Survey Paper For Enabling Dynamic Operations Of Outsourced Data On Cloud Storage System

Arya Govind S
Mtech, Computer Science
Nehru College of Engineering and Research Centre
Pampady, Thrissur, Kerala

Shinu Acca Mani
Mtech, Computer Science
Nehru College of Engineering and Research Centre
Pampady, Thrissur, Kerala

Abstract **--Cloud service providers or CSP are the services that provide storage as a service for the organizations to store their sensitive data on server. The maintenance cost of these data is reduced by storage as a service or SaaS. A data owner can compensate any misbehaviour committed by the CSP. Also, the CSP must be protected from any false accusation that may be claimed by the owner. This scheme allows the owner to perform full block-level dynamic operations like block modification, insertion, deletion, and append on sensitive data. It must provide the recent version of the data to the authorized user and the owner can grant or revoke the access to the stored data. This scheme also reduces communicational cost and computational cost. However, for protecting data, cloud services encrypt the data before outsourced to the commercial public cloud. Through this scheme, proposed system work as-strong-as-possible with no storage overhead.**

I. INTRODUCTION

Cloud environment can store large amount of sensitive data on cloud server. The data owners store more data on remote servers than on private computer systems by using the outsourcing data scheme. Better disaster recovery is provided by the CSP. Thus from the different geographic area, authorized users can access the data. The cloud storage system provides the space for storing data. The data owner chooses the authenticated user for sharing their data and decides which data is to be given to which user.

In the cloud storage system, the data owner outsources the data to the remote cloud storage provider and it takes full advantage of data management. Here normal encryption and decryption techniques are used. There is a direct mutual trust exist between CSP and data owner. But the data owner does not guarantee that he may not get latest version of datas. But, in this storage model, an indirect mutual trust exists between CSP and owner. So the owner gets the recent version of datas and also he can perform dynamic operations on these datas.

The outsourced data are stored in the cloud server in an encrypted format. By using the inter cloud communication, both key and encrypted data should be exchanged between users in different clouds. A proxy re-encryption scheme is used for reducing communicational cost and computational cost of stored data. The security requirements are given bellow.

- Confidentiality
  Outsourced data must be protected from CSP, users and the trusted third party.
- Integrity
  It must ensure that the authorized user gets the data without having any corruption.
- Newness
  The authorized user gets the recent version of the data that are stored in the cloud. If the CSP ignores data updations that are requested by the owner, a detection mechanism is used for identifying this.
- Access control
  There is a provision that the outsourced data could be accessed only by authorized users.
- Defence
  The CSP must be safeguarded against false accusations that may be claimed by dishonest owner.

The remainder of the paper is organized as follows. Section 2 reviews some related works. Proposed work is in Section 3. Finally, the paper is concluded in Section 4.

II. RELATED WORK

Existing research can be found in the area of integrity checking of outsourced data, distributed and cryptographic file systems. Himanshu *et al*. [1] proposed that cloud computing is an internet based technology that store large amount of data in remote server. Computer can store documents, materials and send email. If it does not work then our data will loss. So store documents and materials on cloud. The cloud provider has

- Software as a service
  It refers to software delivered over a browser. It eliminates the need to install and run applications on the customer's own computers and simplifies maintenance, upgrades and support.
- Platform as a service
  It facilitates the development and deployment of applications without the cost and complexity of buying, managing and configuring the underlying hardware, middleware and software layers.
- Infrastructure as a service
  It is the most basic and each higher model abstracts from the details of the lower models.

This describes about the performance of different security algorithm on a cloud network and also on a single processor for different input sizes and advanced encryption standard security algorithm implemented for ensuring security framework. Speed up ratio is defined as the ratio of mean processing time on a single processor to the mean processing time on the cloud.

Xiaoqiang Zhang *et al*. [2] proposed that the rapid development of Cloud computing, Internet of things and social network technologies, the network data is increasing dramatically. The security of massive data interaction has attracted more and more attention in recently years. This discusses the encryption principles, advantages and disadvantages of some mainstream massive data encryption technologies. More and more users are worried about the security of massive data interaction in the network environment.

The third party usually damages these data by the operations of intercepting malicious tampering, unauthorized copying or distribution. The main security requirements of massive data are confidentiality, integrity, copyright protection and encryption efficiency. In massive data encryption based on modern cryptosystem, modern cryptosystem can be classified into symmetric cryptosystems and asymmetric cryptosystems.

For a symmetric cryptosystem, the sender and receiver share an encryption key and a decryption key. These two keys are the same or easy to deduce each other. For an asymmetric cryptosystem, the receiver possesses a public key and a private key. The public key can be published, but the private key should be kept secret. Considering the difference of encryption speed, the symmetric cryptosystem always encrypt a large quantity of text data, and the asymmetric cryptosystem always encrypt the short massages, such as keys.

In massive data encryption based on parallel and distributed computing, parallel computing can be divided into task parallelism and data parallelism. The former would make the coordination and management of tasks very complex. However, the data parallelism divides a big task into several identical subtasks, and then it is much easier to handle these subtasks. Distributed computing is solving a big and complex problem by giving small parts of the problem to many computers to solve and then combining the solutions for the parts into a solution for the problem.

In massive data encryption based on biological engineering, DNA computing of biological engineering is a new natural method of calculation in recent years, and it has attracted more and more attention because of its high degree of parallelism and massive storage capacity. Set DNA as the information carriers, modern biotechnology as the implementation tool, then this method can utilize the inherent high storage density and high parallelism of DNA to achieve encrypted authentication, signature cryptography and other cryptographic functions.

In massive data encryption based on attribute based encryption, attribute based encryption is introduced first by Sahai and Waters, which can enable an access control mechanism over encrypted data by specifying the users' attributes. Attribute based encryption is an extension of the identity based encryption. It takes attributes as the public key and associates the cipher text and user's secret key with attributes, so that it can support expressive access control policies.

Brian Warner *et al*. [3] proposed that Tahoe is a system for secure, distributed storage. It uses capabilities for access control, cryptography for confidentiality and integrity, and erasure coding for fault tolerance. It has been deployed in a commercial backup service and is currently operational. The implementation is open source. Tahoe was designed following the principle of least authority that is, each user or process that needs to accomplish a task should be able to perform that task without having or wielding more authority than is necessary.

The data and metadata in the file system is distributed among servers using erasure coding and cryptography. Tahoe uses the capability access control model to manage access to files and directories. In Tahoe, a capability is a short string of bits which uniquely identifies one file or directory. Knowledge of that identifier is necessary and sufficient to gain access to the object that it identifies. The strings must be short enough to be convenient to store and transmit, but must be long enough that they are unguessable.

The Tahoe file system consists of files and directories. Files can be mutable, such that their contents can change, including their size, or immutable, such that they are writable only once. Each immutable file has two capabilities associated with it, a read capability or read cap for short, which identifies the immutable file and grants the ability to read its content, and a verify capability or verify cap, which identifies the immutable file and grants the ability to check its integrity but not to read its contents.

For mutable files, there are three capabilities, the read write cap, the read only cap, and the verify cap. Users who have access to a file or directory can delegate that access to other users simply by sharing the capability. Users can also produce a verify cap from a read cap, or produce a read only cap from a read write cap. This is called diminishing a capability. When writing, a writer chooses erasure coding parameters n, that is, the total number of shares that will be written and k, that is, the number of shares to be used on read. Using this codec allows an efficient byte wise implementation of encoding and decoding, and constrains the choice of erasure coding parameters to be 1 <= n <= 256 and 1 <= k <= n. Compared to erasure codes such as Tornado Codes, Reed Solomon codes offer optimal storage overheard. On the other hand, Reed Solomon codes suffer from asymptotically worse computational complexity.

Xinwen Zhang *et al.* [4] proposed that security has been considered as one of the critical concerns that hinder public cloud to be widely used. With the separation of data ownership and storage, a data owner has strong motivation to preserve its control of access and usage of shared data, while leverage storage, computation, and distribution functions provided by cloud, and desire that a public cloud should not learn any clear data. It has been widely recognized that data security should be mainly relied on cloud customers instead of cloud service providers.

This proposes, a certificate less proxy re encryption scheme for data sharing with cloud. Here, a data owner encrypts shared data in cloud with an encryption key, which is further encrypted and transformed by cloud, and then distributed to legitimate recipients for access control. Key management is another burdensome task for encryption based access control. Proxy re-encryption scheme augmented with certificate less public key cryptography, which leverages cloud not only for data storage but also for secure key distribution for data sharing.

Monica *et al.* [5] proposed that cloud computing is highly scalable, dynamic and easily configurable more over it can handle multiple request simultaneously. The basic cloud usage models are private cloud, public cloud, hybrid cloud and commodity cloud. Cloud technology is very dynamic, adaptable, scalable and easily accessible this flexible nature of cloud paves some severe drawbacks in terms of data security and it challenging to deduce a mechanism enabling secure data sharing. One of the suitable solutions for this problem is to follow an authenticated data centric approach rather than using communication centric approach. Cloud computing are of different types .These are public cloud, private cloud and hybrid cloud.

- Public Cloud
  Public cloud applications, storage, and other resources are made available to the general public by a service provider.
- Private Cloud
  Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third party and hosted internally or externally.
- Hybrid Cloud
  Hybrid cloud is a composition of two or more clouds that remain unique entities but are bound together, offering the benefits of multiple deployment models.

This model have employed a two step authentication process one is the login password authentication mechanism which is usually adopted scenario for user authentication at the server end for data access in a simple two or three tired client server architecture, with in the authentication phase of this secure hybrid algorithm, integrating an addition digital fingerprint mechanism to enhance the authentication process which is implemented using RSA [5] for digital fingerprint generation. This RSA can be used for the validation at the sender and receiver. Comparisons of above described papers are shown in the table 1 bellow.

Table 1: Comparison of existing systems

Existing Systems	Advantage	Disadvantage
Security issues in cloud	Reduce speed	Computation overhead
Encryption of the massive data	Reduce network cost	Storage overhead
Tahoe file system	Long term storage capacity	Inefficiency
CL proxy re-encryption	Avoid key escrow problem	Time consuming
Cloud security by cryptography	Scalable	Not guarantee that user to get modified data

III. GENERAL SYSTEM MODEL

Higher functionality like data forwarding is achieved in the cloud server in the following manner. Consider a situation when A wants to send a data to another user B. A will encrypt his data by using his own private key and stored on the cloud. This stored data are forward to the user B. After that, A downloads the data and decrypts it. After that A encrypt it by using public key of B. Then the message will stored in the cloud. When B receives the message he downloads the cipher text and then decrypts it by using his own secret key. For providing higher functionality, use this straight forward solution. This requires three communications round for downloading and uploading of A and also downloading of B. When the length of the message increases, the communication cost will also increases. After that the computational cost will increases.

A. Proxy Re-Encryption

Proxy re-encryption [6] is effective tool that reduces both the communicational cost and computational cost. Consider that, in this scheme A is a data owner and B is a user. If A want to send a message to B, then the re-encryption key will be forward to the storage server. And also the re-encryption key will also be sent to user B. Here the partial decryption is done by the proxy server. So that the length of the message will not influence its computational cost and hence the overhead of the user is minimized which will in turn reduces the communication cost.

If the data owner wants to forward the data to a user, then it is shared with the cloud and here the re-encryption key will also be kept. At the time of data forwarding, the re-encryption key will be forward to the proxy server. This is done by the owner. The encryption keys and the public key of the recipients are used for generating the re-encryption keys. The encryption key is used for encrypting the data in the codeword and then encoded. This is send to the recipient. The re-encryption key and data will also be forward to the user. On the receiving side, the data it will be first decoded and then decrypted using the secret key of recipient.

A cloud storage model consists of a collection of storage servers. It has long term storage services over the Internet. When the data is stored in the cloud, then the user will have no control on that data at that time and hence verifying the correctness of the data stored in the cloud is a challenging issue. So that confidentiality of the data stored in cloud is maintained by the data owner. He uses his private key for encrypting the data and this encrypted data will be stored in the cloud.

B. System Architecture

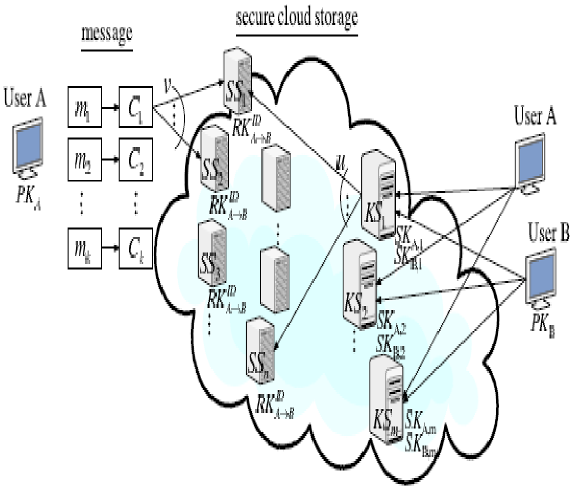

Fig 1: System Architecture

Figure 1 represents that, the system provides security for both the data stored in cloud and the data in transit. When a storage server fails, a new one is added. The new storage server queries k available storage servers, linearly combines the received codeword symbols as a new one and stores it. The system is then recovered.

The cloud storage system is treated as a large scale distributed storage system. It consists of the independent storage server and proxy servers. The storage system consists of users, n storage servers SS_1, SS_2… SS_n, and m key servers KS_1, KS_2… KS_m. Storage servers provide storage services and key servers provide key management services. They work independently. The system manager is responsible for choosing system parameters and publishes them. Each user A is assigned a public-secret key pair (PK_A, SK_A). User A distributes his secret key SK_A to key servers such that each key server KS_i holds a key share $SK_{A,\, i}$, $1 \leq i \geq m$. The key is shared with a threshold t. User A encrypts his message M and dispatches it to storage servers. A message M is decomposed into k Blocks m_1, m_2,….., m_k and has an identifier ID. User A encrypts each block m_i in to a cipher text C_i and sends it to v randomly chosen storage servers. Upon receiving cipher texts from a user, each storage server linearly combines them with randomly chosen coefficients into a codeword symbol and stores it. Note that a storage server may receive less than k message blocks and we assume that all storage servers know the value k in advance. To share the data, user A forwards his encrypted message with an identifier ID stored in storage servers to user B such that B can decrypt the forwarded message by his secret key. To do so, A uses his secret key SK_A and B's public key PK_B to compute a re-encryption key $RK^{ID}_{A\to B}$ and then sends $RK^{ID}_{A\to B}$ to all storage servers. Each storage server uses the re-encryption key to re-encrypt its codeword symbol for later retrieval requests by B.

The re-encrypted codeword symbol is the combination of cipher texts under B's public key. In order to distinguish re-encrypted codeword symbols from intact ones, we call them original codeword symbols and re-encrypted codeword symbols, respectively. Finally to retrieve data, user A requests to retrieve a message from storage servers. The message is either stored by him or forwarded to him. User A sends a retrieval request to key servers. Upon receiving the retrieval request and executing a proper authentication process with user A, each key server KS_i requests u randomly chosen storage servers to get codeword symbols and does partial decryption on the received codeword symbols by using the key share $SK_{A,\ i}$ Finally, user A combines the partially decrypted codeword symbols to obtain the original message M.

In proxy re-encryption scheme, the storage servers act as storage node for storing the message that is encrypted and then encoded. For storing the public and private key of both the data owner and the user, proxy server is used and also it stores the re-encryption key. A template is produced and stored, when a user first registers. After successful login of registered user, one can upload data to the cloud storage system, share the information with other authenticated user and also can download all the data.

C. Key and Message Management

During the login, it checks whether it is an authenticated user. The key will be encrypted using the RSA algorithm and key is stored in the proxy server in an encrypted form in order to ensure the security of that key. Message encrypted with public key can only be decrypted with private key. The data that is uploaded by the data owner will be encrypted using the private key of the data owner and for encrypting, Advanced Encryption Standard algorithm [7, 8] is used.

IV. CONCLUSION

Outsourcing data to remote servers are helpful for many organizations to alleviate the burden of local data storage and maintenance. In this approach, block level data dynamic operations, newness, mutual trust and access control of the sensitive data are offered. Proxy re-encryption scheme provides more security. Encoding, forwarding and partial decryption operations are supported in a distributed way by the proxy re-encryption scheme. Therefore, important features of outsourcing data storage can be supported without excessive overheads in storage, communication, and computation.

REFERENCES

[1] Himanshu , Arun Singh, "Evaluation and comparison of security issues on cloud computing environment", IEEE Trans. Information Theory, vol. 2, no. 5 pp.179-183, June 2012.

[2] Xiaoqiang Zhang, "General survey on massive data encryption" Proc. 12th Int'l Conf. Practice and Theory in Public Key Cryptography, pp. 357-376,june 2009.

[3] Brian Warner,"Tahoe-the least authority file system", Proc.Second USENIX Conf. File and Storage Technologies, pp.29-42, April 2006.

[4] Xinwen Zhang, "A certificate less proxy re-encryption scheme for cloud based data sharing", IEEE Trans. Information Theory, vol. 52, no. 6 pp. 2809-2816, June 2006.

[5] Monica, m.Sudha,"Enhanced security framework for security in cloud by using the cryptography", IEEETrans.Computer science and Applications, vol.1, no.1, 332-337, March 2012.

[6] S. D. C. di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, and P. Samarati, "Over-encryption: Management of access control evolution on outsourced data," in Proceedings of the 33rd International Conference on Very Large Data Bases. ACM, 2007, pp. 123–134.

[7] C. Wang, "Ensuring data storage security in cloud computing," Cryptology ePrint Archive, Report 2009/081, 2009.

[8] M. Sudha , Dr.Bandaru Rama Krishna Rao , M. Monica "A Comprehensive Approach to Ensure Secure Data Communication in Cloud Environment," in International Journal of Computer Applications (0975 – 8887) Volume 12– No.8, December 2010.