Storage Algorithms in Cloud Computing: A Review

[1]Gagandeep Kaur, [2]Sonia Sharma
[1,2]Dept. of CSE, Guru Nanak Dev University, Amritsar, Punjab, India

Abstract
Cloud computing is a buzzword that means different things to different people. For some, it's just another way of describing "outsourcing" others use it to mean any computing service provided over the Internet or a similar network and some define it as any bought-in computer service you use that sites outside your firewall. Storing, accessing and processing of data in cloud environment are the cutting edge technology used all over the world. Storing and sharing is basic need for enhancing the security. Sensitive data must be more secured during the sharing of data. Data encryption, homo-morphic encryption, secret sharing technique and data partition technique are widely used to share the data in cloud computing environment. In this paper we discuss the two techniques: secret sharing algorithm and information dispersal algorithm.

Keywords
Cloud Computing, Storage, Secret Sharing Algorithm, Information Dispersal Algorithm

I. Introduction

Cloud computing is a technology which is mainly about resource sharing with the motive of high availability and scalability, equivalent to providing utilities over a network. Cloud computing as a whole is providing services and resources on demand that is on cloud. Cloud computing, or in other words "the Cloud", focuses mainly on effective resource sharing. Resources in cloud are not only accessed by multiple consumers rather dynamically reallocated based on its demand. This will improve resource allocation in cloud. In cloud computing, multiple consumers access one single server to retrieve data which gives them an advantage of not paying for many applications rather only to those which they consume. Cloud storage is based on cloud computing, is the extension of Distributed Computing. Cloud storage is somehow the same as cloud computing, they consolidate all storage devices by the features like cluster application, grid technology and distributed file system.

Security and privacy are the main concerns in cloud computing. The two capable algorithms such as secrete sharing technique and information dispersal algorithm are measured with respect to their performance are identified for secured data sharing and privacy of the data. These two algorithms split the data into pieces then encrypt. Few pieces are enough to get the actual data. Both algorithms use different techniques to split the data. Secret sharing uses polynomial function whereas information dispersal algorithm uses matrix. This led reduction of transmission overhead and space complexity.

This paper is composed into IV sections. Section I Introduction about cloud computing, Section II is about background work. Section III discussed in detail about secret sharing algorithm. Section 4 discussed about information dispersal algorithm. Section V is about comparison of both algorithms.

II. Background Work

Various approaches are proposed to share and store the data in the cloud environment based on type of storage, access mechanism, user privacy, key distribution etc.

Data encryption translates data into another form, or code, so that only people with access to a secret key (referred to as cipher text, while unencrypted data is called plaintext. Currently, encryption is one of the most popular and effective data security methods used by organizations. Data, or plaintext, is encrypted with an encryption algorithm and an encryption key. The process results in cipher text, which only can be viewed in its original form if it is decrypted with the correct key.

Homomorphic Encryption systems are used to perform operations on encrypted data without knowing the private key (without decryption), the client is the only holder of the secret key. When we decrypt the result of any operation, it is the same as if we had carried out the calculation on the raw data formally called a decryption key) or password can read it. Encrypted data is commonly. The homomorphic encryption is a computationally complex to perform operations [7].

Private information retrieval is one of the secure methods in the database operation. The main aim of this method is to provide security for the user operation. It hides all the user operations in the database. It provides privacy for the user operation of the service provider. A lot of private information retrieval technique has been developed on the basis of user privacy concern. Symmetric private information retrieval, this method provides high security to the user data. It is totally infeasible on single server computation.

The two capable algorithms such as secrete sharing technique and information dispersal algorithm are measured with respect to their performance are identified for secured data sharing and privacy of the data.

III. Secret Sharing Algorithms

Shamir's Secret Sharing [2] is an algorithm in cryptography created by Adi-Shamir. It is a form of secret sharing, where a secret is divided into parts, giving each participant its own unique part, where some of the parts or all of them are needed in order to reconstruct the secret.

Secret sharing (also called secret splitting) refers to methods for distributing a secret amongst a group of participants, each of whom is allocated a share of the secret. The secret can be reconstructed only when a sufficient number, of possibly different types, of shares are combined together; individual shares are of no use on their own. In one type of secret sharing scheme there is one dealer and n players. The dealer gives a share of the secret to the players, but only when specific conditions are fulfilled will the players be able to reconstruct the secret from their shares. The dealer accomplishes this by giving each player a share in such a way that any group of t (for threshold) or more players can together reconstruct the secret but no group of fewer than t players can. Such a system is called a (t, n)-threshold scheme .Counting on all participants to combine the secret might be impractical and therefore sometimes the threshold scheme is used where any of the parts are sufficient to reconstruct the original secret.

Secret-sharing schemes are a tool used in many cryptographic protocols. A secret-sharing scheme involves a dealer who has a

secret, a set of n parties, and a collection A of subsets of parties called the access structure. A secret-sharing scheme for A is a method by which the dealer distributes shares to the parties such that: (1) any subset in A can reconstruct the secret from its shares, and (2) any subset not in A cannot reveal any partial information on the secret. Originally motivated by the problem of secure information storage, secret-sharing schemes have found numerous other applications in cryptography and distributed computing, e.g., Byzantine agreement

Secret sharing schemes are ideal for storing information that is highly sensitive and highly important. Examples include: encryption keys, missile launch codes, and numbered bank accounts. Each of these pieces of information must be kept highly confidential, as their exposure could be disastrous however, it is also critical that they should not be lost.

Traditional methods for encryption are ill-suited for simultaneously achieving high levels of confidentiality and reliability. This is because when storing the encryption key, one must choose between keeping a single copy of the key in one location for maximum secrecy, or keeping multiple copies of the key in different locations for greater reliability. Increasing reliability of the key by storing multiple copies lowers confidentiality by creating additional attack vectors; there are more opportunities for a copy to fall into the wrong hands. Secret sharing schemes address this problem, and allow arbitrarily high levels of confidentiality and reliability to be achieved.

Secret sharing schemes are important in cloud computing environments. Thus a key can be distributed over many servers by a threshold secret sharing mechanism. The key is then reconstructed when needed. Secret sharing has also been suggested for sensor networks where the links are liable to be tapped by sending the data in shares which makes the task of the eavesdropper harder. The security in such environments can be made greater by continuous changing of the way the shares are constructed.

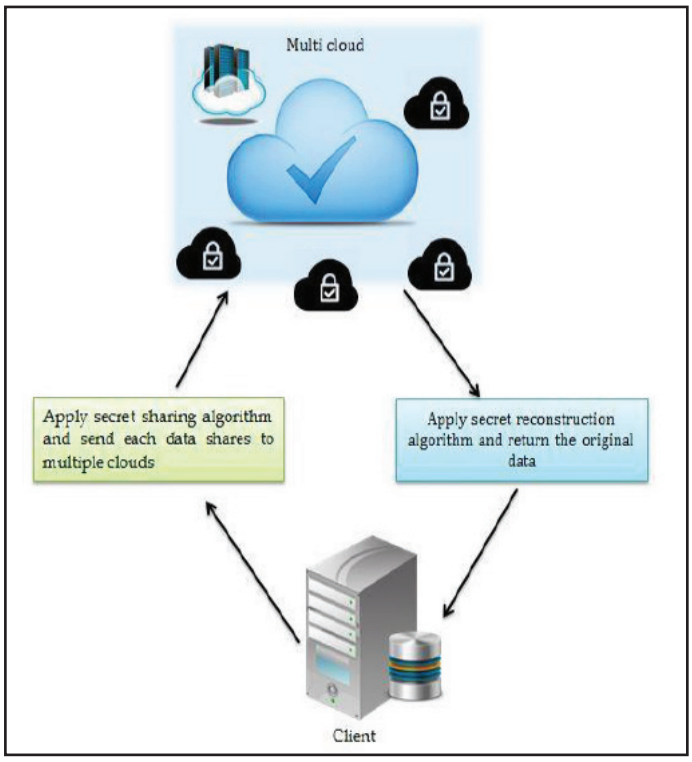

Fig. 1: Secret Sharing Algorithm

Example

Let S=1234

n=6 and k=3 and obtain random integers a1=166 and a2=94

$$f(x) = 1234 + 166x + 94x$$

Secret share points

(1,1494),(2,1942)(3,2598)(4,3402)(5,4414) (6,5614)

We give each participant a different single point (both x and f(x)).

IV. Information Dispersal Algorithms

Naor and Roth proposed an information dispersal algorithm [2] over arbitrary graphs. In their model, an arbitrary file f is distributed among the nodes of the graph in such a way that each node of the graph, by accessing the memory of its own and of its adjacent nodes, can reconstruct the contents of f. Their scheme can be applied to store files in distributed networks. The basic idea of his algorithms is to add to the information some amount of redundancy and then to partition it into n fragments, each transmitted to one of the parties.

Information dispersal algorithms are used to separate data packets into slices so that they are unrecognizable as they sit in storage arrays or traverse the network. Data can be reassembled at the receiving device The efficiency of any information dispersal algorithm, is computed regarding the size of pieces given to each participant.

The matrix is constructed to form the encrypted data:\

M(KxN/K)

M=[s1,s2,….sN/k]

The matrix A(nxm) is constructed using vandermonde matrix property.

A=[]

ai=(ai1,…aik)(1≤i≤n)

Information dispersal algorithms provide a methodology for storing information in pieces (dispersed) across multiple locations, so that redundancy protects the information in the event of a location outage, but unauthorized access at any single location does not provide usable information. Only the originator or a user with a list of the latest pointers with the original dispersal algorithm can properly assemble the complete information. This has been expanded to include peer-to-peer (P2P) file-sharing technologies and protocols, such as those based on the Bit Torrent Protocol, which has proved to be robust on the Internet.

In Cloud Storage, Information dispersal algorithms can be employed to split files into multiple data slices which will be redundantly stored on several storage nodes. IDAs can also enhance the stored data confidentiality. In a cloud storage system employing IDAs, a adversary who wants to read a file, has to compromise minimum slice stores or eavesdrops on slices. He also needs to determine which slices logically belong to the file. The adversary will also have to guess the transform matrix of the file and apply the matrix with the correct sequence of the slices. To achieve all these will be very difficult, in practical. The probability of success is very little. Therefore, the storage system

employing IDAs, rather than that just employing encryption, can more effectively guarantee the confidentiality of the data. To further enhance the security of IDAs-based Cloud Storage, the proxy server can optionally encrypt the file slices before sending them to external storage services.
Information dispersal algorithm has high computational complexity compared to the secret sharing algorithm.

Example:
|D|=32 m=4, n=8
D=b1,b2….b32
D=(b1…b4) (b5…b8) (b29…b32)

M(4×8)
M=(S1S2…S8) = [ ]
A=(8×4)

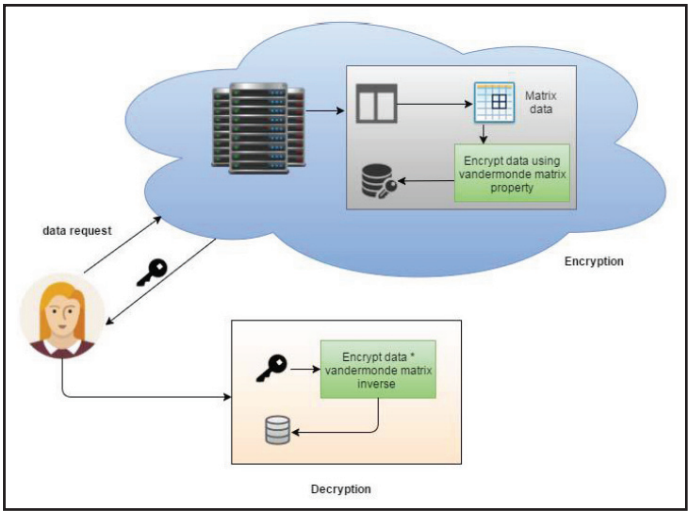

Fig. 2: Information Dispersal Algorithm

V. Comparison of Algorithms

The above two algorithms are differs from each other. The secret sharing algorithm follows the polynomial method for encryption and decryption whereas dispersal algorithm follows matrix method means use of matrix for encryption and decryption. Informational dispersal algorithm has greater computational complexity as compared to the secret sharing algorithm. As compared to traditional encryption techniques , these both algorithms are expensive. Secret sharing algorithm makes the data undetermined only few pieces are available to reconstruct the original data. Information dispersal algorithm produces pieces of data which is less than the original data. If any parts of data are corrupted remaining parts are used to reconstruct the original data. Secret sharing algorithm has fewer disadvantages compare to the information dispersal algorithm. The applications of secret sharing algorithms are voting system, business process whereas Information dispersal algorithm is used in company servers to maintain the huge data with low space complexity. Its usage is mainly where the storage space is main concern. It is vulnerable to attack. It will generate constant pattern so hackers can easily be exploited. By analyzing these two algorithms, new algorithm is proposed for secured data sharing in hybrid cloud environment i.e. modified secret sharing algorithm.

References

[1] C. Blundo, A. De Santis, L. Gargano, U. Vaccaro,"On the information rate of secret sharing schemes.
[2] E.F. Brickell, D.M. Davenport,"On the classification of ideal secret sharing schemes", J. Cryptology
[3] C. Blundo, A. De Santis, L. Gargano, U. Vaccaro,"On the information rate of secret sharing schemes".
[4] Mr.K.A.Muthukumar, Dr. M.Nandhini,"Modified Secret Sharing Algorithm for Secured Medical Data Sharing in Cloud Environment".
[5] [Online] Available: http://www.ibm.com/developerworks/cloud/library/cl-cloudstorage/cl-cloudstorage-pdf
[6] Maha TEBAA, Saïd EL HAJJI, Abdellatif EL GHAZI,"Homomorphic Encryption Applied to the Cloud Computing Security", Proceedings of the World Congress on Engineering, Vol. 1, July 2012.
[7] Purushothama. B., Amberker, B.,"Efficient Query Processing on Outsourced Encrypted Data in Cloud with Privacy Preservation", 2013.
[8] Tebaa, Maha, Saïd El Hajji, Abdellatif El Ghazi,"Homomorphic encryption applied to the cloud computing security", Proceedings of the World Congress on Engineering. Vol. 1. 2012.