Review in Cloud Computing Data Integrity

Anchal Pokharana¹ and Rahul Hada²

Department of Computer Science, Poornima University, Jaipur-INDIA
Email: anchalpokharana258@gmail.com, Rahul.hada@poornima.edu.in

Abstract

The cloud computing is internet based computing, in which large groups of remote servers are networked so as to allow sharing of data processing tasks, centralized data storage, and online access to computer services or resources. The prevalent problem associated with Cloud Computing is the Cloud security. Clouds typically have single security architecture but have many customers with different demands and I attempt to solve this problem. In this I need to provide integrity of data by overcoming many existing problems like Client Data Integrity, Cloud Storage Integrity, Integrity Problem on Transmitted Data, Integrity Problem When Update the Client Data, Integrity among VM Instances and Integrity of Cloud Services which is mostly need to enhance of Integrity of Cloud Data to get better data accessibility over network. To learn about Data Integrity in cloud computing, 40 research papers have been reviewed, which were published in the period of year 2010 to year 2013. The outcome of the review was in the form of various findings, found under the key issue. The findings included algorithms and methodologies used to solve particular research problem, along with their strengths and weaknesses and the scope for the future work in the area.

1 INTRODUCTION

Cloud Computing simply means internet based computing. It allows user to store large amount of data in cloud storage and use as and when required from any part of the world via any terminal equipment. Since cloud computing is rest on internet. It implies sharing of computing resources to handle applications. Cloud computing offers reduced capital expenditure, operational risks, complexity, maintenance and increased scalability while providing services at different abstraction levels.

Since cloud computing is a utility available on net so it brings about not only convenience and efficiency problems but also great challenges in the field of data security and privacy protection and many more like: data theft and leakage, Data confidentiality, Integrity Verification, authentication various hackers attacks are raised. Cloud computing is a great change of information system. Security becomes a bottleneck of cloud computing development ensuring the security has been regarded as one of the greatest problems in the development of cloud computing. Cloud computing is a latest and fast growing technology that offers an innovative, efficient and scalable business model for organizations to adopt various information technology resources like software, hardware, network, storage, bandwidth etc. At the foundation of cloud computing is the broader concept of converged infrastructure and shared services. It has the capability to incorporate multiple internal and external cloud services together to provide high interoperability there can be multiple accounts associated with a single or multiple service provider (SPs). So, Security in terms of integrity is most important aspects in cloud computing environment.

Data integrity refers to maintaining and assuring the accuracy and consistency of data over its entire life cycle. Data integrity can be compromised in a number of ways:

- Human errors when data is entered
- Errors that occur when data is transmitted from one computer to another
- Software bugs or viruses
- Hardware malfunctions, such as disk crashes
- Natural disasters, such as fires and floods

There are many ways to minimize these threats to data integrity. These include:

- Backing up regularly
- Controlling access to data via security mechanisms
- Designing user interfaces that prevent the input of invalid data
- Using error detection and correction software when transmitting data

2 VARIOUS ISSUES IN THE AREA

After reviewing 40 research papers on Cloud Computing Data Integrity we have found following issues, which has to be addressed, while the designing and implementation of the Cloud Computing these issues are:

- Client Data Integrity
- Cloud Storage Integrity
- Integrity Problem on Transmitted Data.
- Integrity Problem When Update the Client Data.
- Data Integrity among VM Instances.
- Integrity of Cloud Services

Keywords: Data Integrity, Cloud Storage, Virtual Instances, privacy.
3 ISSUE WISE DISCUSSION

Issue 1: Client Data Integrity
For the first issue many of the solution approaches are used like data anonymization scheme, an architecture based on CSPs and PDP and A service-oriented solution, called TrustStore, all of these solution approaches are proved by experimental results and algorithms.

Issue 2: Cloud Storage Integrity
For the second issue many of the approaches are used to provide cloud storage integrity like An attack model with transformation technique , Result Verification Mechanism for MapReduce Computation Integrity , A Probabilistic Integrity Checking Approach for Dynamic Data (data blocks checking), Cloud Storage Provable Data Possession(CSPDP) Scheme , A new data integrity check scheme, based on the well-known RSA security assumption , RSA based storage security (RSASS) method, and A remote data integrity checking protocol based on HLA and RSA signature. Research defend the solution by providing proposed architecture and graphical representation.

Issue 3: Integrity Problem on Transmitted Data
For the third issue Blind processing service using trusted computing mechanisms and TVMCM, Trusted VM Clone Model are used to solve integrity problem on transmitted data. whenever data is moving from source to destination these approaches are used to maintain the integrity of data.

Issue 4: Integrity Problem When Update the Client Data
For the fourth issue some solution approaches like A Dynamic Proof of Retrievability (PoR) Scheme (BLS , CMB+T), A Flexible Distributed Storage Integrity Auditing Mechanism (KeyGen, SigGen, GenProof, VerifyProof) And A novel MapReduce framework, Cross Cloud MapReduce (CCMR) are used. Research defend the solution by providing experimental results and graphical representation.

Issue 5: Data Integrity among VM Instances
For this issue a light Weight Centralized File Monitoring Approach (Cryptography) and A Fingerprinting System Call Approach are used to maintain data integrity among VM Instance. These approaches prevents malicious VM users to modify well known frequently executed programs.

Issue 6: Data Integrity of Cloud Services
For sixth issue some solution approaches are used like A Dynamic Proof of Retrievability (PoR) Scheme, security architecture for web-enabled HLA/RTI in the cloud simulation environment, A service integrity assurance framework and an energy-efficient protocol. These solution approaches provides the integrity of cloud services and these are defend by graphical representation and algorithms.

4 ISSUE WISE SOLUTION APPROACHES USED
The solution approaches under the various issues have been shown in the table 1 to 6, which includes solution approaches along with results obtained. The same table also describes the Comparative analysis between various solution approaches.

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Solution Approach</th>
<th>Results</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>A service-oriented solution, called TrustStore</td>
<td>Solve the risk of data Exposure , data tampering, denial of access to data by third parties on the cloud or by the cloud provider , allows the system to store large data volumes cheaply with public storage providers</td>
<td>[33]</td>
</tr>
<tr>
<td>3.</td>
<td>Data Anonymization scheme</td>
<td>It saves the computational power at the client</td>
<td>[6]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Solution Approach</th>
<th>Results</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.</td>
<td>An attack model with transformation technique</td>
<td>Excellent performance against attack for the data privacy with low communication cost.</td>
<td>[20]</td>
</tr>
<tr>
<td>5.</td>
<td>Result Verification Mechanism for MapReduce Computation Integrity</td>
<td>Applicable in large parallel data analysis such as astrophysics, bio-informatics high energy physics, machine learning, genomics and cyber-security, extremely large amount of computations with reduced time and cost of computation</td>
<td>[26]</td>
</tr>
</tbody>
</table>
A Probabilistic Integrity Checking Approach for Dynamic Data (data blocks checking) | Less computation and Bandwidth reduces the processing time for both client and cloud. no need to download the whole file. | [5]
---|---|---
Cloud Storage Provable Data Possession (CSPDP) Scheme | lightweight, efficient and robust Provide Data Integrity with Data Security | [4]
A new data integrity check scheme, based on the well-known RSA security assumption. | Relieves the storage burden in client. | [38]
A Wew Remote Data Integrity Checking Scheme | Provides integrate correct checking, dynamic update and privacy preserving all together. | [31]
RSA based storage security (RSASS) method | data integrity verification not only for static data but also for dynamic data that support Modification, Insertion and Deletion Operation | [7]
A remote data integrity checking protocol based on HLA and RSA signature | support data dynamics, including block level operations of modification, deletion and insertion | [36]

<table>
<thead>
<tr>
<th>S.No</th>
<th>Solution Approach</th>
<th>Results</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Blind processing service using trusted computing mechanisms</td>
<td>A secure communication channel between dedicated processes for transfer data</td>
<td>[35]</td>
</tr>
<tr>
<td>13</td>
<td>TVMCM, Trusted VM Clone Model</td>
<td>Identities verification of involved servers attestation of source VM and destination VM protection of integrity of transmitted data. capable of working with the Xen hypervisor</td>
<td>[3]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S.No</th>
<th>Solution Approach</th>
<th>Results</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>A Dynamic Proof of Retrievability (PoR) Scheme (BLS, CMB+T)</td>
<td>can detect static data corruption ,data update, Public verification together.</td>
<td>[17]</td>
</tr>
<tr>
<td>15</td>
<td>A Flexible Distributed Storage Integrity Auditing Mechanism (KeyGen, SigGen, GenProof, VerifyProof)</td>
<td>avoidance of expenditure on hardware, software, and personnel maintenances. Support scalable and efficient public auditing .Provides a privacy-preserving auditing process</td>
<td>[2]</td>
</tr>
<tr>
<td>16</td>
<td>A novel MapReduce framework, Cross Cloud MapReduce (CCMR),</td>
<td>CCMR can guarantee at least 99.52% of accuracy with 33.6% of overhead when replication probability is 0.3 and the credit threshold is 50.</td>
<td>[8]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S.No</th>
<th>Solution Approach</th>
<th>Results</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>A light Weight Centralized File Monitoring Approach (Cryptography)</td>
<td>Avoiding the need of external database to store the checksum of files. minimum resources needed requires minimal and lightweight Standalone utility application, transparent and platform independent</td>
<td>[23]</td>
</tr>
<tr>
<td>18</td>
<td>A Fingerprinting System Call Approach</td>
<td>Prevents malicious VM users to modify well known frequently executed programs</td>
<td>[21]</td>
</tr>
</tbody>
</table>

5 COMMON FINDINGS

**Issue 1:- Client Data Integrity**

- In the First Issue The Best Approach is An architecture based on CSPs and PDP because of computing speed and efficiency. In this scheme existence of multiple CSPs to cooperatively store and maintain the clients data.
The Worst Approach is truststore because there is a risk of data exposure, data tampering.

**Issue 2: Cloud Storage Integrity**

- In second Issue the best approach is A Probabilistic Integrity Checking Approach for Dynamic Data in Untrusted Cloud Storage. Because this provides supports Dynamic data operations by storing small additional information on the cloud server. less computation and Bandwidth and reduces the processing time for both client and cloud.
- The worst Approach is on-demand logical resource (file) replication scheme, because FRS is a single point of failure in a group.

**Issue 3: Integrity Problem on Transmitted Data.**

- In third Issue Best Approach is TVMCM because three problems are resolved, the local server is able to verify the identity of the remote server before the cloning process begins, the newly generated destiny VM can attest itself to the source VM to be cloned, and the integrity of the transfer data is protected.
- Worst Approach is blind processing service using trusted computing mechanisms.

**Issue 4: Integrity Problem When Update the Client Data.**

- In Fourth Issue The Best Approach is A Flexible Distributed Storage Integrity Auditing Mechanism because The authors defend the solution by providing four algorithms (KeyGen, SigGen, GenProof, VerifyProof).
- The Worst Approach is A Dynamic Proof of Retrievability (PoR) Scheme because it can detect only static data corruption.

**Issue 5: Data Integrity among VM Instances**

- In Fifth Issue The best approach is Fingerprinting System Call Approach for Intrusion Detection because it creates individual array for each system call and Low Complexity, Scalability and adaptability are the advantages.
- The worst Approach is light Weight Centralized File Monitoring Scheme because this scheme is not Deployed yet.

**Issue 6: Integrity of Cloud Services.**

- In Sixth Issue The best Approach is TPM, TNA and trusted measurement system, because it can not only ensure security of both client and server, but also integrity of provided cloud service. And defend the solution by providing the encryption and decryption scheme.

## 6 Scope For the Work In Area

- Three tier security Framework, The future target is to design scheduling Algorithms, to provide security to data and application and optimize the performance of the framework.
- A light Weight Centralized File Monitoring Approach, The deployment of the scheme and further analysis on large scale public cloud is under progress.
- Future Scope of Fingerprinting System Call Approach for Intrusion Detection is for Commercial Cloud Environment.
- Enhance security of data in cloud we can implement various method with RSA.
- New combination of different method with cryptography technique enhance the security of cloud computing.
- VM cloning technology has a brilliant perspective of future other kind cloud computing such as Hadoop
- Protect data through the unsecure networks like the Internet; using various types of data protection is necessary.
- Cloud computing moves the application software and databases to the large data centers, where the management of the data and services may not be fully trustworthy. This unique feature however raises many new security challenges which have not been well understood.
- Need in cloud computing, both data and software are fully not contained on the user's computer; Data Security concerns arising because both user data and program are residing in Provider Premises.
- Clouds typically have single security architecture but have many customers with different demands. Every cloud provider solves this problem by encrypting the data by using encryption algorithms. So we can create a new and efficient security model.

## 7 Conclusion

The review of 40 research papers has been carried out in the area of Data Integrity in Cloud Computing to investigate and find out current challenges and scope of work. After the review, 6 issues have been found namely, Client Data Integrity, Cloud Storage Integrity, Integrity Problem on Transmitted Data, Integrity Problem When Update the Client Data, Integrity among VM Instances, Integrity of Cloud Services. And many of the solution approaches are used to maintain integrity of cloud. The exhaustive review has finally led to extract findings, issue wise findings and common findings in the area of Data Integrity in
Cloud Computing, strengths and weaknesses and also the gaps in the published research work. Virtualization is the core of Cloud Computing, to maintain the integrity of Cloud Storage we can provide the integrity on Virtual Machines and it would be a big effect on cloud storage integrity.

ACKNOWLEDGMENT

I would like to express my deep gratitude and thanks to Prof. Mahesh Bundele (Coordinator, Research), Poornima University for giving me an opportunity to work under his guidance for review of research papers and his consistent motivation & direction in this regard. I would also express my sincere thanks to Mr. Rahul Hada (Associate Professor, CE), Poornima University for their guidance and support.

REFERENCES

2. SALMA T.J. (2013), A Flexible Distributed Storage Integrity Auditing Mechanism in Cloud Computing, IEEE.
4. HuiFeng Wang, Zhanhuai Li, Xiaoan Zhao, Chanying Qi, Qiu Hu and Jian Sun (2013), Ascheme to ensure data security of cloud storage , IEEE,978-1-4799-0530-0/13, 79-82.
27. D.Sureshraj & Dr.V.Murali Bhaskaran(2012), Automatic Dna Sequence GenerationFor Secured Cost-Effective Multi -Cloud Storage, IEEE.
31. Xiangtao Yan, Yifa Li (2012), A New Remote Data Integrity Checking Scheme forCloud storage With Privacy Preserving, National High-tech R&D Program of China, 704-708.

Author(s) Profile

She has done B.tech from Institute of Technology & Management, Bhilwara in the year 2011, and having teaching experience of 2.5 years and industry experience of 9 Months. She has published one paper in “International Organization of Scientific Research” Journal and presented a paper in “National Conference on Mathematical Analysis and Computation” at MNIT Jaipur. She has attended an (IEEE) International Conference on Recent Advances and Innovations in Engineering - ICRAIE 2014. She has also attended several national workshops on Research Methodologies, Neural Networks in MATLAB and has also attended an ISTE E-seminar on Steps 2 Research.

He has done M.Tech from MNIT, Jaipur and having teaching experience of 4.5 years, Industry experience of 18 Months and Research experience of 6 month. He has specialization in Open source, Ad-Hoc network, wireless sensor network, security and cloud computing. He attended a Workshop on Network programming, Exacta Qualnet, Intrusion Detection using Metasploit framework, Effective learning and teaching and learning computer programming. His M. Tech. thesis topic was on “simulation of wormhole attack using ns3”.

Review in Cloud Computing Data Integrity