A New Approach of Providing Data Security in the Cloud

V.V.Satyanarayana Kadali [#1], Naga Raju Katta[*2], Dr V.V.Krishna[*3]

[#1] IInd M.TECH (CSE) Student,[*2] Associate Professor, [*3]Professor
Department of CSE ,
GIET Engineering College,
Rajahmundry, AP, India.

Abstract

A new feature of the cloud computing is that user's data are usually processed remotely in unknown machines that users do not own or operate. As the evolvence of this new emerging technology with a lot of convenience brought by user's, the user's also fears of losing control of their own data (particularly financial data and health information) can become a crucial barrier to the wide adoption of cloud services. In this paper, we mainly focus on cloud data storage security, which has always been an important aspect of QOS. To find out complete assurance of user data in the cloud with maximum correctness of user's data in the cloud server, we propose a new effective and flexible distributed scheme with two main important key features, opposing to its predecessors. By utilizing the new homomorphic token with distributed verification of erasure-coded data, our proposed scheme is able to identify the presence of misbehaving server(s). By conducting several extensive security and performance experiments we show that the proposed scheme is highly efficient and resilient against several Byzantine failure, malicious data modification attack, and even server colluding attacks.

Keywords

Cloud Computing, Byzantine failure, homomorphic token, distributed scheme.

1. Introduction

In recent days several cloud users regarded cloud computing as an ingenious collaboration of a series of multiple technologies, establishing a simple business model by offering Information Technology services and also by using economies of scale [1], [2]. With this type of cloud computing services several participants who are present in the business chain of cloud computing can benefit from this new model. By using this new simple model cloud customers can save their huge capital investment of IT infrastructure, and can only concentrate on their own core business. Therefore, for this reason many small companies or large scale organizations have been migrating or building newly their business into cloud.

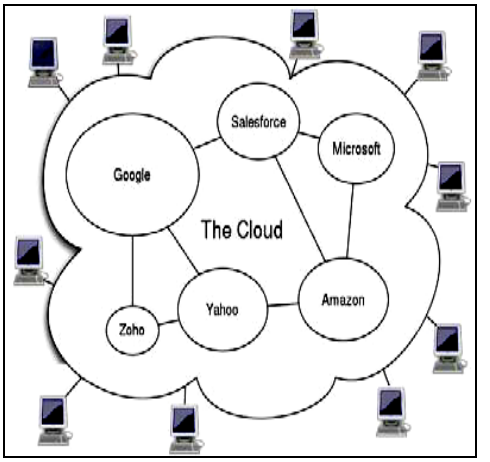

Figure. 1 Overview of a cloud

Cloud Server which is located with various types of services for the customers always provide the ever cheaper and more powerful services, together with the software as a service (SaaS) computing architecture, through which they are mainly transforming a large data centers into several pools of computing service on a large scale. The increase of network bandwidth and reliable nature of cloud server make it possible for the cloud users to subscribe high quality services from data and software that reside solely on remote data centers.

From the user point of view for data security in the cloud, which is always treated as an important aspect of quality of service, Cloud Computing inevitably poses new challenging security threats for a number of reasons as shown in figure 1.

Recently in several research works [3], [4], [5], [6], [7] we have observed the privilege of ensuring the remote data integrity. Along with this approach, researchers have also proposed several distributed protocols [8][9][10] for ensuring the data storage correctness across various servers or network peers. In this current paper, we mainly propose a new effective and flexible distributed scheme with explicit dynamic data support to ensure the correctness of users' data in the cloud.

2. Problem Statement

In this section, we will find some important concepts which are related to cloud computing as well as security issues in the cloud. Later we briefly discuss works which adopt similar techniques as our approach but serve for different purposes.

2.1 Cloud System Model

To demonstrate system model Figure.2 clearly describes the representative network architecture for cloud data storage. They are mainly three different network entities for describing the proposed model, they are as follows:

1. **User Module**: In this cloud system model, the User's are the persons, who have data to be stored in the cloud server and rely on the cloud for data computation, consist of both individual consumers and organizations.

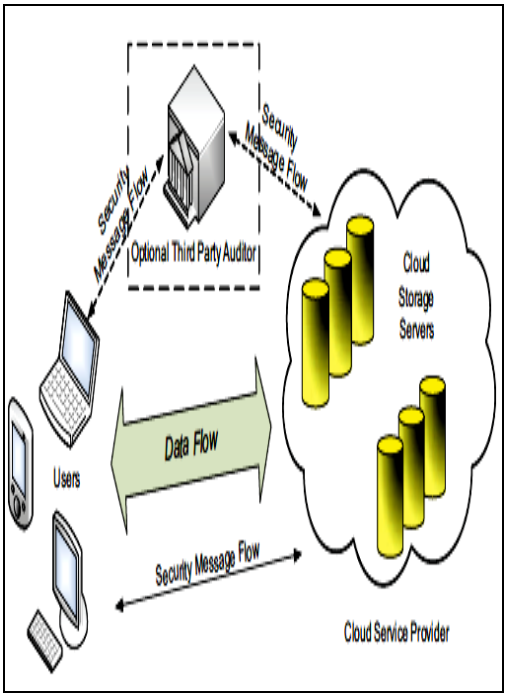

Figure. 2 Cloud data storage architecture

2. **CSP Module:** In the Cloud System model, Cloud Service Providers are those persons, who has significant resources and also expert in building and managing distributed cloud storage servers, they mainly owns the live Cloud Computing systems for operating it.

3. **TPA:** In the Cloud System model, Third Party Authors are the persons who have highest privileges than the normal persons in the cloud environment. He is the only trusted person to assess and expose risk of cloud storage services on behalf of the users upon request.

2.2 Cloud Adversary Model

In this paper we mainly divided the cloud adversary models of two types based on CSP.They are as follows

1. Cloud Weak Adversary Model

An adversary model is said to be cloud weak adversary model, if the appropriate user is mainly intended in corrupting the other user's data files which is stored on their individual servers. If such type of intended fault is created the router obviously becomes compromised, so that an adversary person can easily misuse the original data files by modifying or introducing its own fraudulent data to prevent the original data from being retrieved by the user.

2. Cloud Strong Adversary Model

A cloud model is said to be strong adversary model if we assume that the adversary node or adversary person can compromise all the complete data which is stored on the individual storage servers so that he/she can intentionally modify the data files as long as they are internally consistent. We can also compare the same to the case where all servers are colluding together to hide a data loss or corruption incident.

3. Ensuring Cloud Data Storage

In this section we mainly propose important methods for ensuring cloud data storage. The first part of this section is mainly deals with a review of basic programming tools from coding theory that is needed in our scheme for file distribution across cloud servers. We used java as coding technology for this current research work the homomorphism token is introduced after finishing the first part. We mainly consider the token computation function from a family offset of universal hash function [11], chosen to preserve the homomorphic properties, which can be perfectly integrated with the verification of erasure-coded data [8] [12].

3.1 File Distribution Preparation

In cloud data storage, we mainly rely on the file distribution technique to distribute the data file of F redundantly across a set of n = m + k distributed servers.

Algorithm 1 Token Pre-computation

procedure
  Choose parameters l, n and function f, ϕ;
  Choose the number t of tokens;
  Choose the number r of indices per verification;
  Generate master key K_{prp} and challenge k_{chal};
  for vector $G^{(j)}$, $j \leftarrow 1, n$ **do**
    for round $i \leftarrow 1, t$ **do**
      Derive $\alpha_i = f_{k_{chal}}(i)$ and $k_{prp}^{(i)}$ from K_{PRP}.
      Compute $v_i^{(j)} = \sum_{q=1}^{r} \alpha_i^q * G^{(j)}[\phi_{k_{prp}^{(i)}}(q)]$
    end for
  end for
  Store all the v_is locally.
end procedure

An set of (m, k) which is also known as Reed-Solomon erasure-correcting code is mainly used to create k redundancy parity vectors from m data vectors in such a way that the original m data vectors can be reconstructed from any m out of the m + k data and parity vectors.

3.2. Challenge Token Pre-Computation

Our proposed token pre computation scheme entirely relies on the pre-computed verification tokens. The main idea for this technique is as follows: before file distribution the user pre-computes a certain number of short verification tokens on individual vector $G^{(j)}$ ($j \in \{1, \ldots, n\}$), each token covering a random subset of data blocks.

Algorithm 2 Correctness Verification and Error Localization

procedure CHALLENGE(i)
  Recompute $\alpha_i = f_{k_{chal}}(i)$ and $k_{prp}^{(i)}$ from K_{PRP};
  Send $\{\alpha_i, k_{prp}^{(i)}\}$ to all the cloud servers;
  Receive from servers:
  $\{R_i^{(j)} = \sum_{q=1}^{r} \alpha_i^q * G^{(j)}[\phi_{k_{prp}^{(i)}}(q)] | 1 \le j \le n\}$
  for $(j \leftarrow m+1, n)$ **do**
    $R^{(j)} \leftarrow R^{(j)} - \sum_{q=1}^{r} f_{k_j}(s_{I_q,j}) \cdot \alpha_i^q, I_q = \phi_{k_{prp}^{(i)}}(q)$
  end for
  if $((R_i^{(1)}, \ldots, R_i^{(m)}) \cdot \mathbf{P} == (R_i^{(m+1)}, \ldots, R_i^{(n)}))$ **then**
    Accept and ready for the next challenge.
  else
    for $(j \leftarrow 1, n)$ **do**
      if $(R_i^{(j)} ! = v_i^{(j)})$ **then**
        return server j is misbehaving.
      end if
    end for
  end if
end procedure

4. Problem Implementation Modules

Implementation is the stage of the project where the theoretical design is completely turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. As our application is implemented in java as the chosen technology with Java Swings as front end user interfaces. The following paper is divided into following five modules. They are as follows:

4.1 CAM

This is the main important module in our implementation of this project also known as Client Authentication Module, the client requests the file from the server. Once the query sent by the client is successfully received by the server. The server then checks the requested file and then sends that file to the client. To process all these activities initially client authentication should be done by the server side.

4.2 CSM

This is also one of the main important modules in our implementation of this project also known as Cloud System Model, we will identify three different network entities, and they are as follows:

- User
- CSP
- TPA.

These three different entities are the key entities for designing the system model.

4.3 CDSM

This is also one of the main important modules in our implementation of this project also known as Cloud Data Storage Module, which is mainly used for storing valuable sensitive data into the individual set of cloud storage servers .This storage facility is given access with the help of content service providers. In our proposed model, we have created a network which is of point-to-point communication channels in a secured and authenticated manner between each and every individual cloud server and the individual participating user.

4.4 CAS

In this module the Cloud Authentication Server is having some additional new behaviors added when compared with traditional client-authentication protocol. This is the latest authentication server which is not at all used in any of the existing systems.

4.5 UDMCM

This is also one of the main important modules in our implementation of this project also known as Unauthorized Data Modification & Corruption Module, by using this module we can effectively detect any unauthorized data modification and data corruption which is altered or attempted by any compromised node.

5. Conclusion

In this paper, we mainly investigated the problems that are caused during providing data security in cloud data storage, which is essentially

in a distributed storage system. To ensure whether user data is inserted correctly in cloud data storage, we proposed a new effective and flexible distributed scheme with explicit dynamic data support, including block update, delete, and append.

6. References

[1] P.T. Jaeger, J. Lin, and J.M. Grimes, "Cloud Computing and Information Policy: Computing in a Policy Cloud?," J. Information Technology and Politics, vol. 5, no. 3, pp. 269-283, 2009.

[2] S. Pearson and A. Charles worth, "Accountability as a Way Forward for Privacy Protection in the Cloud," Proc. First Int'lConf. Cloud Computing, 2009.

[3] A. Juels and J. Burton S. Kaliski, "PORs: Proofs of Retrievability for Large Files," *Proc. of CCS '07*, pp. 584–597, 2007.

[4] H. Shacham and B. Waters, "Compact Proofs of Retrievability," *Proc. of Asiacrypt '08*, Dec. 2008.

[5] K. D. Bowers, A. Juels, and A. Oprea, "Proofs of Retrievability: Theory and Implementation," Cryptology ePrint Archive, Report 2008/175, 2008, http://eprint.iacr.org/.

[6] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable Data Possession at Untrusted Stores," *Proc. of CCS '07*, pp. 598–609, 2007.

[7] 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.

[8] 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.

[9] M. Lillibridge, S. Elnikety, A. Birrell, M. Burrows, and M. Isard, "A Cooperative Internet Backup Scheme," *Proc. of the 2003 USENIX Annual Technical Conference (General Track)*, pp. 29–41, 2003.

[10] K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A High-Availability and Integrity Layer for Cloud Storage," Cryptology ePrint Archive, Report 2008/489, 2008, http://eprint.iacr.org/

[11] L. Carter and M. Wegman, "Universal Hash Functions," *Journal of Computer and System Sciences*, vol. 18, no. 2, pp. 143–154, 1979.

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

7. About the Authors:

V.V.Satyanarayana Kadali is currently pursuing his M.Tech (CSE) in Computer Science & Engineering Department, GIET Engineering College, Rajahmundry. His area of interests includes Networks, Cloud Computing.

Naga Raju Katta is currently working as an Associate Professor in Computer Science & Engineering Department, GIET Engineering College, Rajahmundry. His research interests include Networks, Data Mining and Cloud Computing.

Dr V.V.Krishna is currently working as a Professor for Computer Science & Engineering Department, GIET Engineering College, Rajahmundry. He is awarded with PhD in related field. His research interests include Image Processing, Data Mining & Warehousing, Networks and Security, Software Engineering.