Encryption performance

File Count	Existing System DES	Proposed system AES
10 Files	42 Milli Seconds	31 Milli Seconds
20 Files	65 Milli Seconds	55 Milli Seconds
30 Files	90 Milli Seconds	70 Milli Seconds
40 Files	124 Milli Seconds	112 Milli Seconds

Table 4.1 Encryption performance

The Table 4.1 and Figure 4.1 shows that the period of time taken for encryption. The AES algorithm takes minimum time for encryption and decryption. It helps the data owner to send the data securely. The encryption time is reduced when compared with the existing system.

B. DATA LEAKAGE

Data leakage is defined as the accidental or intentional distribution of private or sensitive data to an unauthorized user. Ostrovosky scheme is used to detect data leakage in the proposed system. If the user can download the file without getting permission from data owner. The proposed system can detect the unauthorized accessing using Ostrovosky scheme.In the existing system no method is used to find data leakage. The proposed system used dynamic secret key generation protocol and data user authentication protocol to avoid Data leakage.

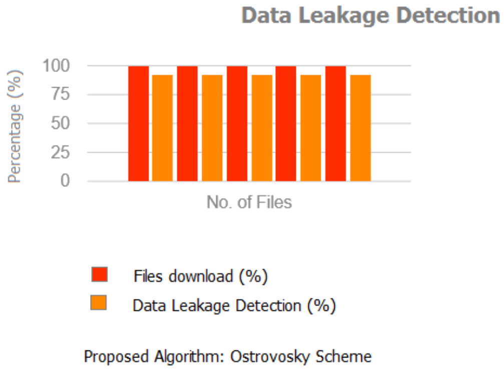

Fig 4.2 Data leakage detection

File Count	**Data Download (BDRMS Searching)**	**Leakage Detection Level (Ostrovsky scheme)**
10 Files	100 %	92 %
20 Files	100 %	92 %
30 Files	100 %	92 %
40 Files	100 %	92 %
50 Files	100 %	92 %

Table 4.2 Data Leakge Detection

TheFigure 4.2 and Table 4.2 shows the performance of ostrovsky scheme. The information about data leakage that can be only viewed by data owner .It can detect the data leakage at minimum level up to 100 files it shows the best performance.

C. RANKING SEARCH

In the computer field, most user are in need of cloud data accessory. The data owners are motivated to outsource their data from local sites to the commercial public cloud for sharing the data with great flexibility. The existing system used apriori algorithm for ranking. The proposed method used association rule for ranking. Association rule mining is primarily focused on finding frequent co-occurring associations among a collection of items. Association rules analysis is a technique to uncover how items are associated to each other. Association rules are used to find the relationships between the objects which are frequently used together.

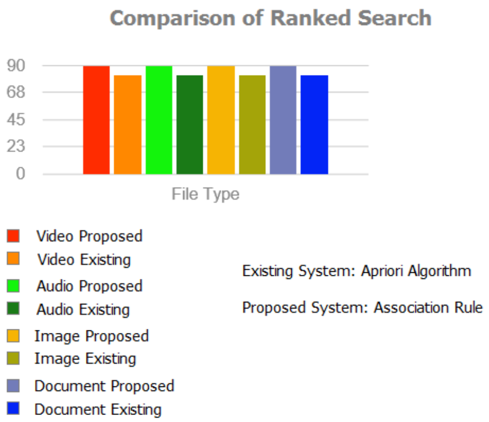

Fig 4.3 Ranking Search

File Type	**Existing system(Apriori scheme)**	**Proposed system(Association rule)**
Video File	82 %	90 %
Audio File	82 %	90 %
Image File	82 %	90 %
Document File	82 %	90 %

Table 4.3 Ranking search

The Table 4.3 and Figure 4.3 shows the ranking performance of association rule when searching, Association rule allows a user to retrieve an item from a server database as many as possible. This comparison shows the safe search over un-trusted server. The charts provide the difference between the apriori and association rule ranking.

D. SEARCHING TIME

The semantic search more smart in multi-keyword ranked search over cloud searching time of the document considerably reduced. It reduces searching timing and gives more accurate (semantic) output than existing Method. Association Rule is used to match the keyword as much as possible search. Existing method has taken 2000 ms(Milli Seconds) to search 10 MB file. But the proposed method takes 1380 milliseconds to search 10 MB file.

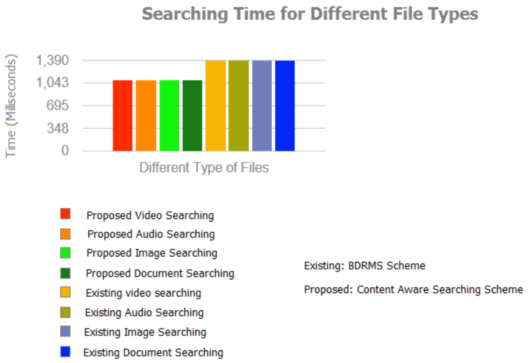

Fig 4.4 Searching Time

File Type (each file size 10MB)	Time taken by Existing system (BDRMS Scheme)	Time taken by Proposed system (Content aware searching scheme)
Video File	1390 ms	1080 ms
Audio File	1390 ms	1080 ms
Image File	1390 ms	1080 ms
Document file	1390 ms	1080 ms

Table 4.4 Searching Time

The above Figure 4.4and Table 4.4 shows the efficient searching of the proposed search. In the proposed system Content aware searching scheme index tree is largely reduced by defining the key terms while uploading the file.

E. STORAGE CONSUMPTION

In Cloud Computing, space allocation normally takes actual size of the uploading file. To reduce the file size it can compressed before uploading files. The Following table and chart shows cloud Storage consumption. The consumption of file size is reduced.

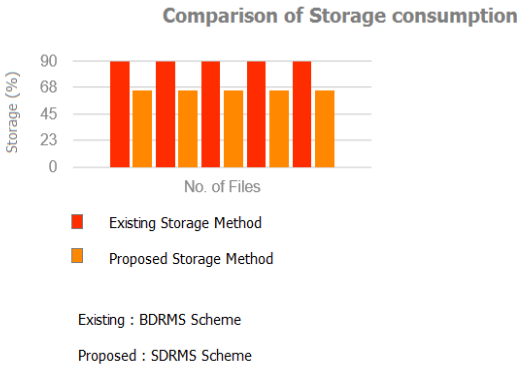

Fig 4.5 Storage Consumption

File Count	Time taken by Existing system(BDRMS Scheme)	Time taken by Proposed system (SDRMS scheme)
10 Files	90 %	65 %
20 Files	90 %	65%
30 Files	90 %	65 %
40 Files	90 %	65 %
50Files	90%	65%

Table 4.5 Storage consumption

The Table 4.5 and Figure 4.5 shows the storage consuming of Smart Dynamic Ranked Multi Keyword Search(SDRMS) scheme when storing the data. The SDMRS scheme helps the data owner provide the keyword related document which they are uploaded it. The storage consumption of the document is considerably reduced.

5. CONCLUSION

This proposed research work discussed about data leakage, encryption and ranked search. Existing system used encryption and ranked search. The proposed system enhanced the encryption by using AES encryption algorithm and improved the ranking with association rule. The proposed method introduced ostrovsky scheme to detect data leakage in cloud environment. The proposed schemes enable authenticated data users to achieve secure, convenient, and efficient searches over multiple data owners and detect attackers who steal the secret key and perform illegal hacking. In future work relationships among query keywords have to be considered to enhance the system and introduce the keyword weight to the search protocol design.

REFERENCE

[1] Chi Chen, "An Efficient Privacy-Preserving Ranked Keyword Search Method" IEEE Transactions on Parallel and Distributed Systems, TPDS.2015.2425407, 2015.

[2] Ming Liu and Rafael A. Calvo, "Secure ranked keyword search over encrypted cloud data," in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst. (ICDCS), Jun. 2012, pp. 253–262, 2012.

[3] Ning Cao, Cong Wang[2014], "Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data" in Proc. Of EDBT, 2014, pp. 287–298, 2014.

[4] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing,"IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5, pp. 847–859, 2009.

[5] Reza Curtmola[2006], "Efficient multi-keyword ranked query over encrypted data in cloud computing," Future Generat. Comput. Syst., vol. 30, pp. 179–190, 2006.

[6] RuixuanLi Zhiyong Xu [2014] "Efficient multi-keyword ranked query over encrypted data in cloud computing" IEEE Trans.Parallel Distrib. Syst., vol. 25, no. 1, pp. 222–233, Jan. 2014.

[7] R.V.V Murali Krishna and Ch. Satyananda Reddy, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proc. 8th ACM SIGSAC Symp. Inf., Comput. Commun. Secur., 2012, pp. 71–82.

[8] P.Sridevi et al [2017] "An Efficient Encryption-Then-Compression System using Asymmetric Numeral Metho", International Journal of Engineering and Technology (IJET), Vol 9 No 5 Oct-Nov 2017, ISSN (Print) : 2319-8613,ISSN (Online) : 0975-4024.

[9] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A secure and dynamic multikeywordranked search scheme over encrypted cloud data," IEEE Trans.Parallel Distrib. Syst., vol. 27, no. 2, pp. 340–352, 2015.

[10] Zhihua Xia,"A Secure and Dynamic Multi-keyword RankedSearch Scheme over Encrypted Cloud Data", Ieee Transactions On Parallel And Distributed Systems Vol: Pp No: 99 Year 2015.