EMPOWER DATA PROTECTION AND DATA STORAGE IN CLOUD COMPUTING USING SECURE HASH ALGORITHM (SHA1)

A.William Walls
Research Scholar Department of Computer Science
SSM College of Arts and Science
Komarapalayam, India
williamwallsmsc@gmail.com

Mrs.N.Chandrakala
Head & Professor
Department of Computer Science
SSM College of Arts and Science
Komarapalayam, India
n chandrakala15@gmail.com

***Abstract*- Cloud computing has been enabling its users to access the shared resources through internet reduced cost resource utilization. The security for the data in cloud is very important. Data integrity refers to maintaining and assuring the accuracy and consistency of data over its entire life-cycle, and is a critical aspect to the design, implementation and usage of any system which stores, processes, or retrieves data. The term data integrity is broad in scope and may have widely different meanings depending on the specific context – even under the same general umbrella of computing. Fault tolerance and the integrity checking of data are difficult. The data are stored in multiple servers. Data integrity protection scheme is used for code regeneration. It is used to find the fault tolerance and repair traffic saving. Data integrity protection enables the client to verify the integrity of the outsourced data.**

Keywords- Data integrity, Retrieves data, Fault tolerance,

Repair traffic saving

I. INTRODUCTION

L.Ronald et al., [L.Ronald et al.,2010] say to cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources such as hardware and software that can be rapidly provisioned and released with minimal management effort or service provider interaction. Cloud computing is a technology that uses the internet and central remote server to maintain data and applications. Cloud computing allows users to use application without installation and access personal file at any computer with access internet. The cloud computing allows much more efficient centralizing data storage, processing and bandwidth. Cloud Computing has been envisioned as the definite and concerning solution to the rising storage costs of IT Enterprises Cloud computing is a scalable and managed infrastructure and payable as per its usage .The cloud computing technology has been evolved as business cloud models to provide computing infrastructure, data-storage, software applications, programming platforms and hardware as services.

A. Cloud Services

Sikder Sunbeam Islam et al., [Sikder Sunbeam Islam et al., 2012] say cloud computing can offer the services dealing with the data, software and computation through the internet and hence there are three services in cloud Figure A, Explain the cloud such as SaaS, PaaS and IaaS.

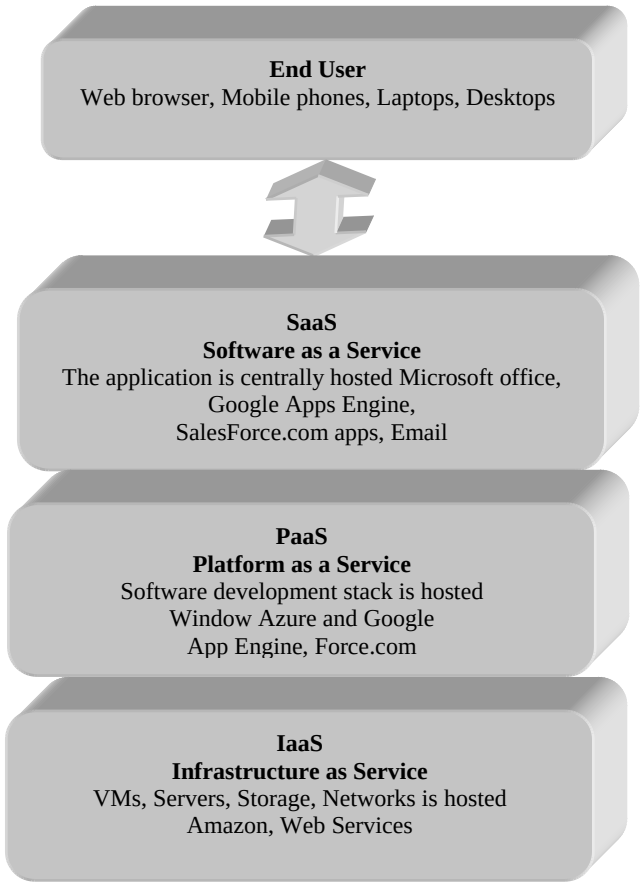


Figure A: Cloud Services

1. (SaaS) Software as a Service

SaaS Software as a Service provides the entire application as a service to the clients through the internet on demand. The user need not to bother about the hardware or software components needed to run the application. E-mail is a perfect example for SaaS. The customers of these services do not managed or control the underlying infrastructure and application platform only limited user-specific configurations are possible. Some of the SaaS vendors are SalesForce.com, Oracle and IBM service.

2. (PaaS) Platform as a Service

Platform as a Service provides a computing platform as a service to the user's .The entire software and hardware that the client needs to run an application will be offered as a services. Enable the client to concentrate on the application rather than the required Google Application Engine, Force.com, Microsoft Azure are few of the leading companies offering PaaS.

3. (IaaS) Infrastructure as a Service

Infrastructure as a Service provides to the computing and storage resources as per the requirements of clients. The clients do not have to purchase any servers or datacenters and they have to pay for the amount of time they use the resources. Example includes Amazon, Elastic Compute Cloud (EC2), and Microsoft Azure etc.

B. Cloud Models

Barrie Sosinsky et al., [Barrie Sosinsky et al, 2011] say based on the location where the cloud is hosted, we can classify clouds in to four type's private, public, hybrid and community cloud.

1. Public Cloud

The public cloud is a cloud computing deployment scheme that is generally open for use by the general public. The public cloud in as infrastructure used is owned by a cloud services vendor's organization. Examples of public cloud deployment vendor offerings include Amazon, Web Services, Google Application Engine, Salesforce.com, and Microsoft Windows Azure.

2. Private Cloud

The private cloud infrastructure is operated solely for one organization and is not shared with other organizations. The purpose of private cloud is to meet the internal computational needs within an organization. This cloud offers more security as it is implemented within the internal firewall. Every aspect of cloud implementation is fully controlled by the organization and hence security wills been enhanced.

3. Community Cloud

Community cloud shares infrastructure between several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted internally or externally.

4. Hybrid Cloud

Hybrid cloud is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound to gather, offering the benefits of multiple deployment models. For example, an organization might use a public Cloud service, such as Amazon Simple Storage Service (Amazon S3).

C Data Integrity In Cloud Computing

Data Integrity is an essential in databases similarly integrity of Data Storages is an essential in the cloud; it is a major factor that affects on the performance of the cloud. The data integrity provides the validity of the data, assuring the consistency or regularity of the data. It is the complete mechanism of writing of the data in a reliable manner to the persistent data storages which can be retrieved in the same format without any changes later. As described above, in cloud, the complete storage of data provided by the end-user is done at the data centers or data storages and the security and integrity of the data lies on the vendor storing data in the data centers but not the cloud hosts. Cloud Storage is gaining popularity for the outsourcing of day-to-day management of data. Therefore just storing data at cloud data storages or data centers doesn't ensure the integrity of data, but some mechanisms are to be implemented at each storage level to ensure the data integrity. Data Integrity is most important of all the security issues in cloud data storages because it not only ensures completeness of data but also ensures that the data is correct, accessible, consistent and of high quality.

II. PROBLEM DEFINITION

For checking the integrity of the huge amount of archived data the whole-file checking method becomes prohibitive. Proof of irretrievability and proof of data possession are used to verify the integrity of the large data. If the data is stored in the cloud means the data integrity confirmation is very important. It has more chance to modification in the data. Then we should repair the corrupted data and restore the original data. Store all the data in a single server is vulnerable to the single point-of-failure problem and vendor lock-ins. POR and PDP are used in the proposed for the single-server case. The Erasure coding has a lower storage overhead than replication under the same fault tolerance level. It results in permanent data loss in the server. These schemes can only provide the detection of corrupted data. MR-PDP and HAIL extend integrity checks to a multi server setting using replication and erasure coding. Time consumption for find the original data is very high.

III. LITERATURE REVIEW

A literature review can be just a simple summary of the sources, but it usually has an organizational pattern and combines both summary and synthesis. B. Chen et al., [B. Chen et al., 2010] say the RDC – NC, a novel secure and efficient RDC scheme for network coding based distributed storage systems. RDC – NC mitigates new attacks that stem from the underlying principle of network coding. B.Schroeder et al., [B.Schroeder et al., 2010] say the Latent sector errors

(LSE) refer to implemented and experimentally show that it is computationally inexpensive for both clients and servers.LSEs are a critical factor in data reliability, since a single LSE can load to data loss when encountered during RAID re-construction after a disk failure. LSEs happen at a significant rate in the welds, and are expected to grow more frequent with new drive technologies and increasing drive capacities. Y.Xiangtao et al., [Y.Xiangtao et al., 2012] say to an economic solution is looked upon so that the companies and organizations can focus on their businesses rather than on infrastructure. Well! The very concern of the presented literature review is regarding one aspect of cloud security that is named as data integrity maintenance especially in dynamic cloud environment. R.Saravana Kumar et al., [R.Saravana Kumar et al., 2011] say to developed an integrity checking scheme which gives a proof of data integrity in the cloud which could be utilized by the customer to check the correctness of data in the cloud. Yang Lie Wu et al., [Yang Lie Wu et al., 2012] say the method compressed effectively the check value for data integrity to reduce storage, and improved the check efficiency of multi-data objects. Zhou Hao, et al., [Zhou Hao et al., 2011] says a Remote data integrity checking mechanism that did not include any third party auditor. Data insertion, modification, and deletion at the block level, and also public verifiability were also promoted by this protocol. Wenjun Luo et al., [Wenjun Luo et al., 2011] say to addressed a remote data integrity checking protocol based on HLAs and RSA signature with the support public verifiability. Also, this very mechanism was very satisfactory of cloud storage systems. Q. Wang et al., [Q. Wang et al., 2011] say they concentrated on this very fact because earlier works on ensuring data integrity often lacks the support of either public dynamic operations data.

Henry C. H. Chen et al., [Henry C. H. Chen et al., 2011] say the protect outsourced data in cloud storage against corruptions, adding fault tolerance to cloud storage, along with efficient data integrity checking and recovery procedures, becomes critical. Regenerating codes provide fault tolerance data across multiple servers, while using less repair traffic than traditional erasure codes during failure recovery practical data integrity protection (DIP) scheme is designed and implemented for a specific regenerating code, while preserving its intrinsic properties of fault tolerance and repair-traffic is saving.

IV. PROPOSED METHOLOGY

First of all the client have to be created for use the cloud storage. Then the client selects the particular data that need to be store. Afterwards the data is splitting and then it is stored in the cloud servers. These data can be retrieved by the clients on any time. These data may contain the personal information of the client and splitter data stored in multiple servers in cloud. Each server in cloud contains different data of the clients. If any one of the server is failed means then the data in the server is lost.

The data in the failed server is not retrieved from that failed server. Finding the failure node is very important, because from this information the data can be retrieved. The failed server details are sending to the other servers in the cloud. Integrity to the data in cloud server is very important. Because of integrity the variations in the data are confirmed. Each data's integrity is confirmed by the Message Authentication Code the Secure Hash Algorithm (SHA1) is used. If the attacker made any change in the data means then the cloud server can find out this. This is because of the message authentication code. In repair operation the data in the failed server is retrieved by FMSR.It is retrieved by get the same data from the other servers. The data is stored in multiple servers in cloud. The whole data is not stored in multiple servers. The splitting data is stored in multiple servers in cloud. The integrity of the data is also confirmed by the server and provides the data to the client who needs the data.

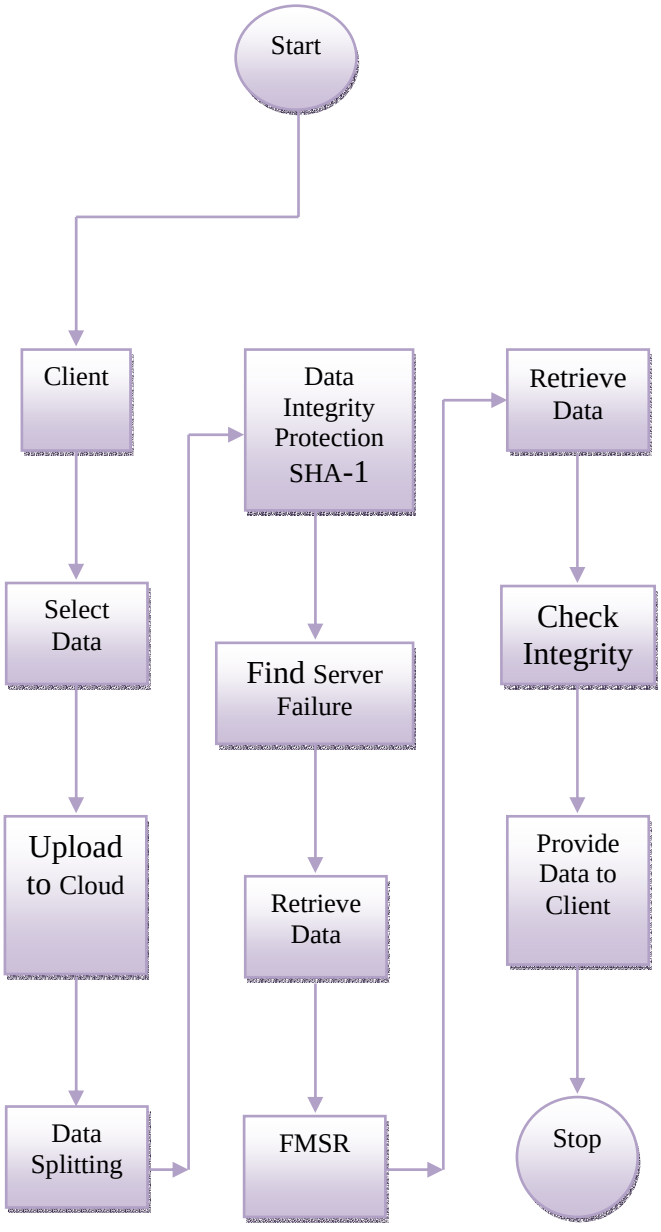


Figure IV: Flow Diagram

The figure IV deals the flow diagram proposed methodology such as Data Sharing In Cloud, Find Failure Servers, Repair Operation and Data Integrity Protection.

1. Data Sharing In Cloud

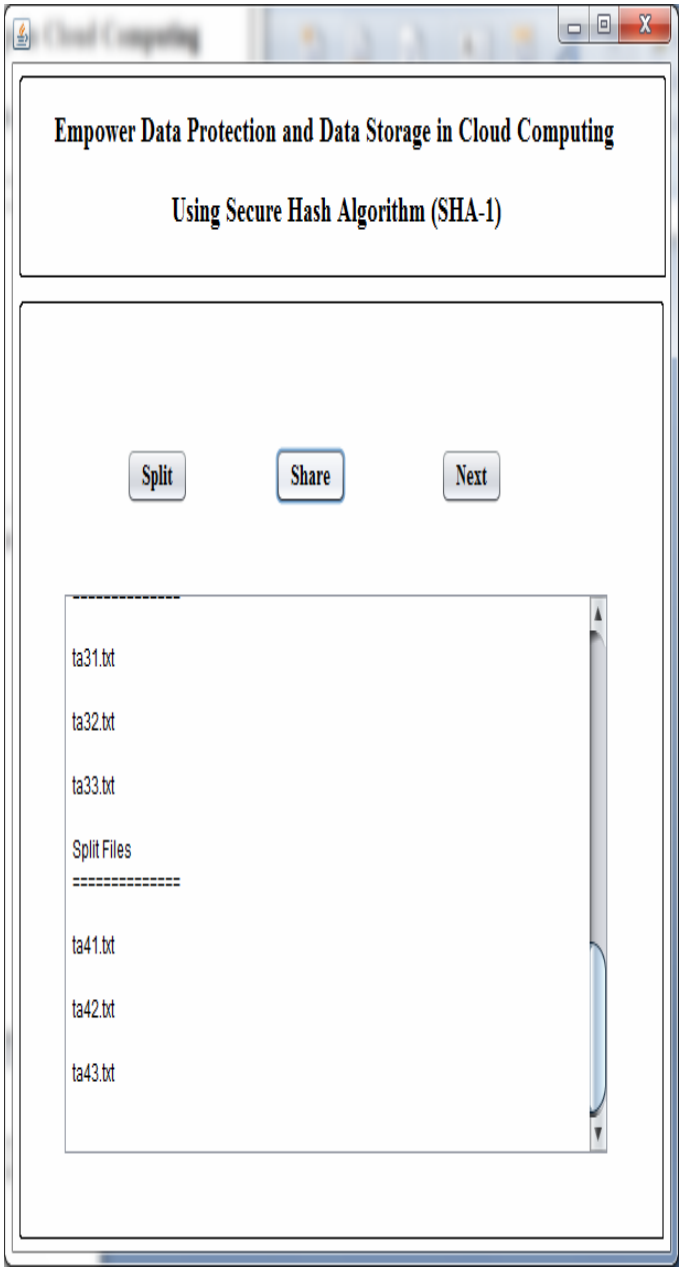


Figure 1: Data Sharing In Cloud

As shown in the figures 1 the client data are uploaded to cloud. Then the client selects the particular data that need to be store .The cloud receives the data uploaded by the client. Then the data is splitting and then .it is stored in the cloud servers. This is done for the security purpose. These data can be retrieved by the client on any time .These data may contain the personal information of the client so the security for the data is very important. The splitter data store in multiple servers in cloud.

2. Find Failure Servers

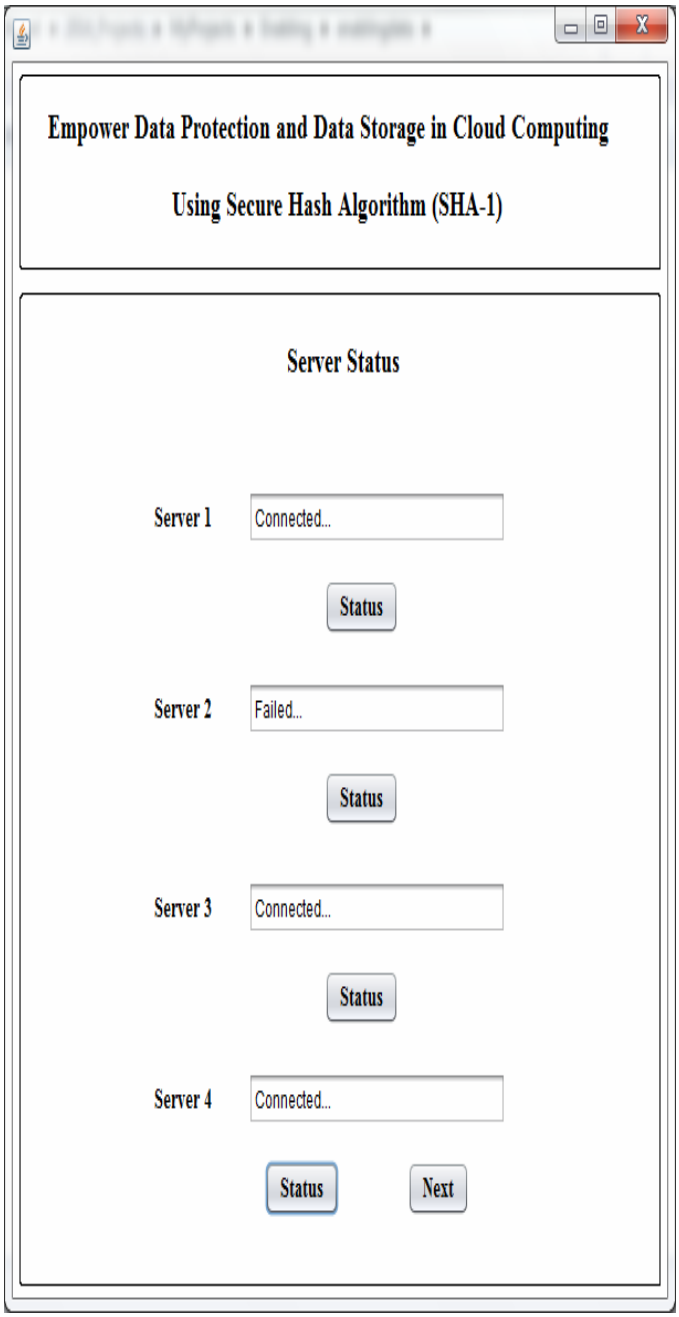


Figure 2: Find Failure Servers

As shown in the figure 2 to source each server in cloud contains different data of the clients and then the capacity of the servers are also different .If any one of the server is failed means then the data in the server is lost .The data in the failed server is not retrieved from that failed server. That is the disadvantage in the existing system but in our proposed system finding the failure node is very important, because from this information the data can be retrieved the failed

server details are send to the other servers in the cloud by finding the failure node.

3. Repair Operation

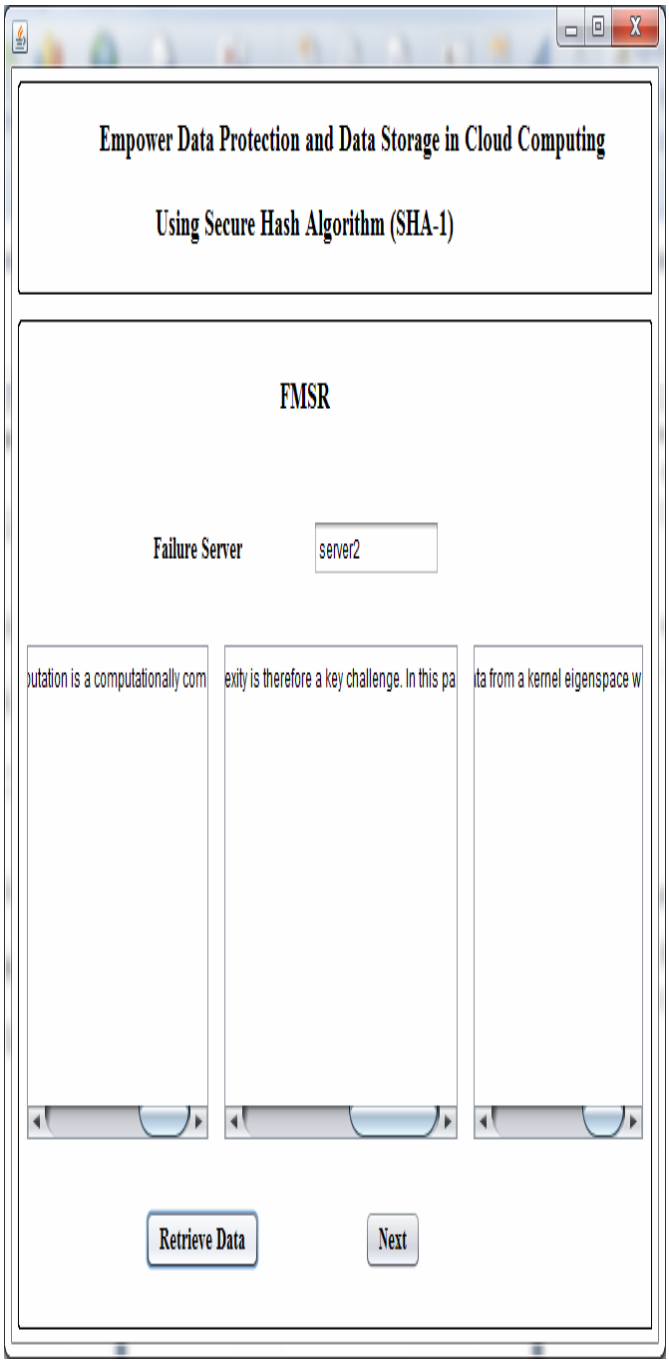


Figure 3: Repair Operation

As shows in the figure 3 in repair operation the data in the failed server is retrieved by FMSR. It is retrieved by get the same data from the other servers .Because the data is stored in multiple servers in cloud. The whole data is not stored in multiple servers. The splitting data is stored in multiple servers in cloud and then the integrity of the data is also confirmed by the server and provides the data to client who needs the data. In that the integrity checking is very important.

4. Data Integrity Protection

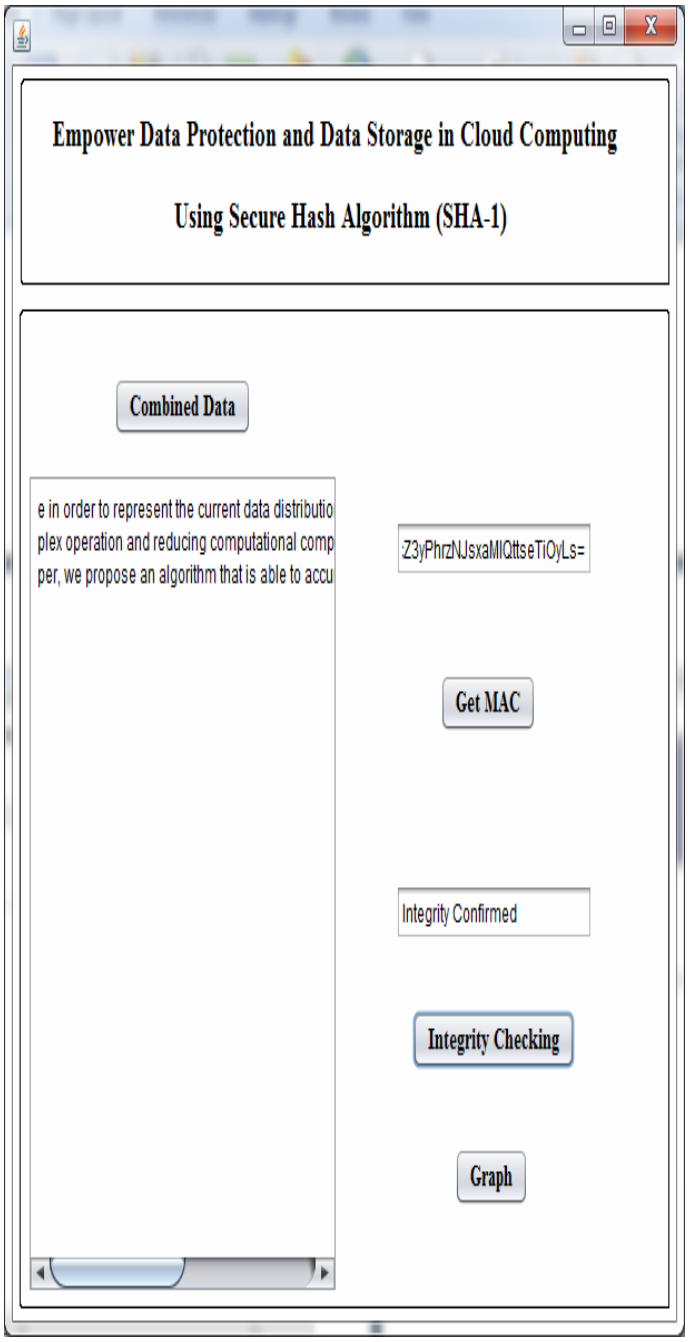


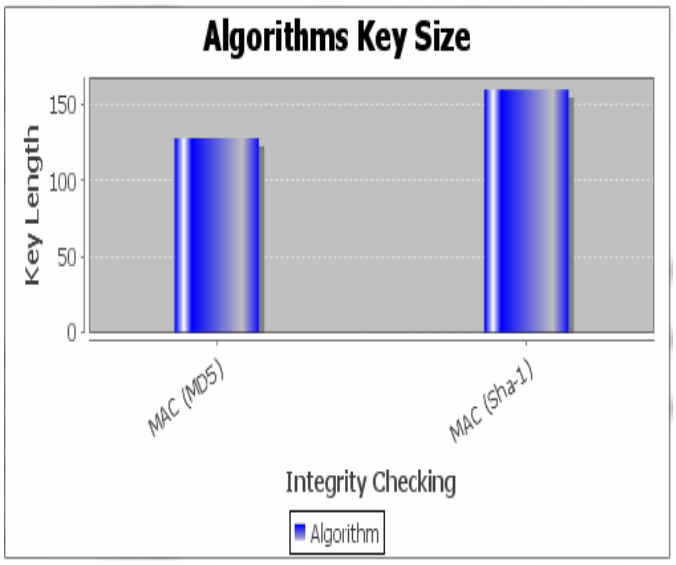
Figure 4: Data Integrity Protection

As shows in the figure 4 integrity checking to the data in cloud server is very important this is also one type of the security because of integrity the variations in the data are confirmed, each data's integrity is confirmed by the message authentication code the secure hash algorithm(SHA1)is used It is better than the MD5. If the attacker made any change in the data means then the cloud server can find out this. .This is

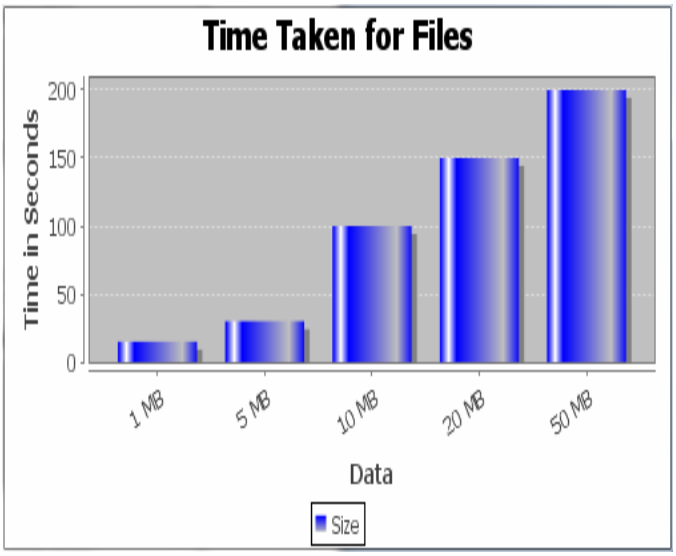
because of the messages authentication code, so that it is used for the security purpose.

V RESULT and ANALYSIS

The key sizes of the algorithms are shown as a graph. MAC has two algorithms. They are MD5 and sha-1. For MD5 the key size and for sha-1 the key size is compared. Using this algorithms the integrity of the data are verified. The according to the size of the data time is consumed. Time should be in seconds. Times taken for all the size of the data are displayed



A) This graph shows comparisons of MAC (MD5) and MAC (SHA1)



B)This graph time token for files sizes

VI. CONCLUSION

Outsourcing the data in the cloud is very useful for clients and the clients need to verify the integrity of their data in the cloud. The Data integrity protection scheme for the FMSR codes is used in the multiserver system. FMSR-DIP codes are used for the fault tolerance and repair traffic saving properties of FMSR codes. The security strength of the FMSR-DIP is increased very much. It is evaluated by the mathematical modeling.

REFERENCES

1. L.Ronald, Kurtz, Russell Dean Viness,"Cloud Security Comprehensive guide to Secure Cloud Computing', 2010
2. Sikder Sunbeam Islam Muhammad Bagar Mollah Md Imanual Huq,Md. Aman Ullah, "Cloud Computing Future Generation of Computing Technology",2012
3. Barrie Sosinsky, "Cloud Computing Bible", 2011
4. B.Chen, R.Curtmola, G.Ateniese, and R.Burns, "Remote DataChecking for Network Coding Based Distributed Storage System,"2010
5. B.Schroeder, S.Damouras, P.Gill, "Understanding Latent Sector Errors "2010
6. Y.Xiangtao Yifa Li, "Remote Data Integrity Checking Scheme for Cloud Storage with Privacy Preserving", 2012
7. R.Saravana kumar, "Data Integrity Proof in Cloud Storage", 2011
8. Yang Lie Wu, Yulin Yan, "Fine Grained Data Integrity Check forPower Cloud Computing", 2012
9. Zhuo Hao,Sheng Zhong, "A Privacy Preserving Remote Data Integrity Protocol with Data Dynamics and Public Verifiability",2011
10. Wenjun Luo, Guojing, "Ensuring Data Integrity In Cloud Data Storage", 2011
11. Q.Wang Cong Wang,Kui Wenjing Lou,JinLi, "Enabling Public Audibility and Data Dynamics for Storage Security in Cloud Computing",2011
12. H.C.H Chen and P.P.C Lee "Enabling Data Integrity Protection In Regenerating Coding Based Cloud Storage", 2012