Study Of Integrity Of Information Stored In Cloud

Arti S. Bhor[1], Smita M. Pathare[2], Khushali J. Solanki[3], Madhuri D. Dhayarkar[4]
[1,2,3,4]Dnyanganga College Of Engineering And Research

***Abstract*— Cloud computing is having importance in current IT world. Cloud computing provides us with shared pool resources which we can access from anywhere without worrying about maintenance and management. It is important that data in the cloud should be correct, consistent and accessible and should have high quality. There is no guarantee that data stored in the cloud is secured and not altered by the Third Party Auditor (TPA).**

In this paper we provide scheme which gives trustworthiness of information stored in the cloud which the customer can employ to check the correctness of his data in the cloud. This proof can be agreed upon by both the cloud and the customer and can be incorporated in the Service level agreement (SLA). It also checks integrity with more accuracy.

Keywords* - *Cloud computing, Trustworthiness of data, TPA, Data integrity, SLA

I. INTRODUCTION

Cloud computing has given a new dimension to the complete outsourcing arena (Software as Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS)) and they provide ever cheaper powerful processor with these computing architecture. The simplest thing that computer does is to store in available space and retrieve information whenever requested by the authenticated user. Cloud system dynamically allocates computational resources in responds customers' resource reservation requests in accordance customers' predesigned quality of service. It helps enterprises to have a dynamically scalable abstracted computing infrastructure that is available on demand and on pay-per-use basis. Storing user data in the cloud despite its advantage has interesting security concerns which need to be extensively investigated to make it a reliable solution to problem avoiding local storage data. Many problems like data authentication and integrity, (i.e. how efficiently and securely ensure the cloud storage server returns correct, complete results in the response to its clients' queries[1])outsourcing encrypted data and associated difficult problems dealing with querying over encrypted domain [2] were discussed in research literature.

From the perspective of data security Cloud Computing inevitably poses new challenging security threats for number of reasons. At first, traditional cryptographic primitive for the purpose of security protection cannot be directly adopted due to the users' loss control data under Cloud Computing. Therefore, we require verification of data storage in the cloud. Considering various kinds of data for each user stored in the cloud and the demand of long term continuous assurance of their data safety, the problem of verifying accuracy of data storage in the cloud becomes even more challenging. Secondly, Cloud Computing is not just a third party data warehouse. The stored data in cloud may be frequently revised by the users, including operations like insertion, deletion, modification, affixing, reordering, etc. To ensure storage correctness under the dynamic data revise is hence of paramount importance.

Data integrity is defined as the accuracy and consistency of stored data, in absence of any alteration to the data between two updates of file and record. Although outsourcing of data into the cloud is economically attractive for cost and complexity of long term large scale data storage, its lacking of offering strong assurance of data integrity, availability impede its wide adoption by both enterprise and individual cloud users [3].

In this paper we deal with the problem of implementing protocol for obtaining the proof of data possession in the cloud sometimes referred as Proof of retrievability (POR). This protocol tries to obtain and verify a proof that the data that is stored by a user at remote data storage in the cloud is not modified by the archive and thereby the integrity of the data is assured. Such kinds of proofs are very helpful in peer-to-peer storage systems, network file systems, long term archives, web service object stores. Such verification systems prevent the cloud storage archives from misrepresenting or modifying the data stored at it without the consent of the data owner by using frequent checks on the storage archives. Such checks must allow the data owner to efficiently, frequently, quickly and securely verify that the cloud archive is not cheating the owner.

II. CLOUD COMPUTATION

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over networking (typically the Internet). The name comes from the use of a cloud shaped symbol as an abstraction for the complex infrastructure include in system diagram. Cloud computing entrusts remote services with user data, software and computation.

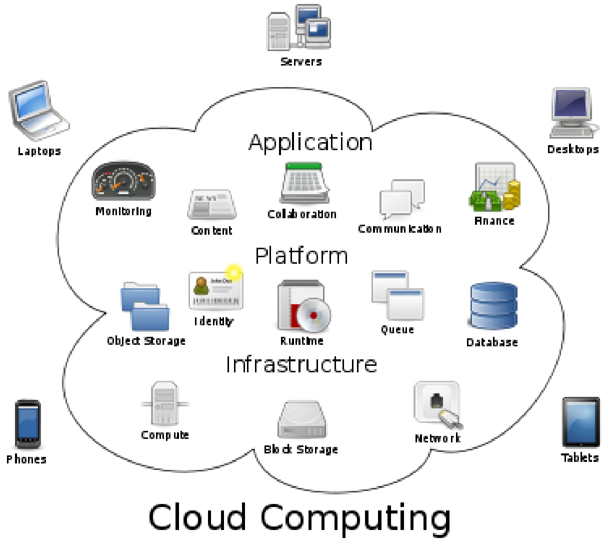

Figure1. Cloud Computing

The origin of the term cloud computing is obscure, but it appears to derive from the practice of using drawings of stylized clouds to denote networks in diagrams of computing and communications systems. The word cloud is used as metaphor for the internet based on the standardized use of a cloud like shape to denote a network on telephony schematics and later to depict the internet in computer network diagrams an abstraction of the underlying infrastructure it represents. The cloud symbol was used to represent the Internet as early as 1994.

III. Data Integrity In Cloud

A. *Data Integrity*

In computing, data integrity refers to maintaining and assuring the accuracy and consistency of data over its entire life cycle [1], and is an important feature of a database system. Data warehousing and business intelligence in general demand that the accuracy, validity and correctness of data is ensured from hardware failures, software errors. Data that has integrity is identically maintained during any operation, such as transfer, storage or retrieval.

B. *Data Integrity Proofs in Cloud*

Cloud storage can be attractive means of outsourcing the day-to-day management of data, but ultimately the responsibility and liability for that data falls on the company that owns the data not the hosting provider. It is important to understand that some of the causes of data corruption, how much responsibility a cloud service provider holds, some basic best practices for utilizing cloud storage safely, and some methods and standards for monitoring the integrity of data regardless of whether that data resides locally or in the cloud.

Integrity checking is essential in cloud storage for the same reasons that data integrity is critical for any data center. Data corruption can happen at any level of storage and with any type of media. Bit rot controller failures, reduplication metadata corruption and tape failures are all examples of different media types causing corruption. Metadata corruption can be the result of any of the vulnerabilities listed above, such as bit rot, but are also susceptible to software glitches outside of hardware error rates. Unfortunately, a side effect of reduplication is that a corrupted file, block, or byte affects every associated piece of data tied to that metadata. The truth is that data corruption can happen anywhere within a storage environment. Data can become corrupted simply by migrating it to a different platform, i.e., sending your data to the cloud. Cloud storage systems are still data centers, with hardware and software, and are still vulnerable to data corruption. One needs to look no further than the recent highly publicized Amazon failure. Not only did many companies suffer from prolonged downtime, but 0.07 percent of their customers actually lost data. It was reported that this data loss was caused by recovering an inconsistent data snapshot of Amazon ESB volumes.

IV. Secure Data Computation Outsourcing In Cloud

Fundamental concern to move computational workloads from private resources to the cloud is the protection of confidential data that computation consumes and produces. Secure computation outsourcings services are in great need to not only protect sensitive workload information but validate the integrity of the computation result. This is, however very difficult task due to number of challenges that have to be met simultaneously. Firstly such a service has to be practically feasible in terms of computational complexity. Secondly, it has to provide sound security guarantee without restriction system assumptions. Thirdly, it also has enable substantial computational savings at the end-user's side as compared to the amount of the efforts that otherwise has to be committed to solve the problem locally. Challenges practically exclude the applicability existing techniques developed in context of secure multi-party computation fully homomorphism encryption.

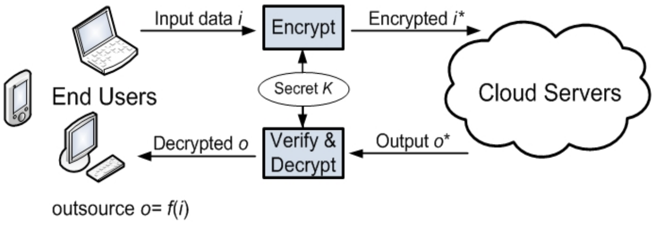

Figure2. Secure Data Computation Outsourcing In Cloud

Our methodology is to decompose computations into public programs and private data and leverage structures of specific computations achieving desirable trade-offs security, efficiency, and practicality. We plan to organize secure outsourcing mechanisms into hierarchy, where computation can be represented at various abstraction levels, such that the aforementioned trade-offs can be flexibly explored in a systematic manner. Two critical applications to be studied this project includes secure outsourcing systems of linear equations [4] and secure outsourcing linear programming [5] in the cloud. These two applications are among the most widely used algorithmic and computational tools in various engineering disciplines that analyze and optimize real-world systems. The study would prepare a solid knowledge base and provide insights for further research on more advanced computation problems, such as secure outsourcing convex programming in cloud.

V. INTEGRITY MAINTENANCE PROCESS

There are two phases in our project to check integrity of data. First is setup phase and second is verification phase. When client stored there data in cloud then this flow starts working.

A. Setup Phase

In setup phase we divided the file into n blocks which is having m no. of bits. Setup phase goes through the following steps:

1. *Creation of Meta-Data:*

Function is used to generate for each data block a set of k bit positions within the m bits that are in the data block. The value of k is in the choice of the verifier and is a secret known only to him. Therefore for each data block we get a set of k bits and in total for all the n blocks we get n*k bits. Figure 4 shows a data block of the file F with random bits selected using the function.

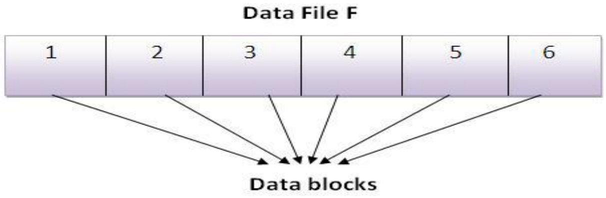

Figure3. A data file F with 6 data blocks

2. *Encrypting the meta data:*

Each of the Meta data from data blocks m is encrypted by using a suitable algorithm to give a new modified Meta data. Without loss of generality we show this process by using a simple XOR operation. Function which generates a k bit integer. This function is a secret and is known only to the verifier.

3. *Appending of meta data:*

All the Meta data bit blocks that are generated using the above procedure are to be concatenated together. This concatenated Meta data should be appended to the file F before storing it at the cloud server. The file F along with the appended Meta data F is archived with the cloud.

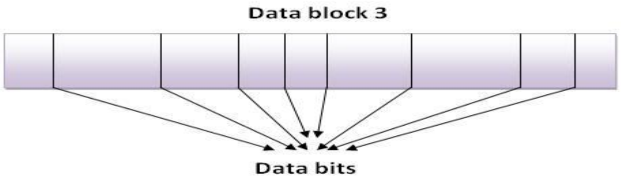

Figure4. A data block of the file F with random bits selected in it

B. Verification phase

Let the verifier want to verify the integrity of the file F. It throws a challenge to the archive and asks it to respond. The challenge and the response are compared and the verifier accepts or rejects the integrity proof. Suppose the verifier wishes to check the integrity of nth block. The verifier challenges the cloud storage server by specifying the block number and bit number generated by using the function which only the verifier knows. The verifier also specifies the position at which the Meta data corresponding the block is appended. This Meta data will be a k-bit number. Hence the cloud storage server is required to send k+1 bits for verification by the client. The meta data sent by the cloud is encrypted by using the number and the corresponding bit in this decrypted meta data is compared with the bit that is sent by the cloud. Any mismatch between the two would mean a loss of the integrity of the clients' data at the cloud storage.

VI. MATHEMATICAL MODEL

1. *Set Theory:*

We will have the following sets.

A = set of data on cloud
B = Set of keys
C = Set of users
D = Set of unique users data on cloud
Here, D = A intersection B intersection C

2. *Venn Diagram:*

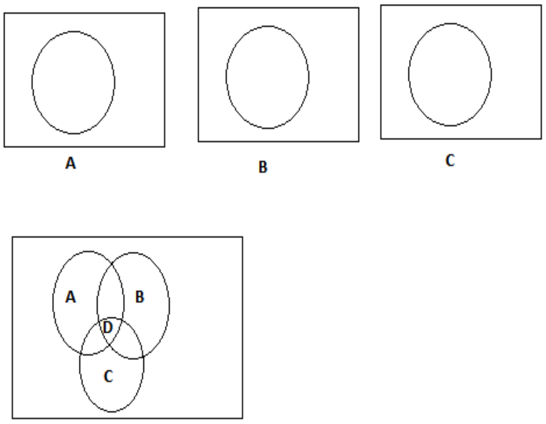

Figure5. Venn Diagram

Type of Problem:

1) Generation of meta-data: Let g be a function defined as follows g(i, j) ! {1..m}, i 2 {1..n}, j 2 {1..k}

Where k is the number of bits per data blocks which we wish to read as Meta data. The function g generates for each data block a set of k bit positions within the m bits that are in the data block.

2) Encrypting the Meta data:

Let h be a function which generates a k bit integer alpha i for each i. This function is a secret and is known only to the verifier V

h : i ! _i, _i 2 {0..2n}

For the Meta data (mi) of each data block the number alpha I is added to get a new k bit number Mi.

Mi = mi + alpha i

In this way we get a set of n new Meta data bit blocks. The encryption method can be improvised to provide still stronger protection for verifiers' data.

VII. Conclusion

In this paper we have tried to give the assurance to the clients that their data is secured in cloud server with the help of proof of data integrity hence confidentiality of users' sensitive files is maintained. This proof can be agreed upon by both the cloud and the customer and can be incorporated in the Service level agreement (SLA). Maintaining the storages can be a difficult task. It transmits the file across the network to the client thus consume heavy bandwidths. Thus we have also tried to reduce this network bandwidth consumption. It also reduces the chance of losing data by hardware failures. Here we are going to use hash function for the encrypted data. This scheme is more advantageous to the mobile phones and PDAs which have limited CPU power, battery power and communication bandwidth. It evaluates the performance of cloud storage as it consumes less computational power.

References

[1] E. Mykletun, M. Narasimha, and G. Tsudik, "Authentication and integrity in outsourced databases," Trans. Storage, vol. 2, no. 2, pp. 107–138, 2006

[2] 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, p. 44.

[3] Cong Wang, Qian Wang, Kui Ren, Wenjing Lou, "Towards Secure and Dependable Storage Services in Cloud Computing," IEEEtransactions on Services Computing, 06 May 2011

[4] Cong Wang, Kui Ren, Jia Wang, and Karthik Mahendra Raje Urs, "Harnessing the Cloud for Securely Solving Large Systems of Linear Equations," The 31st International Conference on Distributed Computing Systems (ICDCS'11), Minneapolis, MN, June 20-24, 2011. (Note: this online version is the extended full paper of the conference camera-ready one.)

[5] Cong Wang, Kui Ren, and Jia Wang, "Secure and Practical Outsourcing of Linear Programming in Cloud Computing", The 30th IEEE Conference on Computer Communications (INFOCOM'11), Shanghai, China, April 10-15, 2011. (Note: this online version is the extended full paper of the conference camera-ready one.)

[6] A.Juels and B.S. Kaliski, Jr., "Pors: proofs of retrievability for large files," in CSS' 07:Proceedings of6 the 14th ACM conference on Computer and communications Security.New York, NY, USA:ACM, 2007, pp.584-597

[7] Boritz, J. Efrim. "IS Practitioners' Views on Core Concepts of Information Integrity". International Journal of Accounting Information Systems. Elsevier. Retrieved 12 August 2011.