ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 3, March 2014

An Efficient and Trusted Data Storage Process for Cloud Computing

Kalyan Singh Meena

M.Tech., Department of Information Technology, Indian Institute of Information Technology , Allahabad, India

Abstract: In recent years due to attractive features of cloud computing, users are storing large amount of data on cloud storage, these data may be users personal or secret. Users remotely store their personal or secret information on cloud storage and enjoy best features of cloud applications without burden on local hardware and software management. After storing data on cloud storage users can access their data by very thin clients. But there is a drawback of outsourcing data on cloud storage is data security risks because users has no control over outsourced data even users are unknown about location of their data. So, in this situation, achieving data confidentiality on outsourced data is an open challenge for researchers. Today some systems are available by using that we can store and retrieve encrypted data but if we have large amount of data to store on cloud than it is difficult to retrieve them efficiently. Another problem is to store and manage encryption keys efficiently. To solve the data confidentiality and key management problems together we are using ranking function on inverted indexes and advanced shamir's secret sharing algorithm.

Keywords: Cloud Computing, Data Encryption, Symmetric Cryptographic Key Management, Cloud Storage.

I. INTRODUCTION

Before five years, the cloud computing was only grumbled about with doubt or debate. Today, technology experts forecast that cloud computing will become the model for accessing back-end applications and collecting information within the next year. Cloud computing allude to applications and services that run on distributed network using virtualized resources and accessed by common internet protocols and networking standards[1]. The main reason of popularity cloud computing is its various advantages like on demand self services, broad network access, resource pooling, rapid elasticity etc.

But till now some organization especially large enterprises they have sensitive data, are not moving data to cloud storage because they do not trust on cloud storage. The main reason for not trusting on cloud computing is security because in cloud computing users do not know about location of data; users do not know how much data are secured. In 2009 gartner group has completed a survey on using cloud computing and find out result, 70% of IT managers believes that security and privacy problem is a biggest problem of cloud computing. In previous years some incidents happened that verify people fears: Media Max went out of business in 2008 after losing 45% of stored client because of fault from system administrator; in 2007, customers emails was accessed by criminals[2] from leading cloud service provider salceforce.com . Today the main challenge for cloud storage service provider is to be sure about confidentiality of all important data. For achieving data confidentiality, sensitive data is usually move to cloud storage in encrypted form which cannot accessed by unauthorized user. In this field lot of work has done. But using this approach is hard because of cryptographic key management problems. On cloud storage, we have sensitive data in encrypted form and location of data could be anywhere, and we need the ability for any application that can access to this data to be able to decrypt it. To decrypting data, we need a way for any application to be able to get the keys that it needs to decrypt data that it gets from the cloud and it must be the responsibility of cloud service provider to ensure the management of cryptographic key so these keys can be used for decryption of data appropriately. In this paper we proposed an architecture by which encrypted data can be store and retrieve successfully and symmetric cryptographic keys can also be managed efficiently, in this way this is a secure, efficient and trusted process for storing and retrieving data.

II. LITERATURE REVIEW

First time on searchable encryption Dawn Xiaodong Song ,David Wagner and Adrian Perrig in 2000[3] proposed a method for searching over encrypted data. This method has various advantages like provable secure in the sense that untrusted server cannot learn anything about the plaintext given only the cipher text. This method provides controlled searching so that the untrusted server cannot search for anything without user's authorization. But this scheme is not suitable for cloud storage services because cloud storage services has large amount of data. On large amount of data ranked based document retrieval is necessary but in this scheme it is not possible to retrieve documents in ranking order. In 2009 Quin Liu et al.[4] proposed a scheme, in this scheme it is not required for user to decrypt all cipher text, user can decrypt a small amount of cipher. In this way there will be fewer loads on server. In this manner this scheme is very useful for protecting user data privacy with efficiently. The main disadvantage of this scheme is that it is suitable for single user system only. Another drawback of this scheme is that searching results is not come in ranking order, so this scheme is not suitable for cloud storage applications.
In 2012 Cong Wang et al.[5] proposed a scheme, By using this scheme we can improve system utility because search

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 3, March 2014

result will come in ranking order instead of sending undifferentiated results. But key management facility is not available in this method. Secret Sharing schemes are one of the widely used methods for key management. In 1979 shamir[6] proposed a method for secret sharing, in this algorithm a data D to be outsourced is split into n parts, such that each data D_i, $i \leq n$, is padded with redundant information to make its size same as that of D. The Data D can be retrieved if k out of n pieces is available. Shamir calls this as threshold (k, n). Shamir's secret sharing method is known for its various advantages like secure, dynamic, extensible and flexible, but when k or more servers collude then security is lost. Another method is proposed by, Rabin [7] to splitting a data D of length $L=|D|$ into n parts such that a person can obtain the data only if $k < n$ of these parts are available, where k is the threshold. Here, each part D_i, $i \leq n$, is of size $|D|/k$, where $|D|$ is the size of the data. The total sizes of all the secrets are $(n/k)*|D|$. But, the security problem in this method is that, if the data shows some repeating patterns, and that the attacker gets hold of $m < k$
Pieces, then there are lot of possibilities for him to get the data D and also unable to work on many environments[8].

III. EXISTING SYSTEM

Various methods have been proposed for searching over encrypted cloud data, by using these schemes user can store encrypted data on cloud storage server and retrieve the same. For encrypting large amount of files user requires various cryptographic keys, but traditional schemes are unable to manage these cryptographic keys. For key management shamir's secret sharing algorithm is helpful but when k or more servers collude then security is lost. We proposed a scheme that will manage symmetric cryptographic keys on cloud based platform. Proposed scheme is based on secret splitting. Cryptographic key management systems include operations like store, generation, distribution, retrieval.

IV. PROPOSED METHOD

In this paper we are targeting to achieve data confidentiality on outsourced cloud data, using such approaches for achieving confidentiality that user can access their data by thin client. In our proposed architecture following methodologies are used:

1) Inverted Index: inverted index also called inverted file is a data structure which is used for documents retrieval systems. Inverted index is created by using different documents and removed stop words. Ranking function is used on inverted index for relevance score on corresponding documents for searching on encrypted data. The advantage of Inverted index is allowing fast full text search.

Ranking Function: in Information retrieval, ranking function is used to calculating relevance score corressponding to given inverted index . For computnig relevance score by ranking function we are using TF*IDF here TF= term frequency, denotes the number of times a given keyword appears in a given document. IDF= inverse document frequency,IDF is calculated by dividing the number of documents in the all collection by the number of documents which has keywords.Following function is used for computing relevance score:

$$Score(\tau, F_d) = \sum_{t \in Q} \frac{1}{|F_d|} \cdot (1 + \ln f_{d,k}) \cdot \ln\left(1 + \frac{W}{f_k}\right).$$

$f_{d,k}$ denotes the TF of keyword k in document F_d , f_k is the number of document that contain keyword k, W is the total number of files in the collection, $|F_d|$ is the length of document F_d.

2)Advanced Shamir's Secret Sharing: Before describing advanced shamir's sharing algorithm we are defining basic algorithm. This algorithm is used for key management and more secure than shamir's basic algorithm. Shamir's basic algorithm is works as following:
For example if data D is available then according to shamir's secret sharing algorithm we can split data D into n parts like D_1, D_2, D_3,……..D_n in such a method that:
By help of K or more D_i pieces makes D easily computable.
Information of any K-1 or some D_i pieces cannot compute D.
This scheme is called (K,N) threshold scheme. In this algorithm when k or more server collude then it create security problem.
In our proposed Advanced shamir's secret algorithm is to divide a key into parts D_1, D_2,D_3,.....D_n. and each part will be store on different server. For example, part D_1 will store on server p_1, part D_2, will store on server p_2 and so on. If K servers are sufficient to compute the key then we ensure that no more than more than K-1 servers are able to collude. By using this approach shamir's secret sharing scheme is considered cryptanalytically unbreakable. This is an effective, robust algorithm for symmetric cryptographic key management in cloud base environment. Figure 1 shows high-level architecture for efficient and robust data storage process in cloud computing. In this architecture symmetric cryptographic keys is managed. Values of inverted index are converted into hash values. Hash function is applies on keywords that is used by users for searching. In this architecture it is assumed that authorization between user U and user U' is already completed efficiently. In proposed method there are three important participants user, cloud storage server and data splitter server, these participants is defined in detail as following:

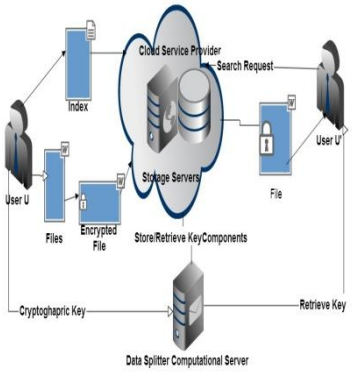

Fig.1 Proposed Architecture

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 3, March 2014

A.User

The user in this architecture is responsible for Create his/her profile, update profile, delete profile,Build inverted index from selected documents, Encryption and decryption of data,Update his documents,Search documents among huge amount of documents,Store and retrieve encryption keys.

B. Cloud Storage Server

Cloud storage is a model of storing data not only on remotely located servers but also virtualized pools of storage. Cloud hosting companies has large data centres and users they want to store their data on cloud storage, buy or lease storage capacity from them. Cloud storage server is responsible for store inverted index, store Encrypted values of document name, Encrypted values of document content,Splitted values of encryption keys for symmetric cryptographic key management purpose.

C. Data Splitter Server

Data splitter server is responsible for Split Encryption keys Into different parts,Store splitted keys into different cloud storage servers, After receving request from user to compute original encrypted key from splitted parts,Send original key to user.

Figure 2 shows the use case diagram for proposed architecture. This diagram represents user's interaction with the system and depicting the specifications of a use case. In this diagram user U and user U' are two different users and it is assumed that user U' is authorised for retrieving data from cloud storage.

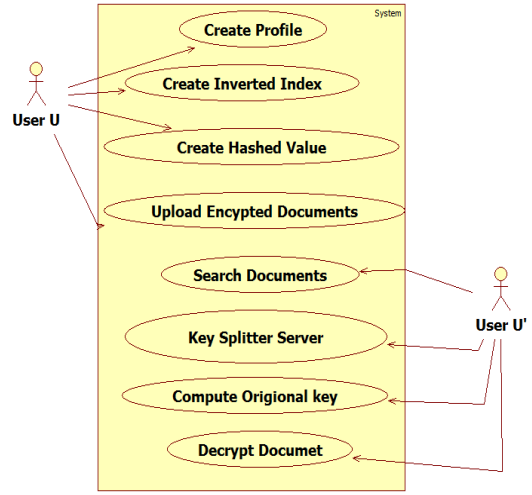

Fig 2 Use case diagam

V. CONCLUSION

In this paper after studying different research papers and research reports we have anylized that Today the biggest problem of cloud computing is security problem. In recent years various researchers tried to solve security problems on clud storage which has outlined systematically in this paper. Achieving data confidentiality on cloud storage is challengaing task for researchers, in this paper first time we tried to solve data confidetiality problem with key management. By using this scheme a user can store large amount of encrypted data on cloud storage and retrieve the same encrypted data efficiently. For retrieving encrypted data efficiently inverted index and ranking function is used. Advanced shamir's secret sharing algorithm is used For solving key management problem.

REFERENCES

[1] Cloud computing bible Wiley India Pvt Ltd (2011) Auther-Barrie Sosinsky,ISBN:978-81-265-2980-3
[2] Chen, Hong Zhao, "Data Security and Privacy Protection Issues in Cloud Computing", International Conference on Computer Science and Electronics Engineering-2012, pp-647-651.
[3] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in SP '00: Proceedings of the 2000 IEEE Symposium on Security and Privacy. Washington, DC, USA: IEEE Computer Society, 2000, pp.1- 44.
[4] Qin Liu, Guojun Wang Jie Wu. "An Efficient privacy Preserving Keyword Search Scheme in Cloud Computing" IEEE DOI10.1109/CSE.2009. International Conference on Computational Science and Engineering-2009, pp. 715-720.
[5] Cong Wang, Ning Cao, Kui Ren, , and Wenjing Lou." Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data", IEEE transaction on parallel and distributed systems, vol. 23, no. 8, august- 2012, pp-1467-1479.
[6] Shamir, A.: How to share a secret. In: Commun. ACM, vol. 22, no. 11, pp. 612–613 (1979)
[7] Rabin, M.O.: Efficient dispersal of information for security, load balancing, and fault tolerance. In:Journal of The ACM 36(2), pp. 335–348 (1989
[8] Resch, Jason; Plank, James (February 15, 2011). "AONT-RS: Blending Security and Performance in Dispersed Storage Systems". Usenix FAST'11, 2011

BIOGRAPHY

Kalyan Singh Meena received the M.Tech. degree in Information Technology from Indian Institute of Information Technology Allahabad , currently he is working as Assistant Manager in Uttar Bihar Gramin Bank. His research interests are cloud computing security and data mining.