ISSN:0975-9646

# A Survey on Data Security of Cloud Storage Services

Gunjal Yogita S.,

*Amrutvahini College of Engineering, India*

***Abstract*— Cloud computing is the convenient, ubiquitous, broad network access, pool of sharing and measured services that can be rapidly provisioned and released with the minimal management efforts or service provider interaction. In the existing IT infrastructure or traditional solutions in which IT services are remain in the physical, logical, personal controls where cloud computing refers to the applications and services run on the distributed network using virtualized resources and accessed by the common internet protocol and networking standards. In this paper focuses on the data integrity which is most challenging part for accurate data store .To ensure the exactness data of the cloud client in the cloud , we proposed some cloud services for the data integrity and data security and access control when cloud clients data outsource sensitive data for sharing on cloud server. In order to address this is new challenge problem and later achieves a secure and dependable cloud storage services. In this proposed system, our scheme achieves the integration of storage correctness verification and error localization that is the identification of the misdemeaning .For this extensive data integrity and analysis shows that the proposed scheme is highly efficient and resilient against the internal and external attacks.**

***Keywords*— Data security, TPA audit ,cloud client, cloud server.**

## I.INTRODUCTION

Cloud computing is the type of computing that relies on sharing computing resources rather than having local server or personal devices to handle applications. In the cloud computing ,the name of cloud it refers to the internet so in short cloud computing means this is the internet based computing. In the cloud computing uses most of processors combine with software as a service in the computing architecture. These are transforming data centres into shared pools of computing service on a large scale. Combining a set of existing and new technology from research areas such as virtualization. In the cloud computing infrastructure resources are provided as services over the Internet. The increasing number of online services such as Amazon, Yahoo!, Google, gmail, etc. These services are most important for storing and maintaining lots of valuable user data. For example uses of this storage include online backup, email, audio, video and users personal data. Transferring data into the cloud and it offers more convenience to since they don't have any care about the complexities of hardware management. There are two well known examples of cloud computing vendors that is amazon simple storage services and amazon elastic compute cloud. This cloud computing is internet based computing provides large amount of storage space and customizable cloud computing resources ,but at the same time it eliminates the responsibility of local machines for data maintenance .For the result cloud user providers that is CSP for data integrity[3][4].On the one side, infrastructure as a service of cloud computing are more powerful than the personal computing, extensive range of both internal and external attacks for data integrity still exits. For example cloud storage services appear outsource of sensitive data from time[5][6],but on the other side cloud user may not keep any local copy of outsource sensitive data, but there are various reasons for cloud service provider to behave faithlessly towards the cloud client regarding of their outsource data.

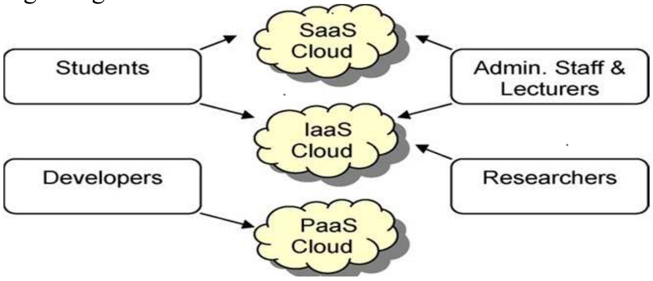

Figure1. Simple example university with an information technology infrastructure that caters for the needs of students, teaching staff and management, research staff and software developers

In above example given, For demonstrating how those services can be used and the processes involved in their utilization. For example, we are taking example a typical university with an information technology infrastructure that caters for the needs of students, teaching staff and management, research staff and software developers As shown in Fig demand for information technology services in this environment is directed shows to the information technology services Department whose job is to provide students and staff with software and provide researchers and postgraduate students with the required special software and hardware to run experiments that are likely to involve a great deal of processing and computation and provide Web developers with the development tools needed to write and host Web applications.

**Advantages**

1. Lower cost needed for data processing services when compared with the previous system of maintaining software and its associated hardware on an internal system.
2. Easily Remote Accessibility.
3. Environmentally Friendly.
4. Better Security.
5. Less time required

**Disadvantages**

1. Additional cost of data transfer fees.
2. Do not have control over the remote servers or their security.
3. It may be impossible to move large amounts of data from the provider.

**Characteristics of the cloud computing:**

1) On demand self services:In the on demand self services cloud client access cloud services through online control panel and manage their own computing resources.
2) Broad network access: In this broad network access we can easily access services anywhere, anytime.
3) Resource pooling: It means that customer draw from a pool of computing resources usually in the remote data centre.
4) Rapid elasticity: In this rapid elasticity means that it is computing terms for ability to provide scalability services.
5) Measured Services: In this measured services, aspect of the cloud services are controlled by the cloud provider.

**Cloud Computing services:**

- Software as a services: In this software as a services user can access both resource and applications .
- Infrastructure as a services: In this infrastructure as a services cloud client completely outsources the storage and resources such as hardware and software that they need
- Platform as a services: In this platform as a services cloud client give subscriber access to the components that they require to develop and operate application over the internet

## II . PROBLEM STATEMENT

In the problem statement, there are two models are as follows : System model and adversary model. A descriptive cloud storage service network architecture is shown in figure. 1.System Model:

In system1 model, there are three different entities can be identified as follows:

1. Cloud Client : Cloud client has data to be stored in the cloud. It relies on the cloud for data storage and computation. Cloud client can be either enterprise or individual customers.

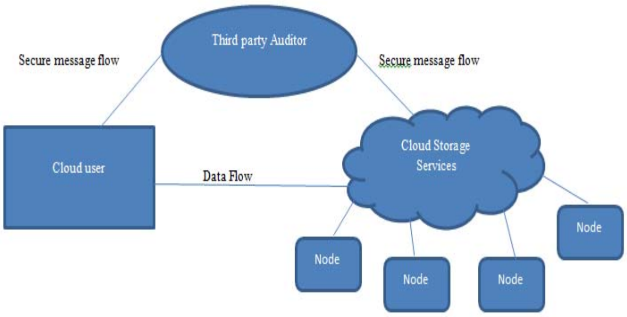

Figure2. Cloud storage services architecture

2. Cloud server: This cloud server managed by cloud service provider to provide data storage service. It has significant storage space and computation resources.
3. Third-Party Auditor: This TPA is an optional. This third party auditor protects from the byzantine failures, this byzantine failure means hide that errors from the cloud client for its own purpose. Third party auditor has expertise and capabilities that users may not have trusted to assess and expose risk of cloud storage services on behalf of the cloud client upon request

**Adversary Model:** In the adversary model , from cloud client viewpoint this model has to capture all types of attacks toward his cloud data integrity. For reason cloud data do not reside at cloud client local site but at cloud storage provider address domain .These attack can come from two different sources:

a) Internal attacks. In internal attacks, a cloud storage provider can be self-interested, malicious. In this internal attack malicious employees at client, malicious employees at cloud provider, cloud provider itself. For example malicious attack ,it is an active attack.
b) In external attacks: In this external attack, data integrity attacks may come from outsiders who are beyond the control domain of cloud service provider. For example network attack, it is an passive attack.

So there is need to make an effective and flexible distributed techniques which will explicitly support dynamic data including various operation on data block such append , delete, modify etc. effective and distributed scheme with explicit changeable data support to sure the correctness of cloud client data in the cloud server. This proposed work is not only sure correcting code in the file distribution preparation but also to support the guarantee of data dependability and redundancies. This structure significantly removes the communication and storage overhead as compared to the traditional duplication based file distribution techniques. By using the homomorphic token with distributed verification of not sure coded data. The scheme achieves the storage correctness verification. as well as data error localization whenever data corruption has been detected during the storage correctness verification. This scheme can almost guarantee the simultaneous localization of data errors as well as the identification of the misbehaving servers. Benefits of the cloud computing are as follows:

a) Scalabilty: Scalability means cloud computing offers unlimited processing power storage capacity.
b) Reliablity: Reliabilty means that it enable access to application and document anywhere in the word via the internet.
c) Efficient :Efficient means it allows the organization to free up

In this system, It's not been protected from the internal attacks. So here i am applying slicing approach in which each data stored in the sliced to provide data integrity it means that every data is uploaded and it will sliced on user attribute and stored so secured from internal attack . In order to implement the quality of cloud storage services and successed the assurance of cloud data capable methods that allow on-demand data correctness verification on behalf of cloud client have to be designed. In fact, in the long time cloud client data in the cloud is not Prohibits the direct approval of traditional cryptographic primitives for the intension of data integrity protection.so therefore the verification of cloud storage exactness must be conducted without clear knowledge of the complete data files[8][9].Sometimes this cloud storage is not only a third party data accessed but also regularly updated by the cloud client including dynamic operations like deletion modification insertion and appending etc.

This cloud computing is internet based computing provides large amounr of storage space and customizable cloud computing resources,but at the same time it eliminates tehe responsibility of local machines for data maintenance.For the result cloud user providers that is CSP for data integrity[4].On the one side,infrastructure as a service of cloud computing are more powerful than the personal computing, extensive range of both internal and external attacks for data integrity still exits. For example cloud storage services appear outsource of sensitive data from time[5],but on the other side cloud user may not keep any local copy of outsource sensitive data, but there are various reasons for cloud service provider to behave faithlessly towards the cloud client regarding of their outsource data.

## III Exisiting Algorithms

### Notation and Preliminaries:

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

1. B – The dispersal matrix used for Reed-Solomon coding.
2. L – 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 de- notes that each Fi is represented as a column 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 B and it 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 B can be written as 1 0 . . . 0 P11 P12 . . .P1k 0 1 . . .0 P21 P22 . . . Pmk
B = ( I|P )= . . . . 0 0 . . . 1 Pm1 Pm2 . . . Pmk

By multiplying **F** by **B**, the user obtains the encoded file: L = F · B = (L(1) , L(2) , . . . , L(m) , L(m+1) , . . . , L(n) ) = (F1, F2 , . . . , Fm , L(m+1) , . . . , L(n) ), Where L(j) = ( l1(j), l2(j), . . . , ll(j) )T ( j ∈ {1,…,n }). noticed, the multiplication reproduces the inventive data file vectors of **F** and the remaining part (L(m+1) , . . . , L(n) ) are k parity vectors generated based on **F**;

### Algorithm 1:Token Precomputation:

Before file distributing cloud clients precomputes a assured number of short verification tokens on individual vector.Cloud client wants to make sure storage exactnessfor the data in the cloud,cloud client challenges the cloud servers with a set of randomly generated block indices.

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

### Algorithm 2:Correctness Verification and Error Localization:

Mixes the exactness verification and error localization using challenge response protocol. The response values from cloud servers for every challenge not only consider the exactness of the distributed storage but also contain information to locate potential data errors. The comparison between precomputed tokens and received response values can assurance the identification of misbehaving servers when the data is corrupted. The cloud client can reconstruct the original file by downloading the data vectors from the first m servers, supposing that they return the correct response values.

Algorithm 2 Correctness Verification and Error Localization:

```
procedure Challenge (i)
    ecompute αi = fkchal (i) and kprp from KPRP
    Send { αi, k( i )prp } to all the cloud server
    Receive from servers: {Ri( j ) = Σrq=1 αi ∗ L(j) [φ (i)(q)]|1 ≤ j ≤n}
    **for** (j ← m + 1, n) **do**
        R( j ) ← R( j ) -Σrq=1 ƒk j (sIq ,j ) · αi , Iq = φkprp(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 ← 1, n) do
            if (Ri ( j ) ! = vi( j ) then
                **return** server j is misbehaving.
            **end if**
        end for
    end if
end procedure
```

## IV CONCLUSIONS

In this paper, we have noticed several problems related to data integrity in cloud data storage, which is necessarily in a distributed storage systems. To get the exactness of cloud data integrity and impose the quality of dependable cloud storage services for cloud client. So there is need to make an effective and flexible distributed techniques which will explicitly support dynamic data including various operation on data block such append , delete.By using the homomorphic token with distributed verification of erasure-coded data, our scheme reaches the integration of storage exactness insurance and data error localization, that is whenever data fraud has been detected during the storage correctness verification across the distributed servers, we can almost assurance the concurrent credentials of the misbehaving servers.

## REFERENCES

1. C. Wang, Q. Wang, K. Ren, and W. Lou, "Ensuring Data Storage Security in Cloud Computing," Proc. 17th Int'l Workshop Quality of Service (IWQoS '09), pp. 1-9, July 2009
2. C. Wang, K. Ren, W. Lou, and J. Li, "Towards Publicly Auditable Secure Cloud Data Storage Services," IEEE Network Magazine, vol. 24, no. 4, pp. 19-24, July/Aug. 2010.
3. A. Juels and B.S. Kaliski Jr., "PORs: Proofs of Retrievability for Large Files," Proc. 14th ACM Conf. Computer and Comm. Security (CCS '07), pp. 584-597, Oct. 2007.
4. Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing," IEEE Trans. Parallel and Distributed Systems, vol. 22, no. 5, pp. 847-859, 2011.
5. K.D. Bowers, A. Juels, and A. Oprea, "HAIL: A High-Availability and Integrity Layer for Cloud Storage," Proc. ACM Conf. Computer and Comm. Security (CCS '09), pp. 187-198, 2009
6. C. Wang, S.S.M. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage," IEEE Trans. Computers, preprint, 2012, doi:10.1109/TC.2011.245
7. M.A. Shah, M. Baker, J.C. Mogul, and R. Swaminathan, "Auditing to Keep Online Storage Services Honest," Proc. 11th USENIX Workshop Hot Topics in Operating Systems (HotOS '07),pp. 1-6, 2007.
8. M.A. Shah, R. Swaminathan, and M. Baker, "Privacy-Preserving Audit and Extraction of Digital Contents," Cryptology ePrint Archive, Report 2008/186, http://eprint.iacr.org, 2008.
9. T. Schwarz and E.L. Miller, "Store, Forget, and Check: Using Algebraic Signatures to Check Remotely Administered Storage," Proc. IEEE Int'l Conf. Distributed Computing Systems (ICDCS '06), pp. 12-12, 2006.