ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 4, April 2013

Data Security And Integrity Of Cloud Storage In Cloud Computing

Yogita Gunjal[1], Prof. J.Rethna Virjil Jeny[2]

ME.student, Dept. Of Information technology, Amrutvahini College of Engineering,Sangamner, Maharashtra, India [1]

Associate Professor, Dept. Of Information technology, Amrutvahini College Of Engineering, Sangamner, Maharashtra,India[2]

Abstract: Cloud computing is the computing paradigm which enable obtaining resources like software, hardware, services over the internet. Most of user store their data on cloud for data security and integrity are prime related. In this article the problem of ensuring data integrity and security of data storage in cloud computing. For ensuring correctness of data, we assume the task of allowing a Third party auditor(TPA) used for exposing risk of cloud storage services on behalf of the cloud client to verify data integrity stored in the cloud. This paper focus on the data security, we proposed erasure correcting code in the file distribution to provide the redundancies and guarantee data dependability. By using homomorphic token with distributed verification of erasure coded data ,our scheme achieve storage correctness as well as error localization. Extensive security analysis show the proposed scheme is highly efficient and resilient against Byzantine failure, malicious data alternation attack and even server colluding attacks.

Keywords: Cloud computing ,Data integrity ,TPA, Cloud Client.

I. INTRODUCTION

Cloud computing is computing paradigm in which task are assigned to a combination of connections software, hardware, services over the internet. Several trends are opening up the era of cloud computing which is use for computer technology. Conceptually, users get computing platform from computing clouds and then inside run their applications. Always cheaper and more powerful processors together with the software-as-a-service computing architecture. These are transforming data centers into of computing service on huge scale. The increasing network bandwidth and reliable yet flexible network connections make it even possible that users can now subscribe high quality service from data and software that reside solely on remote data centers. Cloud offers great convenience to users 'when transforming data into cloud since cloud client don't have to care about the complexities of direct hardware management. In cloud computing there are well example Amazon Elastic computer cloud(EC2)[3].

While these internet based online services do provide huge amount of storage space and customizable computing resources. This IT infrastructure shift, however is eliminating the responsibility of local machines data in cloud computing vendors maintenance at the same time. Result for users are at the mercy of their cloud service providers for the data availability and data integrity[6]. Data integrity defines accuracy and consistent data is stored is indicated by an modification in data between two updates of the data record. This data integrity refers the valid of data .In ensuring cloud data storage , the cloud data storage system ,cloud clients store their data in the cloud and it has no longer possess the data locally in cloud[10]. Recently, the importance of ensuring the remote data integrity[3]. These techniques are, while can be used to ensure the storage correctness instead of having cloud clients possessing data It cannot address all the security threats in cloud data storage, so they are all focusing on individual's server scenario and most of them not consider dynamic data operations. To addressing this problem for ensuring cloud data storage correction chosen to reserve the homomorphic token properties , which can be perfectly integrated with the verification of erasure coded data in cloud[11]. Section II describes with system model and adversary model in background. Related work in section III. Proposed system presented in section IV and conclusion describes in section V.

II. BACKGROUND

In this paper shows two model in problem statement. Those are system model and adversary model.

A. System Model:

In this model shows representative network architecture of cloud services architecture. This is shown in Fig 1. In that there are three network entities are as follows:

- User: User stored their data in the cloud and relies on cloud for data storage and computation task.
- Cloud Server: Which is managed by cloud service provider(CSP) to provide data storage has significant storage space and computation resource.
- Third Party Auditor: TPA, who is trusted and expose risk of cloud storage services on behalf of cloud client upon request.

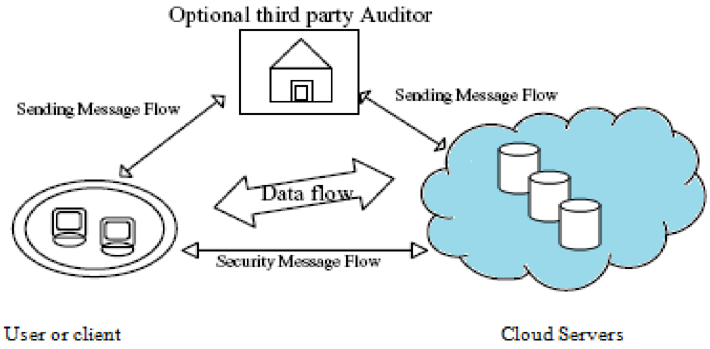

Fig.1. Cloud Storage Service Architecure.

In the cloud storage architecture, cloud make it possible to access our information from any where at any time. In that there are four cloud storages are as follows:

- **Public cloud storage:** in public cloud storage, it can access by any subscriber with an internet connection and access to the cloud space.
- **Private cloud storage:** In private cloud storage, it is established for a specific organizations and limits to access to those organizations.
- **Hybrid cloud storage**: In hybrid cloud storage, it is combination of the public and private cloud storage. It means where critical cloud data located in private cloud while other data is stored and accessed from public cloud.
- **Mobile cloud storage:** In mobile cloud storage, it stores the separate data in the cloud and access it from anywhere at any time.

cloud client stores theirs data in cloud data storage through a cloud service provider into a set of cloud servers, which occurring at the same time and running in cooperated manner. Redundant of data can be employed with technique of erasure precisely code to further tolerate faults or server crash as user's data grows in size. In this paper, this dynamic features makes also traditional integrity insurance techniques useless and requires new solutions[8].Therefore, data storage correctness assurance will be of most necessary in achieving a strong and safely cloud data storage system in the real world[7]. In other words, the cloud data we are assuming is unexpected to be rapidly changing in a relative short period. In this system model shows point to point communication between cloud service provider and cloud client.

B.Adversary Model

This model capture all kinds of data integrity threats and this cloud data not denoted at cloud client side but at the cloud service provide domain address. This can come from two attacks:

- Internal Attack: Cloud service provider can be untrusted.
- External Attack: This attack comes from outsiders and who are beyond control domain of cloud service providers.

III.RELATED WORK

The importance of ensuring the remote data integrity has been highlighted by the following research. Works under different security models. and these can be useful to ensure the storage correctness without having users possessing local data are all focusing on single server scenario[1]. Provable data possession model for ensuring possession of file on untrusted storages[9].Although direct applying these techniques to multiple servers could be straightforward, the resulted verification would be linear to the number of servers.

- Jules et al.[5] defined a formal “proof of retrivabilty”(POR) model for ensuring the remote data integrity, their scheme combines spot-checking as well as error correcting code to ensure both possession and getting of files on archive service systems.
- Bowers et al.[4] extended “proof of retrivabilty”(POR) model to distributed systems, all these schemes are focusing on static data. The effectiveness of their scheme rests mainly on the proposing steps that the user conducts before outsourcing the data file. Any change to the contents of data file ,even few bits must propagate through the error-correcting code and the corresponding random shuffling process ,thus Introducing significant computation and communication complexity, However the token pre-computation of the tags imposes heavy computation overhead that can be expensive for an whole file.

IV. PROPOSED SYSTEM

In this paper, it focus on cloud data storage security, which has always been an most aspect of quality of service. For ensuring the correctness of cloud clients data in the cloud, in this paper propose an highly effective and flexible distributed scheme with two features, apposing to its predecessors. By using the homomorphic token with distributed verification of erasure coded data. In This paper proposed the integration of storage correctness insurance and data error localization most of works, the new scheme further supports secure and efficient dynamic operation on data block including operations. We relies on erasure-correcting code in the file distribution preparation to support redundancy parity vectors for verification of erasure coded data using the homomorphic token, In this paper our scheme achieves the integration of data error localization and storage correctness insurance. This paper proposed highly effective and flexible distributed scheme with explicit dynamic data provide to ensuring the correctness of user’s data in the cloud. our scheme enable the data owner to delegate of data file re-encryption and user secret key update to cloud servers without disclosing data contents .In this paper we achieves this goal by exploiting and uniquely combing techniques that is token precomputation, error correctness verification as well as error localization and error recovery.

In the first reason cryptography services for the intention of data security protection could not be directly adopted due to the users’ loss control of data under cloud computing. So, verification of correct data storage in the must be conducted without explicit knowledge of the entire data. Assuming various kinds of data for every cloud client stored in the cloud and requirements of long term continuous assurance of their data safely, the problem is that verifying exactness of data storage in the cloud becomes even more challenging. This construction drastically decreasing the communication and storage overhead as compared to the based file of replication in distribution techniques. Therefore correctness of data and availability of the data being stored on the distributed cloud servers may be guaranteed. The key issues is to highly detect any unauthorized data alternation and corruption, possibly due to server compromise byzantine failure.

The token computation function we are considering belongs to a family of universal hash function, chosen to storage the homomorphic property. which can be completely integrated with the verification of erasure-coded data. then, it is shown how to derive a challenge-response protocol for verifying the storage correctness as well as identifying misbehaving servers. Finally the process for file recovery and error resurgence based on removal- correcting code is also outlined. It is well known that erasure-correcting code may be used to stand multiple failures in distributed storage space systems. In cloud data storage, we relies on this technique to dissolve the data file F redundantly across a set of n = m+ k distributed servers.

Notation and Preliminaries:

f– The data file to be stored. We consider that F can be denote as a matrix of m equal-sized data vectors, each and every consisting of l blocks. Data blocks are all represent as elements in Galois Field GF (2p) for p = 8 or 16.

1. A – The dispersal matrix used for Reed-Solomon coding.
2. G – The encoded file matrix, which includes a set of n = m + k vectors, each consisting of l blocks.
3. f key (•) – pseudorandom function (PRF), which is defined as f : {0, 1}∗ × key → GF (2p).
4. φ key (•) – pseudorandom permutation (PRP), which is defined as φ : {0, 1}log2 (ℓ) × key → {0, 1}log2 (ℓ).
5. ver – a version number bound with the index for single blocks, which records the times the block has been changed Initially we consider ver is 0 for all data blocks.

Let **F** = (F1 , F2 , . . . , Fm) and Fi = (f1i , f2i, . . . , fli)T (i ∈ {1, . . . , m}). Here T (shorthand for transpose) de- notes that each Fi is represented as a colmn vector, and data vector size denoted l in blocks. All these blocks are elements of GF (2p). The systematic layout with parity vectors is achieved with the information dispersal matrix A, derived from a an m×(m+k) Vandermonde matrix[7] : 1 1 … 1 1 … 1 β1 β2 … βm βm+1 … βn . . . . β1m-1 β2m-1 … βmm-1 … βm+1m-1 βnm-1 where βj (j ∈ {1, . . . , n}) are distinct elements randomly Picked from GF (2p). After a series of basic row transformations, the preferred matrix A can be written as

1 0 . . . 0 P11 P12 . . .P1k 0 1 . . .

0 P21 P22 . . . Pmk

A = (I|P)= . . . .
. . . .
. . 0 0 . . . 1 Pm1 Pm2 . . . Pmk

By multiplying **F** by **A**, the user obtains the encoded file: G = F · A = (G(1) , G(2) , . . . , G(m) , G(m+1) , . . . , G(n)) = (F1, F2 , . . . , Fm , G(m+1) , . . . , G(n)), Where G(j) = (g1(j), g2(j), . . . , gl(j))T (j ∈ {1,…,n }). noticed, the multiplication reproduces the original data file vectors of **F** and the remaining part (G(m+1) , . . . , G(n)) are k parity vectors generated based on **F**.

B.Token pre-computation:

- Before file distribution the cloud clients pre-computes a certain number of short verification tokens on individual vector.
- Cloud client wants to make sure storage correctness for the data in the cloud; he challenges the cloud servers with a set of randomly generated block indices

Algorithm 1 Token Pre_computation

```
procedure
    Choose parameters l, n and function f, $\varphi$;
    Choose the number t of tokens;
    Choose the number r of index per confirmation;
    produce master key Kprp and challenge key kchal;
    for vector G(j) , $j \leftarrow 1, n$ do
        for round $i \leftarrow 1, t$ d
            Derive $\alpha i$ = fkchal (i) and kprp from Kprp
            compute
        end for
    end for
    store all the vis locally
end procedure.
```

B. Correctness Verification and Error Localization:

- Integrates the correctness verification and error localization (misbehaving server identification) using challenge response protocol.
- The response values from cloud servers for every challenge not only determine the correctness of the distributed storage but also contain information to locate potential data error(s).
- The comparison between precomputed tokens and received response values can guarantee the identification of misbehaving server(s) when the data is corrupted.
- The cloud client can reconstruct the original file by downloading the data vectors from the first m servers, assuming that they return the correct response values.

Algorithm 2 Correctness Verification and Error Localization:

```
procedure CHA L L E NG E (i)
    ecompute $\alpha i$ = fkchal (i) and kprp from KPRP
    Send { $\alpha i$, k( i )prp } to all the cloud server
    Receive from servers: {Ri( j ) = $\Sigma$rq=1 $\alpha i * $ G(j) [$\varphi$ (i) (q)]|1 $\le$ j $\le$n}
    for (j $\leftarrow$ m + 1, n) do
        R( j ) $\leftarrow$ R( j ) -$\Sigma$rq=1 $f$k j (sIq ,j ) · $\alpha i$ , Iq = $\varphi$kprp(i) (q)
    end for
    if (( Ri ( 1 ) ,…, Ri ( m ) ) • P = = ( Ri ( m+1 ) , . . ., Ri ( n ) )) then
        Accept and ready for the next challenge.
    else
        for ( j $\leftarrow$ 1, n) do
            if (Ri ( j ) ! = vi( j ) then
                return server j is misbehaving.
            end if
        end for
    end if
end procedure
```

V. CONCLUSION

In this paper, we examine the problem of data security problem stored the in cloud data storage, which is mostly a distributed storage system. We relies on erasure-correcting code in the file distribution preparation to support redundancy parity vectors for verification of erasure coded data using the homomorphic token, In this paper our scheme achieves the integration of data error localization and storage correctness insurance. security analysis shows that our scheme is highly efficient and resilient to Byzantine failure, malicious data change attack, and even server colluding attacks.

REFERENCES

[1] C. Wang, Q. Wang, K. Ren, and W. Lou, "Ensuring data storage security in cloud computing,"in Proc.of IWQoS'09, July 2009, pp.1–9

[2] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik,"Scalable and Efficient Provable Data Possession," *Proc. Of* SecureComm '08, pp. 1– 10, 2008.

[3] Amazon.com, "Amazon s3 availability event: July20,2008,"Onlineathttp://status.aws.amazon.com/s3-20080720.html, July 2008..

[4] K. D. Bowers, A. Juels, and A. Oprea, "Hail: A high-availability and integrity layer for cloud storage," in Proc. *of* CCS'09, 2009,pp. 187–198.

[5] A. Juels and J. Burton S. Kaliski, "Pors: Proofs of retrievability for large files," in *Proc. of* CCS'07, Alexandria, VA, October 2007, pp.584–597

[6] Sun Microsystems, Inc., "Building customer trust in cloud computing with transparent security," Online at https://www.sun.com/offers/details/sun transparency.xml, November 2009.

[7] Q. Wang, K. Ren, W. Lou, and Y. Zhang, "Dependable and secure sensor data storage with dynamic integrity assurance," in *Proc. Of* IEEE INFOCOM'09, Rio de Janeiro, Brazil, April 2009.

[8] Q. Wang, K. Ren, W. Lou, and Y. Zhang, "Dependable and secure data storage with dynamic integrity assurance," in *Proc. Of* IEEE INFOCOM'09, Rio de Janeiro, Brazil, April 2009.

[9] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissinger, Z.Peterson, and D. Song, "Provable Data Possession at Untrusted Stores," Proc. 14th ACM Conf. Computer and Comm. Security (CCS '07), pp. 598-609, Oct. 2007.

[10] T. S. J. Schwarz and E. L. Miller, "Store, Forget, and Check: Using Algebraic Signatures to Check Remotely Administered Storage," *Proc.of ICDCS '06*, pp. 12–12, 2006.

[11] J. Hendricks, G. Ganger, and M. Reiter, "Verifying Distributed Erasurecoded Data," *Proc. 26th ACM* Symposium on Principles of Distributed Computing, pp. 139–146, 2007.

BIOGRAPHY

Yogita Gunjal is doing M.E in Information Technology in Amrutvahini College of Engineerung,Sangamner, Maharashtra. She did her B.E. in Information Technolgy from Amrutvahini College of Engineering, Sangamner, Maharashtra, India.

J.Rethna Virgil Jeny has been serving as a Associate Professor in Information Technology Department department in Amrutvahini College of Engineering,Sangamner, Maharashtra,India**.**She did her B.E and M.E degrees in Computer Science and Engineering from Bharathidasan University, Trichy in 1997 and Annamalai University in 2005 respectively. She is currently doing Ph.D in wireless sensor Networks at MS University. She has received Lady Engineer Award by IEI. She is a member of IEEE, ACM, ISTE, IEI and a senior member of IACSIT. Her research interests include Energy aware routing and Cross layer routing in Wireless Sensor Networks