Framework to Improve Data Integrity in Multi Cloud Environment

Anandita Singh Thakur
M.Tech, Department of CSE
Jaypee University of Information Technology
Waknaghat, Solan, Himachal Pradesh

P. K. Gupta
Assistant Professor (Sr. Grade)
Department of Computer Science and Engineering
Jaypee University of Information Technology
Waknaghat, Solan, Himachal Pradesh

ABSTRACT
Cloud computing is a network based environment where resources, information and software's are shared and provided to the users according to their demands. Though cloud computing is being widely used in various fields, it has certain drawbacks. Security issue is a concern in cloud computing. The confidentiality and data integrity should be guaranteed to the user who is putting their data over the cloud. This paper proposes a framework that would provide integrity of data of multiple users through Third Party Auditor (TPA) and proposes various algorithms to implement this framework. In the proposed framework concept of multi cloud has been used to provide best cost optimization for various requirements of user. The framework is also divided into three platforms namely: platinum for sensitive data storage, gold for less level of security and silver for least level of security on the data. Finally we have implemented different algorithms for the various platforms in the proposed framework. In the obtained results, we can find that the proposed framework is easy to implement and provides better performance by using the traditional algorithms of network security over the various issues of cloud computing.

General Terms
• CDIBA - Cloud data integrity backup agent defines set of rules for backup of cloud data in "cloud zone" using logical grouping of cloud components.

• CSPA - Cloud service provider agent provides security service tasks according to Service Level Agreements.

• CSP - Cloud service provider are the ones who have the resources and are expert in building and managing the distributed servers.

• CS - Cloud servers are managed by the cloud service providers.

• DSA - Digital signature algorithm. The authenticity of a digital message or document is checked using DSA.

• i.e. - in other words

• I/O - input and output

• MAS - Multi agent system. It is a technique where several agents communicate with each other.

• PDP - Prove able data possession. It checks that a file which consists of a collection of n blocks is retained by the outsourced storage site.

• POR - Proof of retrieve ability. It detects data corruption and achieves guarantee of file retrieve ability.

• QOS - Quality of service. It defines the degree to which a provided activity promotes customer satisfaction.

• RSA - Rivest, Shamir and Adleman. The most commonly used algorithm for encryption and authentication is RSA algorithm which is included as part of the Web browsers.

• TPA - Third party auditor. It verifies over the cloud server.

Keywords
AES, Bcrypt, Churning, Cloud computing, data integrity, RSA, security, third party auditor

1. INTRODUCTION
Cloud computing is an emerging trend in the field of technology where resources, software and information is shared [1]. In cloud computing the tasks from individual systems are moved into the cloud where multiple systems can interact with each other at a time. Cloud computing provides a pay-per-use facility i.e. pay for only what the customer uses. This would reduce the customer's expenditure on hardware, software and other services. Though there are various benefits of cloud computing like cost saving, scalability, reliability, maintenance still cloud computing has certain drawbacks [2] which have been shown in Figure 1 below .

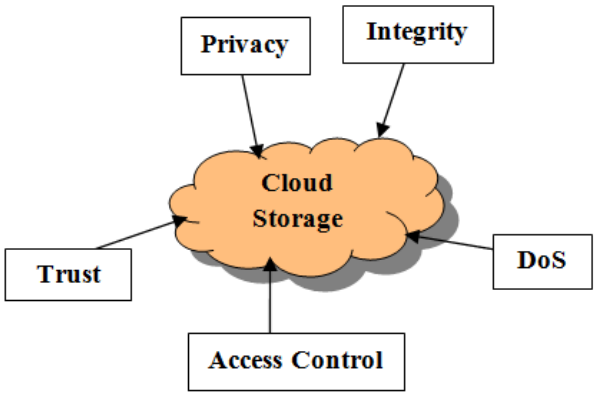


Fig 1: Drawbacks of cloud computing

As multiple users / parties access information on cloud, the integrity and privacy of the information stored is at risk. Cloud provides distribution of data over computers. When data is sent by the user to be processed in the cloud; the control of the data is given to a remote party that may not address security concerns of the user. As a user has no physical access to the data, he is unaware about the location of his data and is not sure whether the integrity of his data is maintained or compromised in cloud. It is important to ensure that the information being processed on cloud is secure and no

tampering of information is done when previously unknown parties may be present. A framework is proposed to provide data integrity using TPA to guarantee the various users that their data is unaltered.

The rest of the paper is classified as follows: Section 2 discusses the literature review, Section 3 describes the proposed model to provide integrity of data, Section 4 describes the algorithm used, Section 5 shows the analysis of the algorithms used and Section 6 presents the conclusion for the work done till now.

2. LITERATURE REVIEW

Cloud computing is an emerging trend in the field of technology. There are various issues related to cloud computing, major ones being the security and integrity of data. Many frameworks have been designed and many algorithms have been proposed to resolve such issues.

Nirmala et al. [2] proposed a new scheme to resolve integrity problem by introducing user authenticator to audit and check the integrity of data. Their research focused on providing solutions to all issues of cloud computing and to develop a model that would provide secure cloud infrastructure which would help to adopt the cloud as and when required. Raju et al. [3] introduced a protocol for integrity checking of cloud storage that would provide integrity protection of user information. This protocol supports public verifiability and is evidenced to be secure against associate un-trusted server. It's additionally non-public against third-party verifiers. Attas and Batrafi [4] proposed an integrity checking model over cloud with help of TPA using DSA algorithm. With the help of the model, user can examine and verify the data from unauthorized people who manipulate with the cloud or extract data. Evaluation of the model was done using Windows Azure project that involved digital signature coding. The results showed that the proposed model worked according to what was claimed. In [5] Pardeep Kumar et al. assess that how cloud providers can gain trust of their customers and provide them with security, privacy and reliability on the data when processing of sensitive data is done by the third party in remote machines located in various countries. Various services are made available to the user by using the concept of utility cloud. The advantage of proposed approach was to incorporate the trusted computing technology into the cloud computing environment to achieve trusted computing requirements for the cloud computing and then fulfill the trusted cloud computing. To resolve the problem of privacy in the clouds Metri and Sarote [6] introduced a threat model. Juels et al. [7] described a formal "proof of retrievability" (POR) model for ensuring the remote data integrity. It uses error correction code in order to check that the data is correct or not. Talib et al. [8] proposed layer architecture based on MAS architecture having two main layers: (a) cloud resource layer [cloud server side] (b) MAS architecture layer [cloud client side]. The MAS architecture has two agents namely Cloud service provider agent [CSPA] and Cloud data integrity backup agent [CDIBA]. This layered architecture collectively was called "Cloud Zone". A prototype of proposed "Cloud Zone" would be designed using Prometheus Methodology and implemented using the Java Agent Development Framework Security (JADE-S). Ateniese et al. [9] considered public audit ability in their defined "provable data possession" model to verify if the client's data is stored at untrusted server. Homomorphic Verifiable Tags are used for data auditing. The model samples random sets of blocks from the server and generates probabilistic proofs of possession which reduces I/O costs. The client verifies the proof by maintaining a constant amount of metadata. The transmission of a small, constant amount of data is done by the response protocol which minimizes network Communication. The PDP model for remote data checking supports large data sets in widely-distributed storage systems. Experimentally it was shown that the previous schemes failed to give any verification for possession of large data sets which has been provided by proposed scheme. In [10] P. Rana et al. proposed an architecture which combines Infrastructure as a service (IAAS) and Platform as a Service (PAAS) framework and remove the drawbacks of IAAS and PAAS. It describes how to simulate the cloud computing key techniques such as data storage technology (Google file system), data management technology Big Table as well as programming model and task scheduling framework using CLOUDSIM simulation tool. Bhosale et al. [11] provides a 3 dimensional framework along with digital signature and RSA algorithm where the user will upload the data over cloud based on the various security levels. Protection ring 1 will provide high level of security, ring 2 will provide less security and ring 3 will provide least level of security. Security of cloud is enhanced by using this framework with RSA and DSA algorithm combination.

3. PROPOSED FRAMEWORK

This section defines the proposed framework to provide data integrity in multi cloud system. Proposed framework is shown in Figure 2 and has following three main roles:

- ***Users***- who will store the data by selecting appropriate layer depending on the level of security needed for the data stored on cloud.
- ***Cloud service provider (CSP)*** - provides the storage of data service with flexible resources to keep the user data. The CSP manages cloud server (CS) which informs the user about the intrusion of data on cloud.
- ***Third party auditor (TPA)*** - verifies the cloud server and checks whether there is any manipulation of user data by the cloud server. It then sends a report to the user stating that the cloud server (CS) was trusted or not.

There are many cloud service providers and each of them provides different storage plan along with different QoS parameters so it becomes a tough task for users to keep moving their data from one cloud to another based on QoS and cost optimization [12]. In the proposed model concept of multi cloud is used to provide best cost optimization for various requirements of user. To give a clear design of the model, use of connectors is made labeled as a, b and c.

The usage of connectors has been made in order to provide better view of the model. It can clearly be seen that depending on type of data the user can move from one service provider to another. e.g.: the user a, user b, and user c can put their data over private, hybrid or public cloud respectively, depending on the security needed for the data stored. The same thing can be applied by other two users.

Depending on the type of data to be stored on various clouds, there are three main platforms in the model namely:

- ***Platinum***- sensitive data will be stored here like data related to transactions of atm, bank account information along with high level of security on the data. The data will be stored on private cloud.
- ***Gold***- data related to simple login on any page like facebook, ebooking and email login is stored. The

level of security needed is not that high. Security only on password is required.

- ***Silver***- data related to only simple browsing of sites, uploading of images, downloading of files like downloading of music files or images is stored. The level of security needed is the least.

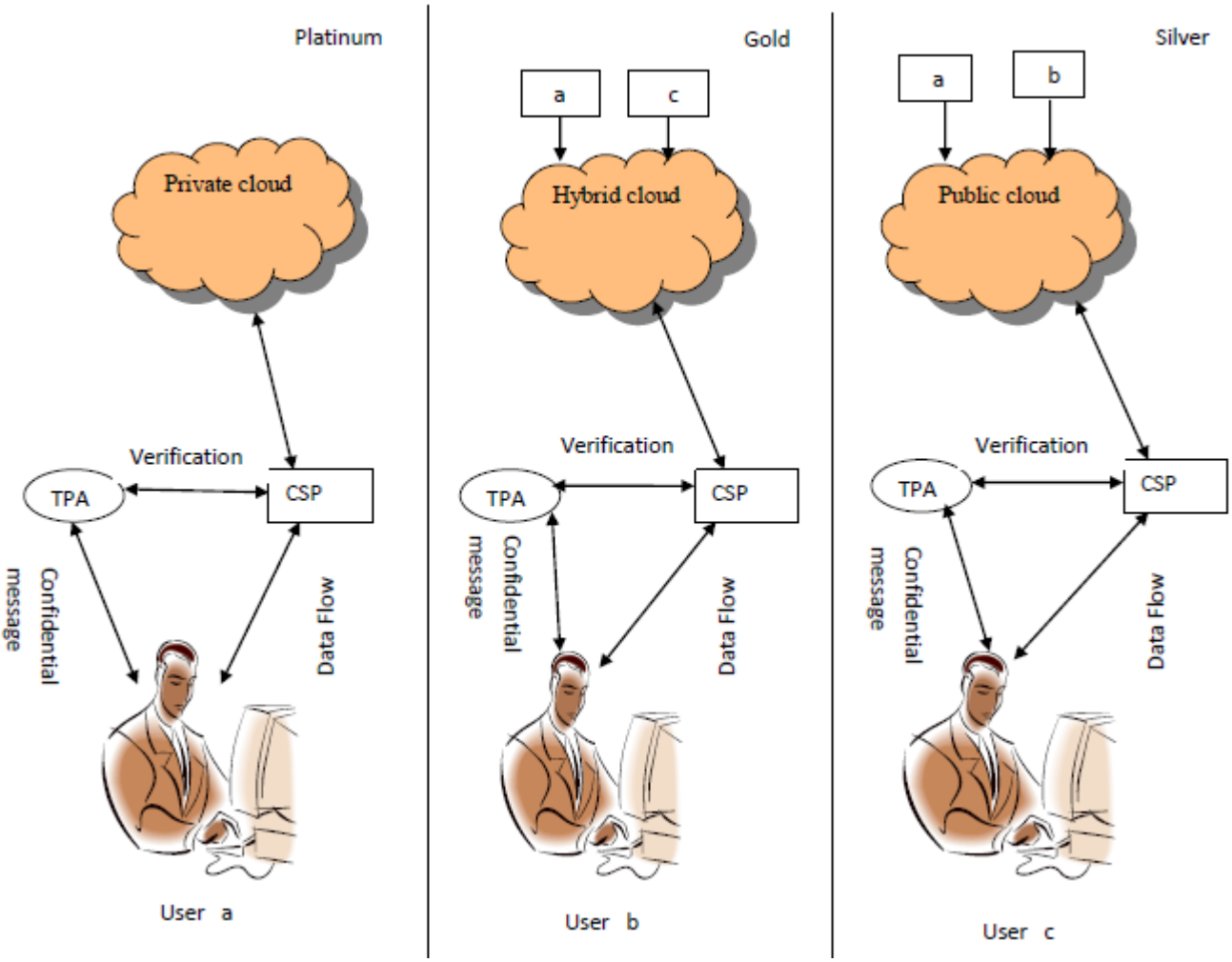


Fig 2: The proposed framework

4. PROPOSED ALGORITHM

In this section we have implemented various algorithms like RSA algorithm, Bcrypt algorithm, and AES algorithms to implement the proposed framework.

/* Variables used for:
User => u,
Platinum => p,
Gold => g,
Silver => s */

Begin

If *u chooses p*

Then *call module m1*

If *u chooses g*

Then *call module m2*

If *u chooses s*

Then *call module m3*

End

Module m1:

Begin

1. *User's data to be stored on cloud*
2. *Data is encrypted using RSA*
3. *Data is verified by CSP using RSA*
4. **If** *data is valid*

Go To Module T

Else

Corrupted data

End

Module m2:

Begin

1. Data stored on cloud

2. Data is encrypted

3. Data is verified by CSP using Bcrypt algorithm.

4. **If** *data is valid*

Go To Module T

Else

Intrusion on data

End

Module m3:

Begin

1. Data is stored on cloud

2. Data is encrypted

3. Verification of data is done by CSP using AES

4. **If** *data is valid*

Go To Module T

Else

Invalid data

End

Module T:

Begin

1. Check the data stored.

2. **If** *proof = direct then*

Report = direct access

Else

Return {1, 0}

1: if integrity of data is verified as correct

0: if integrity of data verified is incorrect

End

In module m1 RSA algorithm is used to provide integrity of data because for storing sensitive information on cloud, hashing algorithms are used. RSA is based on the difficulty of factoring large numbers. There are various advantages of RSA due to which it is preferred over DSA.

- DSA can only be used for authentication while RSA can be used for both authentication and to encrypt a message.
- A bad random number generator will leak DSA key bits.
- Faster at encrypting than DES.

In module m2 Bcrypt algorithm is used for hashing the passwords. A password hashing algorithm should preferably be slow in order to prevent brute force attacks; it should have features which actually decrease the feasibility of a distributed brute force attack on the hashes. The following hashing algorithms are not considered for the purpose:

- MD-5
- SHA-1
- SHA-2
- SHA-3

Bcrypt algorithm is derived from the Blowfish block cipher which uses look up tables that are initiated in memory to generate the hash.

In module m3 AES algorithm is used to provide security on the data stored. AES is asymmetric encryption algorithm in which to encrypt the message sender uses public key of receiver and its private key is used by receiver to decrypt the message.

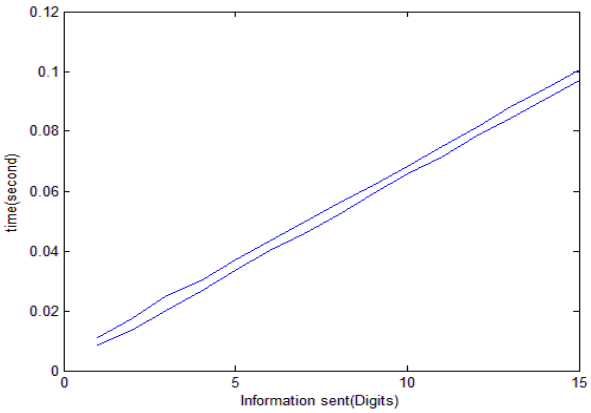
- AES is preferred over DES algorithm as it is more secure.
- AES data encryption is mathematically more efficient and elegant cryptographic algorithm. Key length option is the main strength of the algorithm. Time required to crack an encryption algorithm is directly related to the length of the key used to secure the communication. AES gives an option to choose a 128-bit, 192-bit or 256-bit key, making it exponentially stronger as compared to the 56-bit key of DES.
- Block size of DES is small compared to AES
- A balanced Feistel structure is used by DES while substitution-permutation is used by AES.

5. RESULT ANALYSIS

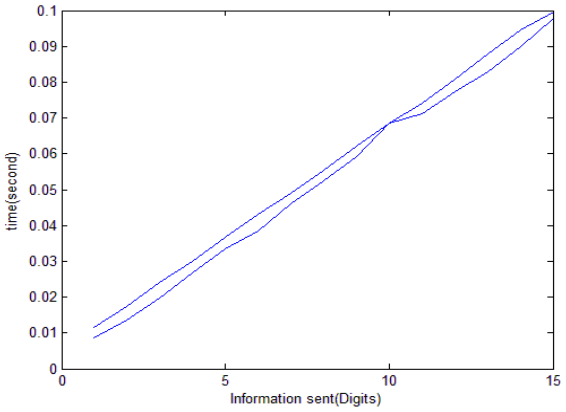
To estimate the output of the programs, matlab tool is used.

5.1 RSA algorithm analysis

RSA algorithm is tested for integer numbers ranging from a single digit message length to 16-digit message length. The execution time t is in seconds. The execution time depends on the values of p and q which are prime numbers. Different values of p and q are taken and depending on these values graph between message length and time are plotted as shown in Figure 3(a) and Figure 3(b).



(a)



(b)

Fig 3: Message length vs. Time for (a) p=3 and q=7 (b) p=23 and q=17

5.2 AES algorithm analysis

AES is used here to provide integrity to data while simple browsing of internet. Plain text is encrypted to hexa decimal format. The change in graph depends on the value of plain text. By using combination of alphabets and digits in a text the time taken increases. A graph between information sent and time is plotted which can be seen in the Figure 4.

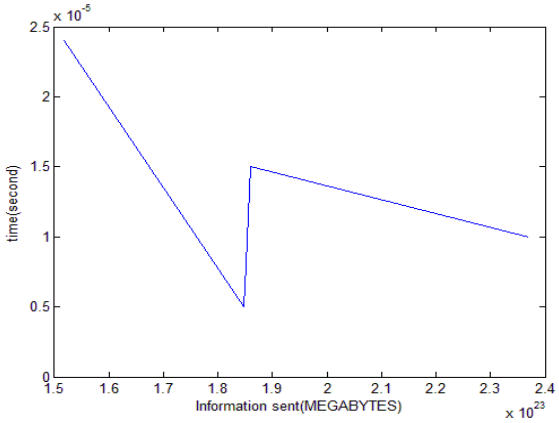


Fig 4: AES algorithm

6. CONCLUSION AND FUTURE WORK

There are many issues in cloud computing, one of them is integrity of data. Due to this issue, many users are apprehensive of using cloud technology as their data security is not guaranteed. Various frameworks have been proposed in order to resolve this issue but no framework had provided full security. In this paper proposed framework resolve the issue of integrity of user data with better performance using the traditional algorithms of network security. Cost is also optimized using multi cloud concept and different platforms for various categories of the users. Simple analysis of the algorithms selected by the users is done using mat lab tool which shows the better improvements in results. In our future work we will implement the hybrid algorithms using this framework and also find the churning effect when the user switches the clouds for their different kind of works. We will also explore the new opportunities for security in multi cloud environment.

7. REFERENCES

[1] Heena I. Syed, Naghma A. Baig "*Survey On Cloud Computing*," International Journal of Emerging Technology and Advanced Engineering, vol. 3, issue 4, April, 2013, pp. 308-312.

[2] Nirmala V., Sivanandhan R.K., and Lakshmi R.S. "*Data confidentiality and Integrity Verification using Authenticator scheme in cloud,"* Proc. of 2013 International Conference on Green High Performance Computing, Nagercoil, March 2013, pp. 1-5.

[3] Raju et al. "*Data Integrity using Encryption in Cloud Computing,"* Journal of Global Research in Computer Science, vol. 4, no. 5, pp. 40-43, May 2013.

[4] Dalia Attas and Omar Batrafi, "*Efficient Integrity checking technique for or securing client data in Cloud computing*," International Journal of Electrical & Computer Sciences, vol. 11, no 5, pp. 43-48, 20.

[5] Pardeep Kumar, Vivek Kumar Sehgal, Durg Singh Chauhan, P. K. Gupta, and Manoj Diwakar, "*Effective Ways of Secure, Private and Trusted Cloud Computing*," IJCSI International Journal of Computer Science Issues, vol. 8, issue 3, no. 2, pp. 412-421, May 2011

[6] Metri P. and Sarote G., "*Privacy Issues and Challenges in Cloud computing*," International Journal of Advanced Engineering Sciences and Technologies, vol. 5, no. 1, pp. 5-6, 2011.

[7] A. Juels and B. S. Kaliski *"PORs: Proofs of retrievability for large files*," Cryptology ePrint archive, June 2007. Report 2007/243.

[8] A. M. Talib, R. Atan, and R. Abdullah, "*CloudZone: Towards an Integrity Layer of Cloud Data Storage Based on Multi Agent System Architecture*," 2011 IEEE Conference on Open Systems (ICOS2011), September 25 - 28, 2011.

[9] Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Lea Kissner, Zachary Peterson, and Dawn Song "*Provable Data Possession at Untrusted Stores*," Proc. of the 14th ACM conference on computer and communications security, pp. 598-609, 2007.

[10] Poonam Rana, P. K. Gupta and Rajesh Siddavatam, "*Combined and Improved Framework of Infrastructure as a Service and Platform as a Service in Cloud Computing,"* Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28–30, 2012, Advances in Intelligent Systems and Computing 236, DOI: 10.1007/978-81-322-1602-5_89, Springer India 2014, pp. 1-8.

[11] Pradeep Bhosale, Priyanka Deshmukh, Girish Dimbar, and Ashwini Deshpande, "*Enhancing Data Security in Cloud Computing Using 3D Framework & Digital Signature with Encryption*," International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 8, October – 2012, pp. 1-8.

[12] S. B. Shivakumar, B. E. Ramesh, G. M. Kavitha, and M. Mala, "*Multi Cloud Architecture for Improved User Experience*," International Journal of Inventive Engineering and Sciences (IJIES), vol.1, issue 7, June 2013, pp. 14-17.