TO VERIFY THE DATA INTEGRITY AND THIRD PARTY AUDITOR AT USER SIDE IN CLOUD COMPUTING

Dissertation submitted to the Central University of Punjab

For the award of

Master of Technology

In

Centre for Computer Science and Technology

BY

Renuka Goyal

Supervisor

Er. Navjot Sidhu

Centre for Computer Science and Technology

School of Engineering and Technology

Central University of Punjab, Bathinda

September, 2014

DECLARTION

I declare that the dissertation entitled “TO VERIFY THE DATA INTEGRITY AND THIRD PARTY AUDITOR AT USER SIDE IN CLOUD COMPUTING” has been prepared by me under the guidance of Er. Navjot Sidhu, Assistant Professor, Centre for Computer Science and Technology, School of Engineering and Technology, Central University of Punjab.

No part of this dissertation has formed the basis for the award of any degree or fellowship previously.

Name: Renuka Goyal
Centre for Computer Science and Technology,
School of Engineering and Technology,
Central University of Punjab,
Bathinda –151001.

Date:

CERTIFICATE

I certify that Renuka Goyal has prepared her dissertation entitled “TO VERIFY THE DATA INTEGRITY AND THIRD PARTY AUDITOR AT USER SIDE IN CLOUD COMPUTING” for the award of M.Tech. degree of the Central University of Punjab, under my guidance. She has carried out this work at the Centre for Computer Science and Technology, School of Engineering and Technology, Central University of Punjab.

Er. Navjot Sidhu
Assistant Professor
Centre for Computer Science and Technology,
School of Engineering and Technology,
Central University of Punjab,
Bathinda – 151001.

Date:

ABSTRACT

To verify the Data Integrity and Third Party Auditor at User Side in Cloud Computing

Name of Student: Renuka Goyal

Registration Number: CUPB/M.Tech/SET/CST/2012-13/07

Degree for which submitted: M.Tech.

Name of Supervisor: Er. Navjot Sidhu

Name of centre: Centre for Computer Science and Technology

Name of School: School of Engineering and Technology

Key words: Cloud computing, Data Integrity, Third Party Auditor, MAC, Digital signatures, Public- auditing.

Now-a-days the concept of Cloud Computing is one of the major theories in the world of IT. Its services are now being applied to several IT scenarios. Cloud Computing is the Internet based computing which provides users with a number of services. One of the services provided by the cloud computing is data storage. Users store their data in the cloud without the burden of local data storage. But as every technology has some drawbacks, cloud computing also has some drawbacks. The main drawback of cloud computing is the security issues that are associated with it. The security issue with which this dissertation is dealing is the threat to data integrity. As the user no longer has physical possession of data so the integrity of data stored on cloud server become the major concern in the cloud computing. Data stored on the cloud server may be getting corrupted and sometimes even the cloud service provider for his own benefit like for more space on data centre can discard the user data which is not used for a longer time. In order to maintain the integrity of data, the user takes the assistance of a Third Party Auditor (TPA). The Third Party Auditor checks the integrity of data on user demand and the released audit reports not only helps the user to evaluate the risk of their services but also helps the cloud service provider to improve their security mechanism. As a part of verification it is assumed that Third Party Auditor is reliable and independent which does not mean that there is no space for the Third Party Auditor to cheat. So there is need to check the integrity and Third Party Auditor at user side. This dissertation highlights the basics of cloud computing, general model and different approaches used for Third Party Auditor. The model to verify the data integrity and Third Party Auditor is implemented in this dissertation and the results found are good according to the user data.

(Renuka Goyal) (Supervisior: Er. Navjot Sidhu)

DEDICATED TO GOD,
MY FAMILY MEMBERS AND MY FRIENDS

ACKNOWLEDGEMENTS

This dissertation in itself is an acknowledgement to the inspiration, drive and the technical assistance contributed to it by many people. It would have never seen the light of day without the help and guidance that it received from them.

First and foremost, I would like to express my sincere gratitude to my guide Er. Navjot Sidhu, Assistant Professor, Centre for Computer Science and Technology, School of Engineering and Technology, Central University of Punjab, for immense help, guidance, and encouragement all the time during this dissertation work. This work would not have been possible without her encouragement. She always provided a monitoring and enthusiastic atmosphere to work with; it was a great pleasure to do this dissertation under her supervision.

I also would like to thank Er. Surinder Singh Khurana, Assistant Professor, Centre for Computer Science and Technology, School of Engineering and Technology, Central University of Punjab, for his precious time, suggestions and support that, he gave me in choosing the field of my research and guided me to do my work.

I would also like to extend my sincere and extreme gratefulness to Prof. A.K Jain, CoC for Computer Science & Technology, faculty of Centre for Computer Science & Technology and the staff of University Computer Centre for providing an exceptionally excellent academic environment.

This acknowledgement would be deficient if I skip mentioning the emotional support, love, blessings and inspiration provided by my family. I would also like to appreciate my classmates for critical peer reviewing of my dissertation.

Last but not the least I bow to the almighty, who gave me strength to accomplish this work with honesty, authenticity, sincerity and dedication.

(Renuka Goyal)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Sr. No.	Full Form	Abbreviation
1.	Amazon Web Service	AWS
2.	Cloud Server	CS
3.	Cloud Service Provider	CSP
4.	Elastic Compute Cloud	EC2
5.	HyperText Markup Language	HTML
6.	Hypertext Transfer Protocol	HTTP
7.	Homomorphic Linear Authenticator	HLA
8.	Infrastructure as a Service	IaaS
9.	Integrated Development Environment	IDE
10.	Institute of Electrical and Electronics Engineers	IEEE
11.	Information Technology	IT
12.	Message Digest 5	MD5
13.	Message Authentication Code	MAC
14.	National institute of Standards and Technology	NIST
15.	Platform as a Service	PaaS
16.	Personal Computer	PC
17.	Personal Digital Assistant	PDA
18.	Provable Data Possession	PDP
19.	Proof of Retrievability	PoR
20.	Rivest-Shamir-Adleman	RSA
21.	Simple Storage Service	S3
22.	Software as a Service	SaaS
23.	Secure Hash Algorithm	SHA
24.	Service Level Agreement	SLA
25.	Third Party Auditor	TPA
26.	Trusted Third Party	TTP
27.	Trusted Control Model of Cloud Storage	TCMCS
28.	Virtual Private Network	VPN

CHAPTER 1

INTRODUCTION

Cloud Computing has gained intention since 2007. It is the general term for anything that involves providing services on the Internet. It moves the data and computing from desktop to large data centres. It is combination of parallel, grid and distributed computing. The wide adoption of cloud computing is restricted to its privacy and security issues. This chapter introduces the basics of cloud computing, its characteristics, deployment models and the service models. The various security issues associated with cloud computing are also described in this chapter. This chapter mainly addresses the issues of data integrity and to check the integrity of data that is stored on the cloud server user take the assistance of Third Party Auditor (TPA).

1.1. Cloud Computing

The term "cloud" originates from the world of telecommunications when providers use virtual private network (VPN) services for data communications. Cloud computing provides computation, software, data access and storage services that may not require user knowledge of the physical location and the configuration of the system that is delivering the services. It is receiving a great deal of attention, both in publications and among users (Jadeja & Modi, 2012).

Cloud Computing is a large pool of systems which are connected via private or public networks, to provide infrastructure for application, data and file storage (Makhija, Gupta & Rajput, 2013). With the use of this technology, the computational cost, application deploying, data storage and delivery are reduced significantly. The basic principle of cloud computing is to make the computing be assigned in a great no. of distributed computer rather than local computer or remote server (Zhang, Chen, Zhang & Huo, 2010).

One way to think of cloud computing is to relate it with email. The email client like Yahoo!, Gmail, Hotmail, and so on takes care of housing all of the hardware and software necessary to support user personal email account. When a user wants to access his email first he opens his web browser, go to the email client and log in with his username and password. The most important part is

having Internet access. The email is not housed on user's physical computer; he accesses it through an Internet connection, and can access it anywhere. If the user is on a trip, at work, even then he can check his email as long as he has access to the Internet. Email is different than the software installed on the computer, such as a word processing program. When a user creates a document using word processing software, the document stays on the device that the user has used to make it unless he physically move it. An email client is similar to how cloud computing works. Except instead of accessing just email, one can choose what information one has access to within the cloud.

According to (Vaquero, Radero-Merino, Caceres & Linder, 2009) Clouds are large pools of easily accessible services and virtualized resources such as virtual machines, storage, memory, network bandwidth, processing which can be dynamically reconfigured to adjust to a variable load. They also allow optimum resource utilization. This pool of resources is typically exploited by a pay-per-use manner in which guarantees are offered by the cloud service provider by means of standardized service level agreement (SLA). The features that closely support this minimum definition of cloud computing are scalability, pay-per-use utility model and virtualization.

Many big companies such as IBM, Google, Amazon, Microsoft, Yahoo and other move their space to develop cloud computing. These companies have launched their own cloud computing infrastructures and cloud services on the basis of their original products and achieved best application results and social impact, such as Amazon's EC2 and S3, Google' Google Apps, Microsoft' Azure and so on (Wang & Mu, 2011).

According to National Institute of Standards and Technology (NIST): "Cloud Computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction". It also defines that a cloud model is composed of:

- 5 essential characteristics
- 3 cloud service models
- 4 cloud deployment models

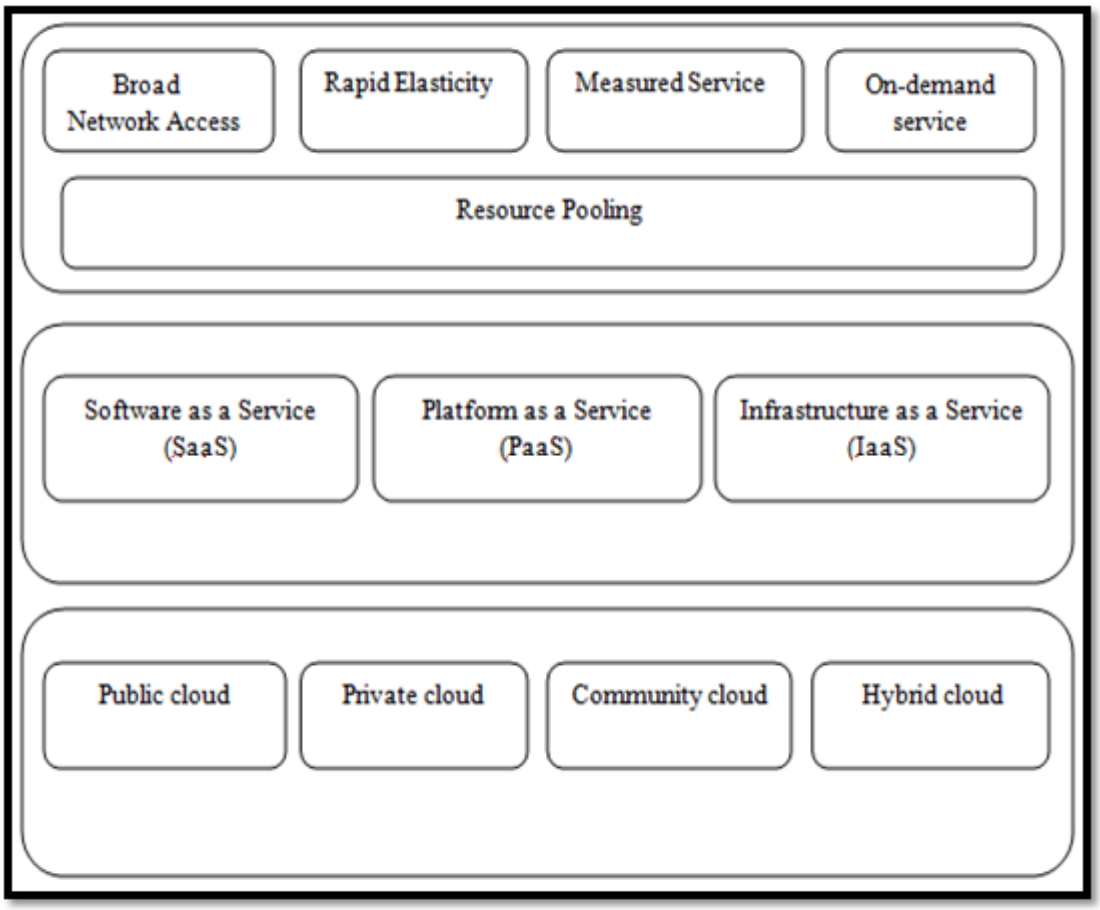

Fig. 1.1: Visual Model of NIST Definition of Cloud Computing

1.1.1. Essential Characteristics of Cloud Computing

Cloud Computing is having many features like platform, location and device independency, high availability, resource reuse, low cost, virtualization at large scale, reliable services, resource sharing, scalability, etc. so it can be easily adopted by all sizes of business particular small and mid-sized. According to (Balakrishnan, Saranya, Shobana & Karthikeyan, 2011) (Idrissi, Kartit & Marraki, 2013) (Piplode & Singh, 2012) (Xiao & Xiao, 2013) following are the five key characteristics of cloud computing:

- **On-demand Service:** The users can get computing capabilities as needed automatically. There is no need for users to directly interact with the cloud server provider. The computing capabilities can be server time, software use, network storage etc.
- **Broad Network Access:** The services are available over the Internet via a standard mechanism that allows the users to access these services through any thin or thick client tools like desktop, laptop, PDA, mobile phone.

- **Resource Pooling:** The cloud service provider employs a multitenant model to serve multiple customers by pooling computing resources like virtual machines, storage, memory, network bandwidth, processing. The different physical and virtual resources can be dynamically assigned and reassigned according to users demand. The users have no knowledge of the exact location of the provided resources.
- **Rapid Elasticity:** The computing resources provided are elastic for the users. The users can rapidly scale up to use whenever needed or scale down to release whenever finished. From the user's point of view the available services are unlimited and can be purchased according to their need at any time.
- **Measured Service:** The cloud server uses a mechanism to measure the usage of resources and services for each individual user. For both the provider and the user's resource usage will be monitored, controlled, metered and reported. The users pay according to the usage.

1.1.2. Cloud Deployment Models

According to (Jadeja & Modi, 2012) (Mollah, Islam & Islam, 2012) (You, Peng, Liu & Xue, 2012) (Idrissi, Kartit & Marraki, 2013) enterprises can choose to deploy applications on four types of cloud that are Public, Private, Hybrid or Community cloud. These deployment models describe who owns, manages and is responsible for the services provided.

- **Public Cloud:** The cloud infrastructure is open to use by the general public. This includes individuals, corporations and other types of organizations. It can be accessed by any user with an Internet connection and access to cloud space. Users need to pay only for the time duration, they use the service, i.e. according to pay per use model. They do not know about the other users who are using the same server or network. It may be owned, managed and operated by a business, academic or combination of them. However, public clouds are less secure compared to other cloud models because the data and applications on public cloud is more prone to attacks. The solution can be that security checks be implemented on both cloud service provider and user side. For example: Amazon, Google Apps, Window azure.

- **Private Cloud:** The cloud infrastructure is used by a single organization. It is established for a specific group or organization and limits access to just that group or organization. The organization has the infrastructure and on the basis of that infrastructure, it can control the way to deploy applications, control how and where the applications run. They can determine which user is allowed to use the infrastructure. It may be owned, managed and operated by the organization, a third party or some combination of them. The advantage here is that it is easier to manage security, maintenance and upgrades and also provide more control over the deployment and use. This is more secure as compared to public as only the users of the organization have access. For example: eBay.
- **Community Cloud:** The cloud infrastructure is used by a specific community of users. The community is made of two or more groups or organizations that have similar cloud requirements. It may be owned, managed by one or more organization in the community, a third party or some combinations of them. For example: zimory and RightScale.
- **Hybrid Cloud:** A hybrid cloud is a combination of at least two clouds infrastructures, where the clouds included are a mixture of public, private or community cloud. The cloud infrastructures will be unique entities, but bound together by technology that enables data and application portability. It is created to fulfill the demand of the organization. There are not many hybrid clouds, but some companies like IBM and Jupiter have introduced their base technologies for hybrid cloud.

1.1.3. Cloud Service Models

Cloud computing provides many services to their users via Internet. According to (Kuyoro, Ibikunle & Awodele, 2011) (Balakrishnan et al., 2011) (Mahmood, 2011) (Wang & Mu, 2011) there are three types of cloud service providers that one can subscribe to: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). These three types differ in the amount of control that user have over his information, and conversely, how much user can expect his provider to do for him. In this IaaS is most basic and each higher model abstracts from the details of lower models. These services can be used independently, but can work together.

- **Software as a Service (SaaS):** The user has the capability to use the service provider applications running on a cloud infrastructure. There is no any need to install and run the application on the desktop computer. The applications can be accessed from various client devices either through thin client interfaces such as a web browser. The user doesn't manage the underlying cloud infrastructure, including the network, servers, Operating system, storage or applications. For example: SalesForce, SAP Business by Design, Google docs etc.
- **Platform as a Service (PaaS):** The user has the capability to deploy onto cloud infrastructure, applications created using some programming languages, libraries and tools supported by the provider. Thus the user need not to go through the burden of installing the software and hardware required for it. The user doesn't manage the underlying cloud infrastructure, including the network, servers, operating system, and storage, but he can control the applications that he has deployed over the cloud. For example: Force.com, Google app engine.
- **Infrastructure as a Service (IaaS):** The user has capability provided to provision processing, storage, networks and their fundamental computing resources so that he can deploy and run arbitrary software, which include operating system and applications. The user doesn't manage or control the infrastructure. He does manage or control the operating system, storage, applications, selected network components. For example: Amazon's EC2, Amazon S3, Rackspace cloud server.

1.1.4. Advantages of Cloud Computing

Cloud computing offers many advantages to the users. According to (Jadeja & Modi, 2010), (Mollah, Islam & Islam, 2012) following are some advantages of cloud computing:

- **Easy Management:** The maintenance of the infrastructure hardware or software is simplified, thus less headaches for the IT team. Also applications that are quite storage extensive easier to use in the cloud environment compared to the same when used by the organization with its own. Also at the user level, what one mostly need is a simple web browser with Internet connectivity.

- **Cost Reduction:** As the applications run on the cloud infrastructure, not on the desktop PC, so there is no need to load any software programs or the files to be saved on the desktop computers. Also the manpower required for such systems is not required. Even simple applications like email can be set up almost free through applications like Google Apps. Also, as most of the time such providers are quite reliable in terms of availability.
- **Green Computing:** Harmful emissions due to extensive use of systems in organizations, electronic waste generated as the time passes and energy consumption is the main disadvantage of the present day computing systems. This can be reduced to some extent by using cloud computing services. This leads to environment preserving. Also the e-waste generated to a minimum extent.
- **Automatic Software Updates:** As the application is web based the updating to the software happens automatically and the updating is available to the user when he uses the cloud application next time. So the user has not to face obsolete software and high upgrade costs. There is also no any problem of format incompatibility when everyone is sharing documents and applications.
- **Limitless Storage Capacity:** Cloud computing provides unlimited storage space to their users. The cloud offers storage up to hundreds of Pbytes which is very big compared to 1 Tbytes hard drive of desktop computers.
- **Increased Data Reliability:** In a desktop computer, if a hard disk gets crashed then it destroys all our valuable data, but in cloud if a computer gets crashed it does not affect the data storage. If the personal computer crashes, all user data are still there in the cloud and is accessible. So cloud computing is safe computing for those users who do not back up their data on regular basis.
- **Universal Document Access:** The user has no any need to take the documents with him they are available where the user is. The user just requires a computer with internet connection to access his documents.
- **Device Independence:** A User is no longer always attached with a single computer or network. In cloud computing the changes that a cloud user made to his computers, applications and documents follow through the

whole cloud. If he moves to a portable device and the applications and documents are still available to the user**.**

- **Easier Group Collaboration:** One of the important advantage of cloud computing is sharing of documents that leads to better collaboration between the cloud users. Many users in the cloud do the collaboration to take advantage of cloud computing. The multiple users can form a group by collaborating easily on documents and projects.

1.1.5. Disadvantages of Cloud Computing

Every technology has its own benefits and drawbacks. Along with the many advantages cloud computing also has some disadvantages. According to (Jadeja & Modi, 2010), (Mollah, Islam & Islam, 2012) following are the some disadvantages of cloud computing:

- **Does not work without internet connection:** The user requires an Internet connection to get services of cloud computing and is not possible without connecting to the Internet. Since user uses the Internet to connect to both his applications and documents, if he does not have an Internet connection he cannot access anything, even his own documents. Also cloud computing is impossible in the areas where there is no Internet connection**.**
- **Cannot work with low-speed internet connections:** Cloud computing does not work well with low speed internet connections which are likely provided by the dial-up services. Web-based applications require large bandwidth to download heavy documents.
- **Stored data might not be secure:** User stores their data on a cloud but they do not know where their data is located or stored. The unauthorized users can gain access to confidential data.
- **Features might be limited:** The applications on the cloud may not have the full features like the desktop based applications which are full-featured. For example: Microsoft PowerPoint provides more features to user than the Google Presentation's web-based offering.
- **Limited control and flexibility:** As the services and applications run on virtual environments, the cloud users have limited control over the execution of the applications and on the functioning of the hardware and

software. The remote software used lacks the features of an application that runs locally.

- **Stored data can be lost:** As we know on cloud data stored is safe due to replication of data across multiple machines. But if sometime by chance data goes missing, users have no physical or local backup on their personal computers. In that case, simply relying on the cloud puts users at risk if the cloud gets down.

1.2. Cloud Computing Security Issues

Cloud computing offers many benefits to their customers like universal data access with independent geographical locations, relief of the burden for storage management and avoidance of capital expenditure on hardware and software maintenances, etc. but there are also some security issues related with the cloud computing. Due to the opaqueness of the cloud the user may not know the internal details of cloud service providers which bring many security issues in cloud computing. According to (Behl, 2011) (Wang,2011) (Chen & Zhao, 2012) (Behl & Behl, 2012) (Dillon, Wu & Chang, 2010) (Tianfield, 2012) (Mahmood, 2011) following are the basic security issues in cloud computing:

- **Confidentiality:** Confidentiality implies that the user data must be kept confidential from other users and cloud providers. Only the authorized parties or systems having the ability to access data. In cloud computing due to increased number of parties, devices and applications involved, the number of access points increases. Users do not have control over their data. Their data can be accessed and misused by the cloud providers and other users.
- **Data Integrity:** Data integrity means the data should be correctly stored on the cloud server without any modification and if any violations, i.e. if the data is getting lost, altered or compromised can be detected. But the integrity of data is at risk in cloud server. Data can get modified by other users or even sometimes cloud service provider for his own benefit can behave unfaithfully towards the others users regarding outsourced data. For example cloud service providers for more space on data centre can discard the user data which have not been or rarely accessed by the user for a

longer time or even can hide the data loss incidents to maintain his reputation.

- **Availability:** Availability is an important issue as the core function of cloud computing is to provide on demand service of different levels. The services must be available to users at any time and from any place. If certain service is no longer available or the quality of service cannot meet the service level agreement customer may lose faith in cloud system. Availability not only refers to data, software, but also refers to hardware. Denial of service attacks, equipment outages and natural disasters are all threats to the availability.
- **Privacy:** Privacy means to control the disclosure of personal information. Cloud computing brings a number of legal challenges towards privacy issues involved in data stored in multiple locations in the cloud. Instead of user's data being stored on the company's servers the data in cloud computing are stored in the service provider's data center, which could be anywhere in the world. Privacy is a core issue in cloud computing. There is need to protect identity information, transaction histories, etc. Storing the data to shared infrastructure increased the risk of unauthorized access and exposure to data.
- **Access Control:** It is probably for the confidential data to be illegally accessed due to the astringent access control. Unauthorized access may exist if the security mechanism is not adequate. Entities in the service chain can easily utilize the vulnerability to access users' data. As data usually exists in the cloud for a long time, the risk of illegal access is higher.
- **Audit:** As the cloud service provider has the powerful control of the data, so it is probable for cloud service provider to copy, move and edit the data in the procession. So there is a need to have a mechanism to audit the process. Auditing is the process of reviewing and examining the authorization and authentication records. Users need to make sure that all the activities are traceable so that they can trust the environment. However, there are several aspects unsolved, for example, as data in the cloud is in huge amount, it is not possible to audit all the data. Sometimes the auditing task is assigned to a third party who audits the data on user behalf.

- **Data Location:** The cloud computing offers a high degree of data mobility. The client does not know the actual place where the data is saved because it is distributed over many places that led to the confusion among the users. When an enterprise has some sensitive data that is kept on a storage device in the cloud, they may want to know the location of it. This requires an agreement between the cloud provider and the user that data should stay in a particular location. The issue is user sometimes not aware of the implication of this and no such contract is agreed.

1.3. Third Party Auditor: An integrity Checking Technique

One of the major concerns in cloud computing is the integrity of data as a user having no physical control over his data and does not know the exact location of this data. So there is a need to ensure the integrity by making the user, capable to check over the cloud data from any unauthorized modification. One solution is to first download the files whose integrity have to check, but downloading the files requires high transmission cost. So to maintain the data integrity and to minimize the storage risk it is important to take assistance of a Third Party Auditor (TPA) who checks the data integrity for the cloud user and helps the user in minimizing his risk (Attas & Batarfi, 2011).

The Third Party Auditor is a kind of inspector. He has the resources and experience that a user does not have and check the integrity that is difficult for users to check. The Third Party Auditor should not put any additional burden on the users and should honestly check the integrity of user data without demanding any local copy of the data. The released audit reports not only help the user to evaluate the risk of their services, but also help the cloud service provider to improve his security platform (Wang ,Wang, Cao, Ren & Lou, 2012) (Gowrigolla, Sivaji & Masilliamani, 2010).

1.3.1. Model of Cloud Data Storage

According to (Paigude & Chavan, 2013) (Luo & Bai, 2011) (Muneshwara &Chandarl, 2012) there are three different network entities in the cloud storage model which is users, cloud service provider and Third Party Auditor. They all play a specific role in the cloud storage model.

1. **Cloud Users:** These are active participants in the model. They have data to be stored on the cloud server and rely on the cloud service provider for data maintenance and computation. Both individual consumers and organizations can be the users.
2. **Cloud Service Provider (CSP):** It is the most important part of the cloud architecture. It has significant storage space and computation resource to maintain the user data and provide all the services available for a client in pay as you use manner.
3. **Third Party Auditor (TPA):** It has more capabilities and experience than the user and checks the integrity of data for the user upon users request and his audit reports helps the users in evaluating the risk.

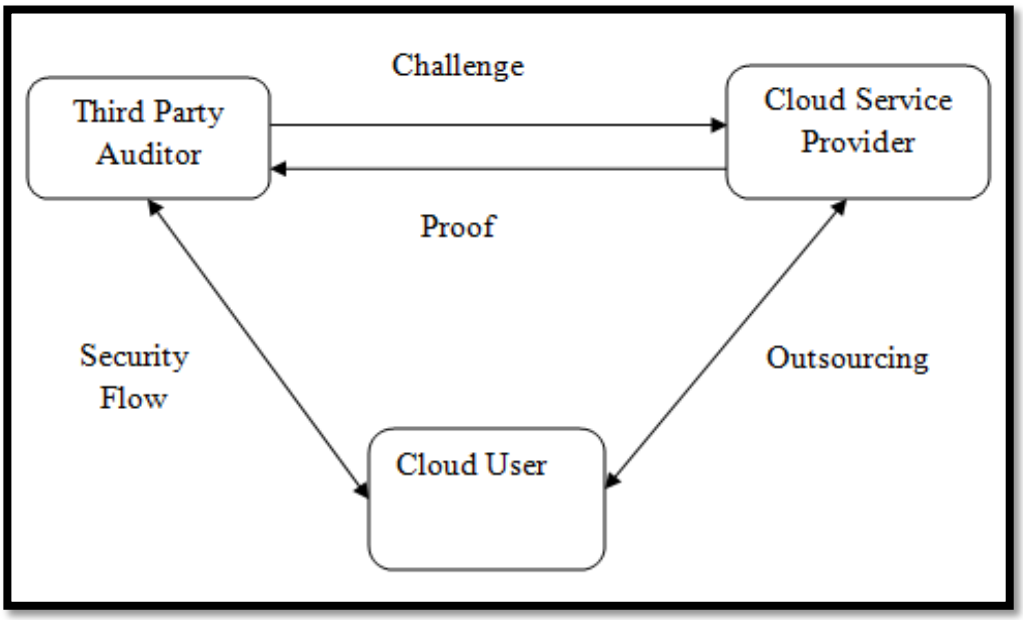

Fig. 1.2: Cloud Data Storage Model

1.3.2. Features of Third Party Auditor:

According to (Patel & Patel, 2012) (Gowrigolla, Sivaji & Masilliamani, 2010) (Balakrishnan et al., 2011) standard Third Party Auditor in cloud environment should take following functionalities into consideration:

- **No data leakage or data learning:** The Third Party Auditor should neither learn any information about the data file from the message it receives from client/server nor leak the same to any unauthorized entity.
- **Audit without downloading:** The Third Party Auditor should audit without asking for entire file from the server, not even in encrypted form. Third Party

Auditor should audit the user data without asking for the local copy of the data or even learning the data contents.

- **Integrity Verification:** One of the important security concerns is to verify integrity of data that is stored on the cloud server. Third Party Auditor should verify the integrity of client's data stored on a cloud with low communication overhead.
- **High Performance:** Performance of Third Party Auditor is also an important issue as it is a central component of the cloud system, where there are thousands of clients and multiple servers. Third Party Auditor should not become a bottleneck of an entire system and performance of overall system should not be compromised due to heavy load on Third Party Auditor.
- **Scalability:** As cloud is a completely dynamic environment, any number of users can come in or go out. Also, it is expected to have huge data storage on cloud servers. Functionalities of Third Party Auditor should not be affected by number of cloud clients, servers, number of data files stored on the cloud or the overall size of the entire storage.

1.3.2. Different Schemes used for Third Party Auditor

Different authors gave different models or schemes for the Third Party Auditor. All the techniques have their own merits and demerits. Following are the some of the schemes.

1.3.2.1. MAC Based solutions:

There are two types of MAC based solutions. First solution does not ensure privacy preserving the second one suffers from auditor statefulness and other demerits (Wang ,Wang, Ren & lou, 2010).

In first solution user first divides the files into blocks and calculate the message authentication code (MAC) for each block. Users transfer the file blocks and the MACs to the cloud service provider, and share the secret key to Third Party Auditor. Third Party Auditor demands for a random no. of blocks and theirs MACs from the cloud service provider. Then Third Party Auditor uses the secret key to verify the correctness of stored data on the cloud server (Wang, Wang, Chow, Ren, & Lou, 2013).

The drawbacks of this system are:

- Third Party Auditor requires retrieval of data blocks for verifying the correctness of data blocks which is not privacy preserving.
- It only works with the static data.
- Computation and Communication complexity are proportional to the sample size.

In the second solution to avoid the requirement of data retrieval in Third Party Auditor verification, one can improve the solution as: first user chooses S random keys, pre-computes S MACs for the whole data file f and publish these verification metadata (the keys and the MACs) to Third Party Auditor. Then he outsources the data. The Third Party Auditor can reveal a secret key to the cloud server and for each audit he demands a new keyed MAC for comparison (Wang, Ren, Lou & Li 2010).

As it is impossible to recover F so this scheme can be called as privacy preserving. However, it suffers from the following drawbacks:

- The user can only check the integrity of a file only for the number of times as there are the secret keys that must be fixed a priori. As all the secret keys get exhausted then one has to retrieve data in full to recompute and republish new MACs to Third Party Auditor.
- In this scheme it is difficult for the Third Party Auditor to maintain and update state for the all audits. As there are large number of audit requests from the multiple users so maintaining such states can be error prone for Third Party Auditor.
- This scheme does not support dynamic data this only deals effectively with static data.

1.3.2.2. Public Auditing Scheme for Third Party Auditor:

According to (Paigude & Chavan, 2013) (Yang & Jia, 2013) (Gowrigolla, Sivaji & Masilliamani, 2010) the working of Third Party Auditor Consists of the four algorithms: KeyGen, SigGen, GenProof, VerifyProof.

- **KeyGen:** Key generation algorithm is run by the cloud user for the setting of scheme. In this the user generates his public/private key pairs.

- **SigGen:** This algorithm is run by the cloud user for the generation of verification metadata, which can be signatures or other information used for auditing.
- **GenProof:** This algorithm is for the generation of proof of correctness of data storage and is run by the cloud server.
- **VerifyProof:** This algorithm is run by the Third Party Auditor for the auditing of the proof generated by the cloud server.

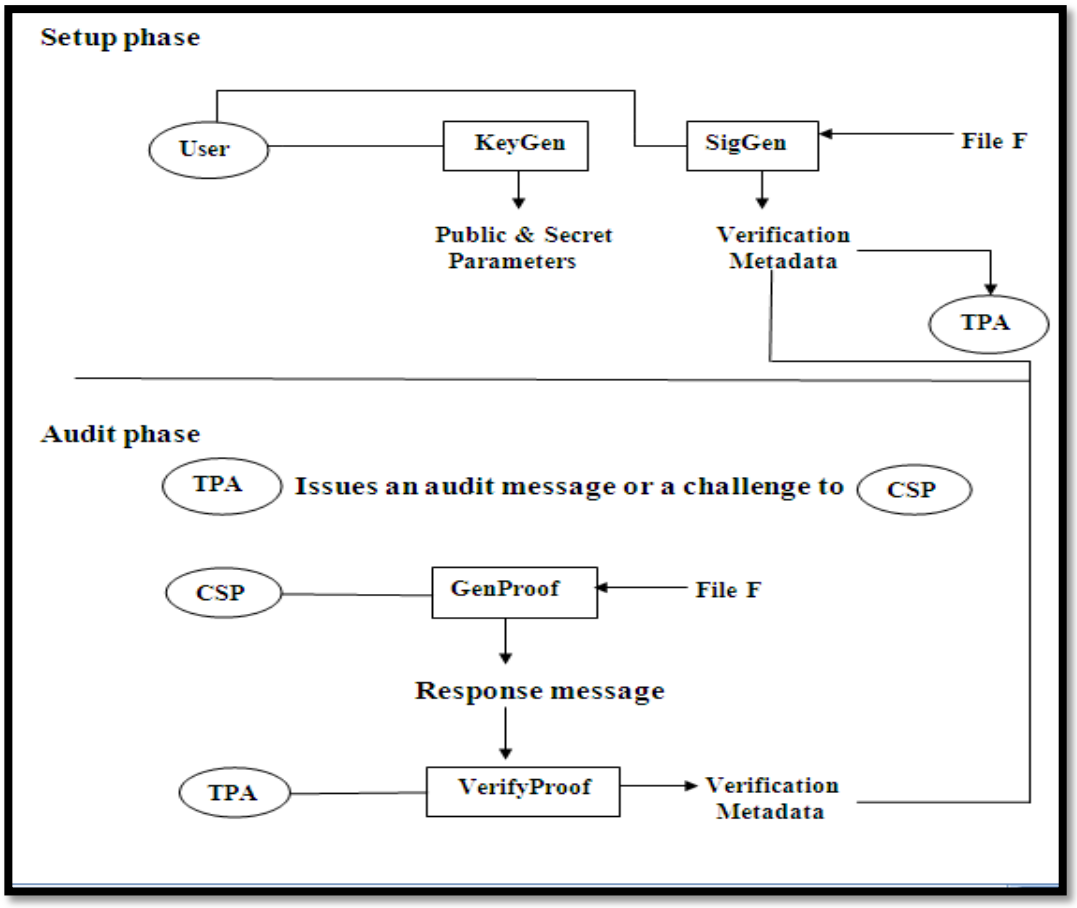

Fig. 1.3: Public Auditing Scheme for Third Party Auditor

The above four algorithms work in two phases. First is the setup phase in which KeyGen and SigGen algorithms are implemented. Second is the audit phase in which GenProof and VerifyProof algorithms are implemented.

- **Setup Phase:** The cloud user first generate the public and private keys by running KeyGen algorithm and then generate the verification metadata to preprocesses the files by using SigGen algorithm. Then the user upload the original data files to the cloud server and delete the local copies of files from

his computer.Then the verification metadata is sent to the Third Party Auditor for later audit by user (Yang & Jia, 2013).

- **Audit Phase**: The Third Party Auditor sends an audit message to the cloud service provider to give the proof of correctness of data stored on the cloud. The cloud service provider responds to message by executing GenProof algorithm. Then the Third Party Auditor verifies the response by executing VerifyProof algorithm using the verification metadata (Attas & Batarfi, 2011) (Bhagat & Sahu, 2013) (Piplode & Singh, 2012).

1.3.2.3. HLA (homomorphic linear authentication) based solution:

To support public auditability without retrieving the data blocks the HLA (homomorphic linear authentication) based solution is used. Like the MACs, HLAs are also some unforgeable verification metadata that is used to check the integrity of data. The only difference is that to authenticate a linear combination of the individual data blocks one can compute aggregated HLAs. It allows efficient auditing and consumes constant bandwidth. But this solution may reveal user data information to Third Party Auditor and violates privacy preserving (Wang et al. 2012) (Balakrishnan et al., 2011).

1.3.3.4. Privacy- Preserving Public Auditing Scheme:

To achieve privacy preserving public auditing, public key based homomorphic linear authentication with random masking is used. Third Party Auditor checks the integrity without demanding the actual copy of the data. When the cloud service provider gives response message then the data blocks that are present in the response message are masked by the server. Due to the random masking, the Third Party Auditor is not able to extract user data, no matter how many linear combinations of the same set of file blocks can be collected (Wang et al., 2013).

1.3.3.5. Digital signatures based solution:

The digital signature method is used to protect the privacy and integrity of data. It uses the RSA algorithm for encryption and decryption purpose and then the hash is calculated followed by the process of digital signatures for message authentication. In this protocol there are three main participants (i) Third Party Auditor (TPA) (ii) User (iii) Cloud Service Provider. As the initial requirement the

user and the TPA generate their own private key and public key with respect to the strong RSA algorithm. The public keys have been shared between them as the part of the SLA or in some other ways. Then, with respect to the protocol the message is encrypted as well as signed in an unique way.

With the generated public key sets, they get exchanged between the user and the Third Party Auditor. At first the data is signed with the user's private key then the cipher is again encrypted with the TPA's public key. This package is now sent to the Cloud and also the Third Party Auditor. The Third Party Auditor now decrypts the encrypted message with his private key and then de-signs the cipher with the user's public key to recognize the data. Then the same process of decryption is carried out in the cloud by the Third Party Auditor to verify the correctness by comparing the data which he has with the stored one. Then, as per the result the Third Party Auditor indicates the user (Govinda, Gurunathparsad & K Kumar, 2012).

1.4. Problem Statement

One of the major risks of the cloud computing is the threat to data integrity. Data integrity states that data should be honestly stored on the cloud servers. There should not be any modifications done by either the cloud provider or the other user. It is an issue where there may be some unauthorized alteration of the data without the consent of the data owner.

There are many mechanisms to check the integrity of data. One of the solutions is that the cloud user assigns the work of integrity checking to a Third Party Auditor. The Third Party Auditor has an experience that the cloud user does not have and also Third Party Auditor checks the integrity that is difficult for the user to check.

There are many schemes given by the different authors for the implementation of Third Party Auditor but In all the existing systems, there is an assumption that the Third Party Auditor is reliable and independent, but which does not mean there is no space for the Third Party Auditor to cheat. So there is a need to build a mechanism in which the user can verify the data integrity and the work done by the Third Party Auditor so that the threat to data integrity gets minimized.

1.5. Proposed Model

As mentioned in the problem statement that there is a possibility that the Third Party Auditor can cheat the cloud user by generating the false audit reports. So we propose a model in which cloud user checks the integrity of data and the Third Party Auditor. In the proposed model user himself check the integrity of some files that he thinks are most important to him but the Third Party Auditor checks for all the files that are stored on the cloud server. In this way the cloud user checks the integrity of some files and also cross check the work of the Third Party Auditor for that file. In the proposed model as a Third Party Auditor is checking the integrity of all the files so the burden on cloud user decreases and also as a user is checking the integrity of some files by his own the threat to data integrity decreases. This model is also the new technique for integrity checking for the users who not want a Third Party Auditor for integrity verification. They can check the integrity of all files with their own.

1.6. Objectives of Dissertation

The objectives of this research work are:

1. To Analyze and implement the model of Third Party Auditor.
2. To propose and analyze a new architecture to check data integrity and Third Party Auditor.

1.7. Dissertation Organization

This section gives the organization of the dissertation. The dissertation is divided into five chapters including this one. The Chapter 1 explains the basic introduction to Cloud Computing including its service and deployment models, essential characteristics, security issues related with it and its advantages and disadvantages. Various schemes for the Third Party Auditor which is an integrity checking technique are explained. This chapter also covers the problem statement, objectives and the proposed work of the dissertation.

The Chapter 2 is Review of literature which provides what different researchers or authors have described about the Cloud Computng, its security issues and Third Party Auditors. The Chapter 3 describes the environment and algorithms used for the implementation of proposed work. The Chapter 4 gives the

implementation details and the results obtained in the form of snapshots. The Chapter 5 concludes the whole work and gives the future scope related to the proposed dissertation.

Cloud Computing is emerging as a big and beneficial technology of the present day and future. Much of work is being put in it and one can expect more progress in cloud computing technology. However, in terms of security it is not so beneficial for medium and small enterprises to adopt cloud computing. Data stored on cloud must be kept confidential and integrated. The Third Party Auditor that is used to the check the integrity of data must not add any new vulnerability to the cloud.

CHAPTER 2

REVIEW OF LITETATURE

Literature survey is the most important step in any research or software development process. It gives a theoretical base for the research and helps us to determine the nature of our research. This chapter presents what different researchers or authors have described about the cloud computing, its security issues and Third Party Auditor.

2.1. Background

The John McCarthy has introduced the concept of cloud computing in the 1960s. He said that "computation may someday be organized as a public utility". The Douglas Parkhill gives the characteristics of cloud computing for the first time in 1966 in his book, “The Challenge of the Computer Utility”. The history of word "cloud" originates from the world of telecommunications when providers use virtual private network (VPN) services for data communications at reasonable prices.

Cloud computing is an extension to cover servers and underlying network infrastructure. Many big companies such as IBM, Google, Amazon, Microsoft, Yahoo and other move their space to develop cloud computing. These companies have launched their own cloud computing infrastructures and cloud services on the basis of their original products and achieved best application results and social impact, such as Amazon’s EC2 and S3, Google’ Google Apps, Microsoft’ Azure and so on (Wang & Mu, 2011).

2.2. Literature Review

2.2.1. Cloud Computing

Vaquero et al. (2009) have explored the concept of cloud computing to give a perfect and complete definition of cloud. They have described the cloud computing using the essential features that are associated with this cloud computing. The authors have studied more than 20 definitions to extract a consensus and minimum definition of cloud computing which covers all the essential characteristics. The paper also gives the similarities and dissimilarities

between the cloud computing and grid computing. The authors have compared the features of grid computing with the main characteristics of cloud computing.

Zhang et al. (2010) have introduced the background, principle, character, style and future of cloud computing. The authors have discussed the various applications and merits of cloud computing. The paper have analyzed some questions and hidden troubles, puts forward some solutions and discussed the future of cloud computing. It is concluded that the clouds will grow in size as soon as available bandwidth and the corresponding service model mature enough, cloud computing will bring a revolutionary change in the Internet.

Jedeja & Modi (2012) have discussed the concept of cloud computing. The authors have described the cloud computing as a computing environment in which computing needs of one party are fulfilled by another party and access to computing power or resources needed are provided via the Internet. The authors have also described the layers, architectural design, deployment and service models, advantages and issues of cloud computing.

Mollah, Islam & Islam (2012) have discussed about platforms, issues and challenges, architecture, advantages, applications, platforms, future and research options of cloud computing. They also have discussed about the four generations of computing such as mainframe based computing, personal computing, client server based computing and web server based computing respectively. As there are several advantages over present web server based computing, such as fast microprocessor, reliable system architecture, huge memory, high-speed network, etc., the cloud computing will provide the next generation of computing services.

Idrissi, Kartit & Marraki (2013) have proposed a better understanding of cloud computing aspects and explore directions for research and technological trends in order to use and implement cloud infrastructure networks. The authors have described the essential characteristics, service and deploy models, various issues like confidentiality, integrity, resource self-monitoring, energy management etc. The papers also present a survey of some of the leading cloud computing products which are launched by different compnies like IBM, Google, Amazon, Microsoft, Yahoo.

2.2.2. Security Issues in Cloud Computing

Wang & Mu (2011) have introduced the concept and characteristics of cloud computing. They have defined the services and deployment models of cloud computing. The authors have described the various security issues like network attacks, data security, lack of standards, privacy. The paper have also summarized some solutions for these issues like strengthen the anti-locking capability, information encryption, uniform safety standards, selecting reputable service providers etc.

Bhel & Bhel (2012) have discussed the various security issues in cloud computing. There are many benefits of cloud computing, but also some issues are also there in cloud computing. The paper have introduced a detailed analysis of the cloud security problem and described the various issues like cloud security, vendor lock-in, multi-tenancy, Service Level Agreement (SLA) management. The authors have examined the security problem on the perspective of the cloud architecture, cloud characteristics, cloud delivery model. Based on various issues the paper presents the key features that any proposed security model should cover.

Piplode & Singh (2012) have provided the overview and study of cloud computing. The authors have included various security and challenging issues of cloud computing. They also have provided some solution for the issues like encryption algorithm, backup, customer satisfaction, investigation support, etc. They have defined the twenty security management models. They have defined the requirements of these security models in cloud that a cloud service provider should adopt to develop their programs.

2.2.3. Third Party Auditor

Ateniese, Burns, Curtmola, Herring, Kissner, Peterson & Song (2007) have included public auditability in their Provable Data Possession (PDP) model to ensure control of files on untrusted cloud storages. They make use of RSA-based homomorphic tags to achieve public auditability to assess outsourced data. The authors claim to verify data integrity with exchange of small data size which is independent of the original data file. However, dynamism in data storage operations is not included in the work. Further, the mechanism may not work if

users' data are deleted, intentionally or accidentally, under the circumstances where cloud server provider is not trusted.

Juels & kalliski (2007) have illustrated a Proof of Retrievability (PoR) model, which makes use of codes of spot-checking and error-correcting for ensuring data possession and retrievability by inserting special blocks randomly into the data to detect unauthorized alteration. The model ensures two aspects of data storage on the cloud that is possession and retrievability. Further, to the special blocks, the data file is encrypted. However, the scheme has some limitations such as a fixed priori, dynamic data updates and not support of public auditability.

Shah, Baker, Mogul & Swaminathan (2008) have proposed a model in which Third Party Auditor (TPA) maintains the data integrity. He first encrypts the data files and then sends a number of pre-computed symmetric-keyed hashes of the encrypted data files to the auditor. The auditors have verified both the integrity of the data file and server's possession of a previously committed decryption key. This model only works when the data files are encrypted. it suffers from the audited statefulness in which when the keyed hashes are used up it puts the burden on users..

Itani, Kayssi & Chehab (2009) have proposed Privacy as a Service (PaaS), a set of security protocols to ensure the privacy and legal issues related to data stored on a cloud, with the help of cryptographic coprocessors as Trusted Third Party (TTP). The authors have categorized the data based on its sensitivity. Temper-proof cryptographic coprocessor are assumed to be physically and logically protected from unauthorized users.

Wang et al. (2009) have proposed public auditability in cloud computing under the circumstances of full dynamic data operations. They aim to achieve public auditability (including batch auditing), support for fully dynamic data operations, blockless verification, efficiency, scalability, security and performance. The authors do not consider the issue of data privacy. They have used the bilinear aggregation signature technique to support multiple auditing tasks and classic Merkle Hash Tree Construction is manipulated for block tag authentication.

Wei, Zhu, Cao, Jia & Vasilakos (2010) have proposed SecCloud, an auditing scheme to secure cloud computing that is based on probabilistic sampling technique to provide secure data storage, computation and privacy preservation. The authors have also demonstrated how to minimize the auditing cost by optimizing the sampling size by considering multiple servers and multiple cloud service providers. Trusted Third Party (TTP) is in the form of designated authority which does the job of data verification through the designated verification protocol. The authors claim to achieve both computing and data security together.

Kumar, Subramnian & Selvam (2010) have proposed an efficient and flexible distribution verification protocol to describe the data storage security problem in cloud computing. The authors rely on erasure code for availability and reliability of data. They have used token pre-computation using a Sobol sequence for integrity checking of erasure coded cloud data. They claim to have better performance and more security against unauthorized data modification attacks. The usage of Third Party Auditor (TPA) is optional.

Wang et al. (2010) have identified the requirement of efficient method that enable on demand data correctness verification on behalf of data owner and propose a publicly auditable cloud data storage system. The authors have highlighted the requirement of Trusted Third Party (TTP) and describe approaches and system requirements for the same along with challenges to be resolved.

Wang et al. (2010) have identified public auditability for data storage security as one of the critical issues. Public key based homomorphic authenticator with random masking is used to achieve the privacy preserving public cloud data auditing system. They give the bilinear aggregate signature technique to handle of multiple auditing tasks. In this multi user setting, Third Party Auditor (TPA) is able to perform multiple auditing tasks simultaneously.

Tian & Wu (2012) aim to address the issues with their Trusted Control Model of Cloud Storage (TCMCS), where authors recommend encrypting the data before sending it to the cloud servers and managing the encryption keys with the help of Third Party Auditor. The model utilizes symmetric, asymmetric encryption algorithms and encoding techniques to protect privacy and integrity of the users' sensitive data.

Mohta, Sahu & Awasthi (2012) have proposed a scheme to protect the data from unauthorized access and to ensure that data are intact which solve the problem of privacy, consistency, integrity and unauthorized access. They first present a network in which cloud architecture, users and Third Party Auditor (TPA) are shown after that they describe how file is retrieved. Then they suggest a scheme for retrieval of file, encryption and decryption of file, how to check the integrity of data from Cloud Service Provider (CSP) and how to give control to Third Party Auditor. Later they had defined the properties that will be given by scheme. Further challenging issues for public auditing services that need to be focused on are discussed too.

Wang et al. (2012) have proposed a flexible distributed storage integrity auditing mechanism which makes use of homomorphic token and distributed erasure-coded data and claims to allow user to audit the data with low computation and communication cost. The authors also take the real situations into account such as the multiple servers and dynamic data operations.

Zhu, Hu, Ahn, Yau (2012) have addressed the construction of an interactive Provable Data Possession (PDP) protocol to prevent the leakage of verified data. The authors also claim the reduction of the cost of auditing per verification. They have implemented abnormal detection through probabilistic queries and periodic verification. The authors also give an accurate method for the selection of an optimal parameter value for the minimization of the computational overheads that occurs during cloud auditing services.

Paigude & Chavan (2013) have proposed a new innovative idea for Privacy Preserving Public Auditing with watermarking for data storage correctness in cloud computing. It supports data dynamics in which various operations like insert, update and delete are performed by the user on the data files. It also supports batch auditing in which requests from multiple users are handled simultaneously to reduce communication and computing cost.

Wang et al. (2013) have proposed a secure cloud storage system which supports the privacy-preserving public auditing. In this model the Third Party Auditor (TPA) is enabled to handle audit requests from multiple users simultaneously and efficiently. Extensive security and performance analysis show the proposed schemes are provably secure and highly efficient. The preliminary

experiment conducted on Amazon EC2 instance further demonstrates the fast performance of the design.

Vinaya & Sumathi (2013) have presented a data model for secure integrity verification scheme and with data update protocol for dynamic data modification by introducing effective Third Party Auditor. They provide a mechanism for data integrity checking in which a cloud user can check the integrity of his data stored on the cloud. This can be agreed upon by the both cloud service provider and cloud user. This mechanism can be incorporated in the Service Level Agreement (SLA).

The literature survey helped to gain a better insight with reference to cloud computing, different service models, deployment models and current security issues of cloud computing. The different schemes proposed by different researchers for the Third Party Auditor has been studied. During the survey, it is noted that a lot of research is going on in cloud computing security issues and how to overcome the security issues and to gain cloud users' confidence in cloud computing.

CHAPTER 3

METHDOLOGY AND ENVIRONMENT

This chapter introduces the various algorithms like RSA and MD5 that are used in implementing the model. It also introduces the environment Visual Studio 2010 that is used to implement the model of Third Party Auditor and Proposed model. Visual Studio is a software development environment which is used to build software. The Technology or Language used in Visual Studio to create the web pages for different entities is ASP.NET with C#.

3.1. Algorithms Used

The RSA algorithm is asymmetric key algorithm that is used to generate the public and private keys and MD5 is the message digest algorithm is used to calculate the hash of the files that is also called a message digest which is used for verification purpose.

3.1.1. RSA

In asymmetric key algorithm each user has a pair of keys: a public key and a private key. Public key is used for encryption and Private key is used for decryption. RSA is most popular asymmetric key algorithm. It is developed by Ron Rivest, Adi Shamir and Len Adleman of MIT in 1977. Since then Rivest-Shamir-Adleman (RSA) scheme has become the most widely adapted and implemented general purpose approach to public key encryption (Zhou & Tang, 2011).

The RSA is based on the fact that it is easy to find and multiply large prime numbers together, but it is difficult to factor their product. The private and public keys in the RSA are based on very large prime numbers. The real challenge in the case of RSA is the selection and generation of public and private key. The RSA scheme is a block cipher. Each plaintext block is an integer between 0 and $n - 1$ for some n. The typical size for n is 1024 bits (Cao & Fu, 2008). The RSA algorithm works in three steps:

- Key Generation
- Encryption
- Decryption

Key Generation:

1. Choose two large prime numbers P and Q.
2. Calculate N=P*Q.
3. Calculate (P-1) * (Q-1)
4. Select the public key E such that it is not the factor of (P-1) and (Q-1)
5. Select the private key D such that equation (D*E) mod (P-1) * (Q-1) = 1 will satisfy.
6. Public Key: (E,N)
7. Private Key: (D,N)

For Example:

1. Let P=7, Q=17
2. N=7*17=119
3. (P-1) * (Q-1)= (7-1)*(17-1) =96
4. The factors of 96 are 2 and 3 as (96=2*2*2*2*2*3). Thus we choose E as 5 which is not factor of 96.
5. D*5 mod 96=1.So we choose D =77 as (77*5) mod 96=1
6. Public Key: (5,119)
7. Private Key: (77,119)

Encryption:

For encryption, calculate the cipher text CT from plain text PT as

CT=PT^E mod N.

For Example:

Let PT=10

CT=10^5 mod 119

100000 mod 119=40

Decryption:

For decryption, calculate the cipher text PT from plain text CT as

PT=CT^D mod N.

For Example:

CT= 40

PT=40^77 mod 119

3080 mod 119=10

3.1.2. MD5

A message-digest algorithm is also called a hash function. It accepts a message as input and generates a fixed-length output, which is generally less than the length of the input message. The output is called a hash value, a fingerprint or a message digest (Jarvinen, Tommiska Skytta, 2005).

Properties of a message-digest algorithm:

- It should be one-way. Given the message digest, it is hard to get the original message for which the message digest was created.
- Given both input and output, it is difficult to find another input message which generates the same output.
- It should be collision-resistant. It is computationally infeasible to find two messages, which generates same message digest. This property is not same as the second property. It is easier to make attack on this property than on the second property.

MD5 is a message digest algorithm developed by Professor Ronald L. Rivest in 1991. MD5 message-digest algorithm takes as input a message of arbitrary length and produces as output a 128-bit "fingerprint" or "message digest" of the input. The input message is processed in 512-bit blocks. The output is a set of four 32 bit blocks, which make up the 128-bit message digest. The MD5 algorithm is intended for digital signature applications, where a large file must be "compressed" in a secure manner before being encrypted with a private (secret) key under a public-key cryptosystem such as RSA (Cao & Yang, 2010).

Process of MD5: The process of MD5 consists of following five steps:

Step1. Append padding bits:

- The input message is padded so that its length (in bits) equals to 448 mod 512. Padding is always performed, even if the length of the message has been already 448 mod 512.
- Padding is performed as follows: a single "1" bit is appended to the message and then "0" bits are appended so that the length in bits of the padded message becomes congruent to 448 mod 512. At least one bit and at most 512 bits are appended.

Step2. Append length:

- A 64-bit representation of the length of the message is appended to the result of step1. If the length of the message is greater than 2^64, only the low-order 64 bits will be used.
- The resulting message has a length that is an exact multiple of 512 bits. The input message will have a length that is an exact multiple of 16 (32-bit) words.

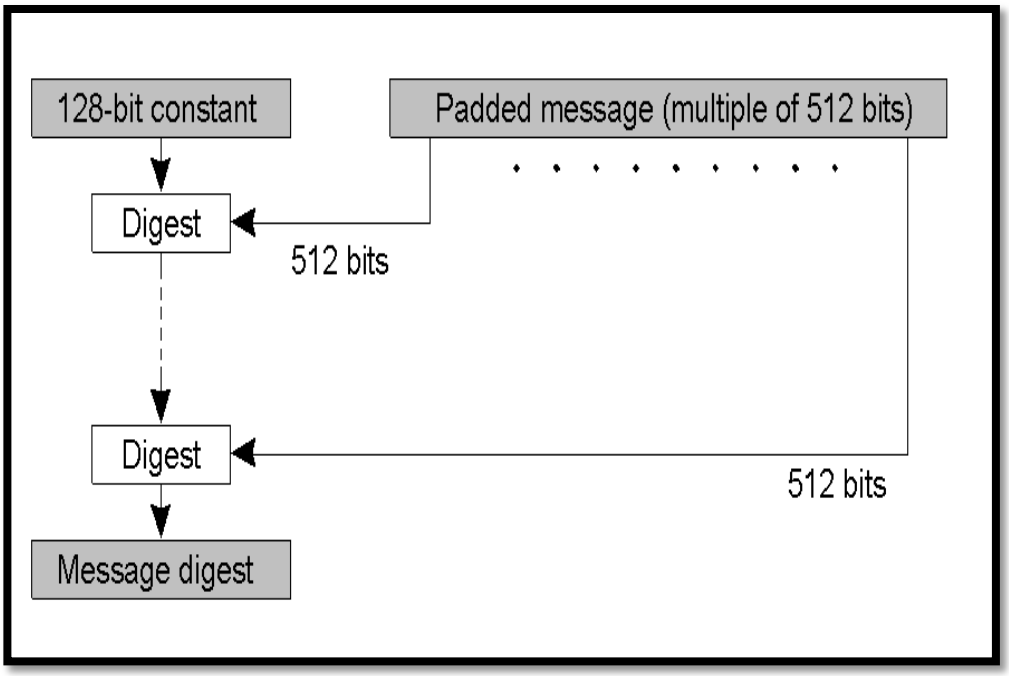

Fig. 3.1: Process of MD5

Step3. Divide the input into 512-bit blocks:

- Now divide the input message into blocks and the length of each block is 512 bits.

Step4. Initialize MD buffer

- A four-word buffer (A, B, C, D) is used to compute the message digest. Each of A, B, C, D is a 32-bit register. These registers are initialized to the following values in hexadecimal, low-order bytes first):
  - word A: 01 23 45 67
  - word B: 89 ab cd ef
  - word C: fe dc ba 98
  - word D: 76 54 32 10

Step5. Process blocks

- First copy the four MD buffers into four corresponding variables a, b, c, d.
- Divide the current 512-bit block into 16 sub blocks. Thus each sub block contains 32 bits.
- Now there are four rounds. In each round all the 16 sub blocks belonging to a block are processed.
- The input to the each round is
  - all the 16 sub block
  - the variables a, b, c, d
  - some constants as t

A single MD5 operation is: a = b + ((a + F (b, c, d) + X [k] + T [i]) <<< s).

3.2. Introduction to Visual Studio 2010

Visual Studio is a software development environment (also known as an Integrated Development Environment or IDE). Software Developers uses this to build websites, Software products and utilities. Visual Studio is a collection of the tools and services to create a wide variety of apps, both for the Microsoft platform and beyond. Visual Studio also connects all projects, teams and stakeholders. Teams can work with greater agility from virtually anywhere irrespective of development tool, including Eclipse and Xcode. Designing mission-critical .NET apps, writing blazing fast code with C++ AMP, or testing and debugging a cloud-connected HTML/JavaScript app that runs on many devices, join millions of developers worldwide in choosing Visual Studio as user essential development environment.

Visual Studio IDE:

The new project window allows choosing an application template from the available templates. When one starts a new web site, ASP.NET provides the starting folders and files for the site, including two files for the first web form of the site. The file named Default.aspx contains the HTML and asp code that defines the form and the file named Default.aspx.cs (for C# coding) or the file named Default.aspx.vb (for vb coding) contains the code in the language that have chosen and this code is responsible for the form's works. The primary window in the Visual Studio IDE is the Web Forms Designer window. Other supporting

windows are the Toolbox, the Solution Explorer, and the Properties window. We use the designer to design a web form, to add code to the control on the form so that the form works according to need, using the code editor.

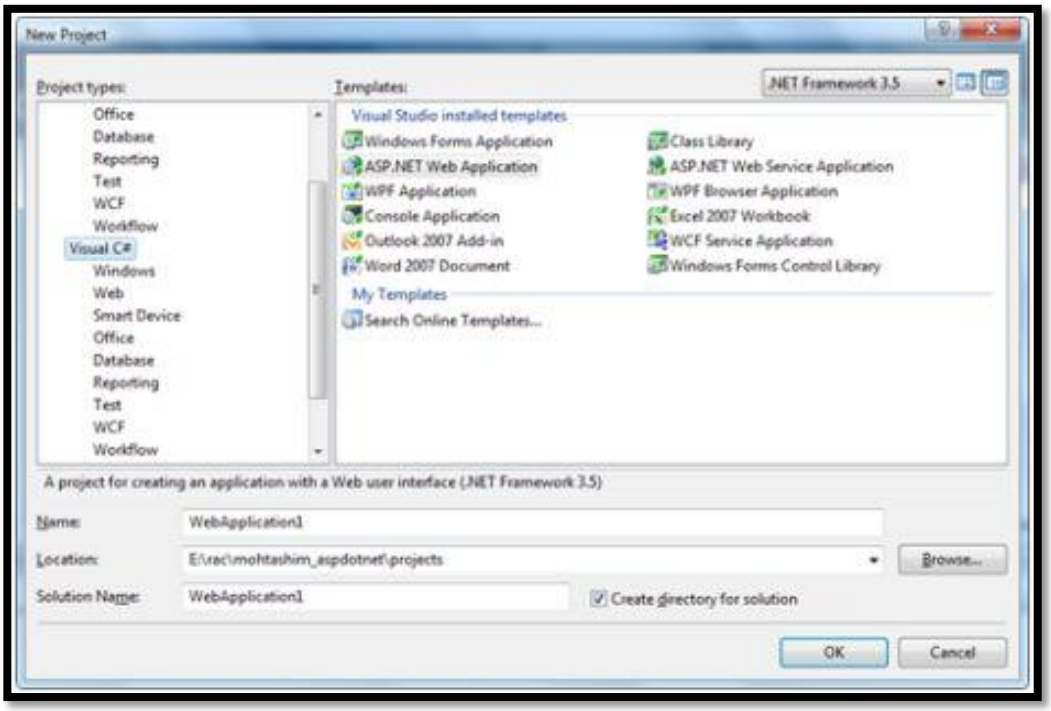

Fig. 3.2: Visual Studio IDE

Adding folders and files to web site:

When a new web form is created, Visual Studio automatically generates the starting HTML for the form and displays it in source view of the web forms designer. The Solution Explorer is used to add any other files, folders or any existing item on the web site.

- To add a standard folder, right-click on the project or folder under which we are going to add the folder in the Solution Explorer and choose New Folder.
- To add an ASP.Net folder, right-click on the project in the Solution Explorer and select the folder from the list.
- To add an existing item to the site, right-click on the project or folder under which we are going to add the item in the Solution Explorer and select from the dialog box.

Projects and Solutions:

A typical ASP.Net application consists of many items: the web content files (.aspx), source files (e.g., the .cs files), assemblies (e.g., the .dll files and .exe

files), data source files (e.g., .mdb files), references, icons, user controls and miscellaneous other files and folders. All these files that make up the website are contained in a Solution. Solutions may contain one or more projects. A project contains content files, source files, and other files like data sources and image files. Generally the contents of a project are compiled into an assembly as an executable file (.exe) or a dynamic link library (.dll) file.

Building and Running a Project:

The application is run by selecting either Start or Start without Debugging from the Debug menu, or by pressing F5 or Ctrl-F5. The program is built i.e. the .exe or the .dll files are generated by selecting a command from the Build menu.

3.3. Introduction to ASP.NET

ASP.NET is a technology that allows users to build and control dynamic Web pages easily. It also provides many enhancements to take advantage of new technology as one can interact with databases, personalize Web pages for visitors, display page on mobile devices and even build an entire e-commerce site from scratch. It is a web development platform, which provides a programming model, a comprehensive software infrastructure and various services required to build up robust web application for PC, as well as mobile devices.

ASP.NET works on top of the HTTP protocol and uses the HTTP commands and policies to set a browser-to-server two-way communication and cooperation. It is a part the of Microsoft .Net platform. Its applications are complied codes, written using the extensible and reusable components or objects present in .Net framework. These codes can use the entire hierarchy of classes in .Net framework.ASP.NET is used to produce interactive, data-driven web applications over the internet. It consists of a large number of controls like text boxes, buttons and labels for assembling, configuring and manipulating code to create HTML pages.

The ASP.NET application codes could be written in either of the following languages:

- C#
- Visual Basic .Net
- Jscript

- J#

ASP.NET Web Forms Model:

ASP.NET web forms extend the event-driven model of interaction to the web applications. The browser submits a web form to the web server and the server returns a full markup page or HTML page in response. All client side user activities are forwarded to the server for stateful processing. The server processes the output of the client actions and triggers the reactions. Now, HTTP is a stateless protocol. ASP.NET framework helps in storing the information regarding the state of the application, which consists of:

- Page state
- Session state

The page state is the state of the client, i.e., the content of various input fields in the web form. The session state is the collective obtained from various pages the user visited and worked with, i.e., the overall session state. The ASP.NET runtime carries the page state to and from the server across page requests while generating the ASP.NET runtime codes and incorporates the state of the server side components in hidden fields.

ASP.NET Component Model:

The ASP.NET component model provides various building blocks of ASP.NET pages. Basically, it is an object model, which describes:

- Server side counterparts of almost all HTML elements or tags, like <form> and <input>.
- Server controls, which help in developing complex user-interface for example the Calendar control or the Gridview control.

ASP.NET is a technology, which works in the .Net framework that contains all web-related functionalities. The .Net framework is made of an object-oriented hierarchy. An ASP.NET web application is made of pages. When a user requests an ASP.NET page, the IIS delegates the processing of the page to the ASP.NET runtime system. The ASP.NET runtime transforms the .aspx page into an instance of a class, which inherits from the base class Page of the .Net framework. Therefore, each ASP.NET page is an object and all its components, i.e. the server-sisde controls are also objects.

CHAPTER 4

RESULTS & DISCUSSION

This chapter provides the results obtained after implementation of model of Third Party Auditor and proposed model. The system is implemented by building the project using Visual Studio 2010. The languages used in programming the system are ASP.NET with C#. Using the ASP.NET with C# web pages each concern with a specific job in the system: the cloud user who stores his data on the cloud server, Cloud service provider who provides space to store user data, Third Party Auditor who check the integrity of data on behalf of the user are created.

4.1. Login page of Admin or Cloud Service Provider

Fig 4.1 is the login page where different users login by entering their username and password. Firstly admin login to the login page by providing his username and the password. The admin plays most important role in the system it activates all the users that sign up to the cloud.

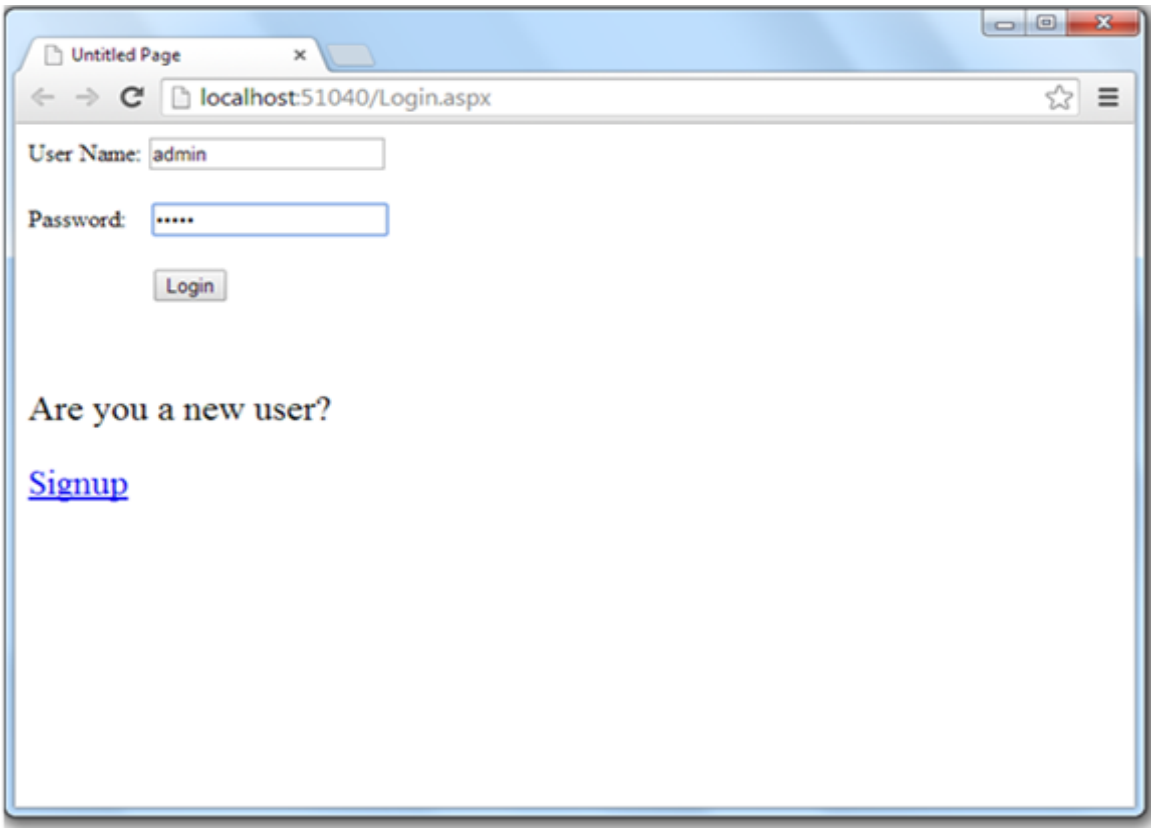

Fig. 4.1: Admin Login Page

4.2. Home page of Admin or Cloud Service Provider

Fig 4.2 is the admin home page it contains the information about the different users who sign up to the cloud server along with their role. The different users who want to store their data to the cloud server first sign up to the cloud server. The admin activates the users who sign up. By clicking on the select button of any user it shows the different files that are stored by that user on the cloud server. As shown in the Fig. 4.2 clicking the select button of user ‘renuka’ it shows the files stored by the ‘renuka’ on the cloud

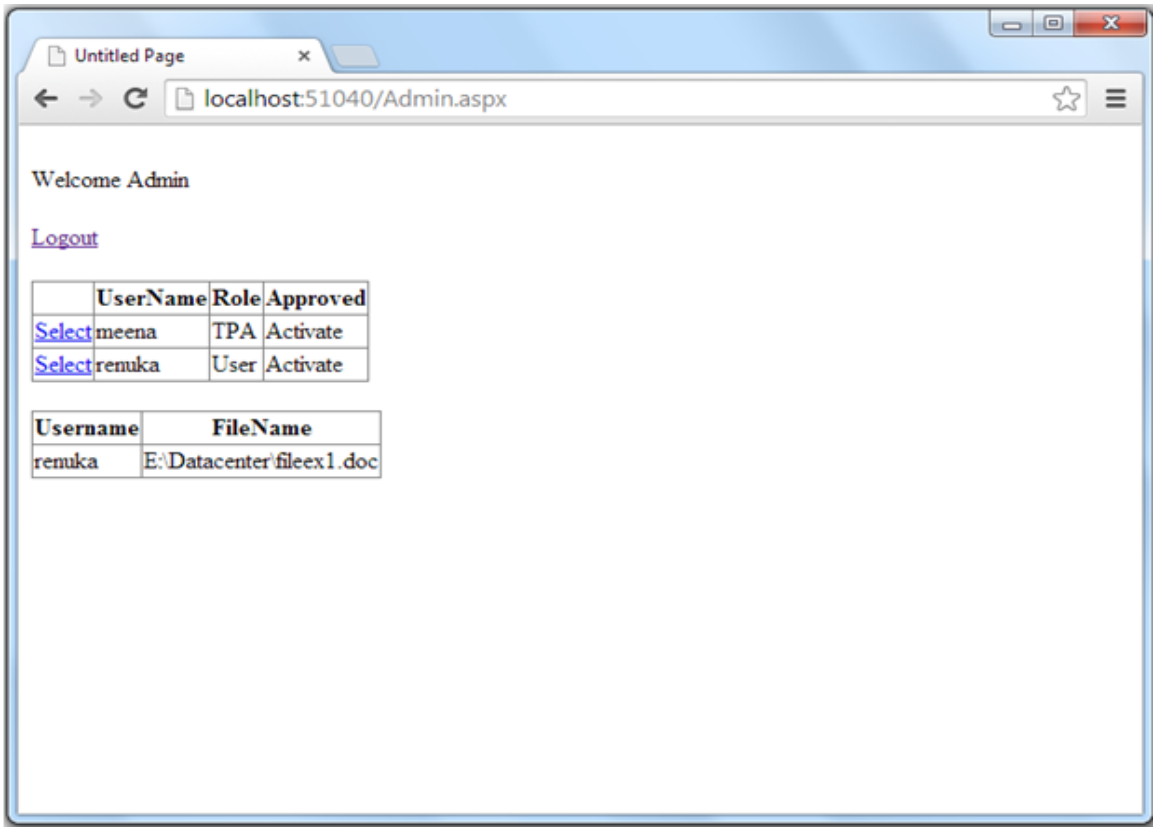

Fig. 4.2: Admin Home Page

4.3. Sign up page of Cloud User

Fig 4.3 is the user sign up page. The user first registers to the cloud server to get the services by signing up. As user click on sign up, he goes to sign up page where he enters his personal details like first name, last name, middle name, email id, address, Date of Birth, etc. He chooses the user name and password which he later uses to login into the system. He selects his role whether it is cloud user or Third Party Auditor. Then he submits all his details. He can login to the system after the cloud service provider activated him.

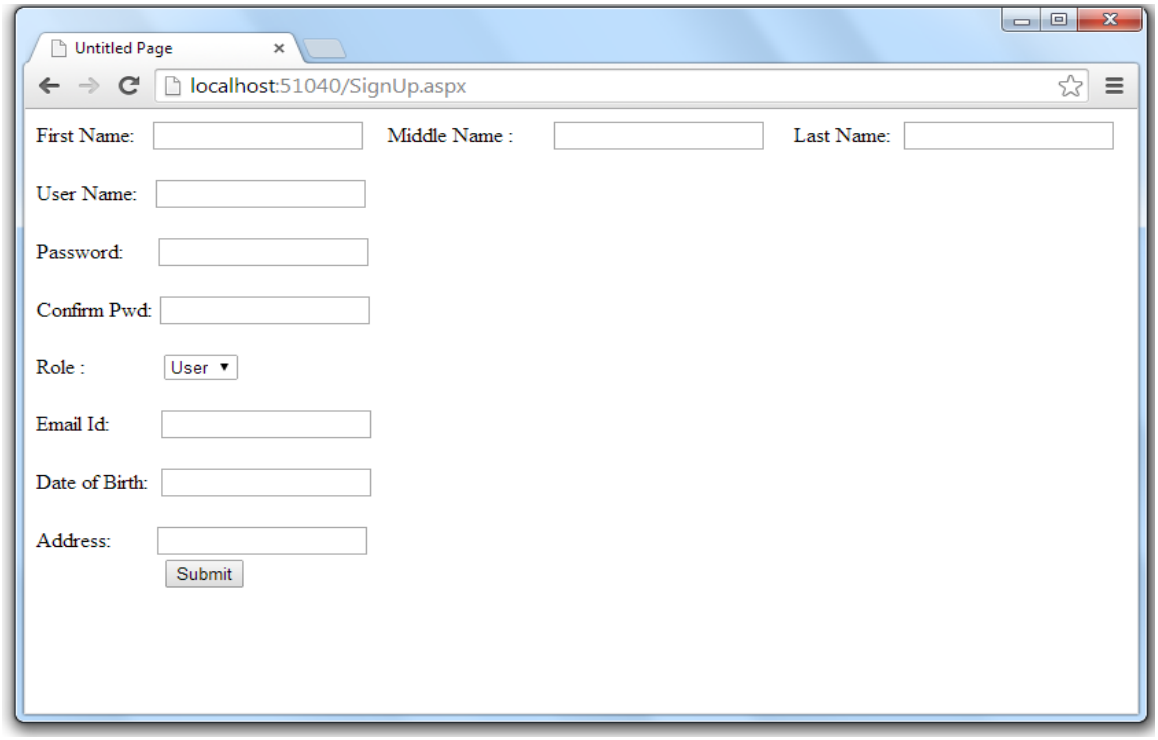

Fig. 4.3: User Sign up Page

4.4. Login page of Cloud User

Fig 4.4 is the user login page. The users login to the cloud server by using the username and password which he chooses earlier.

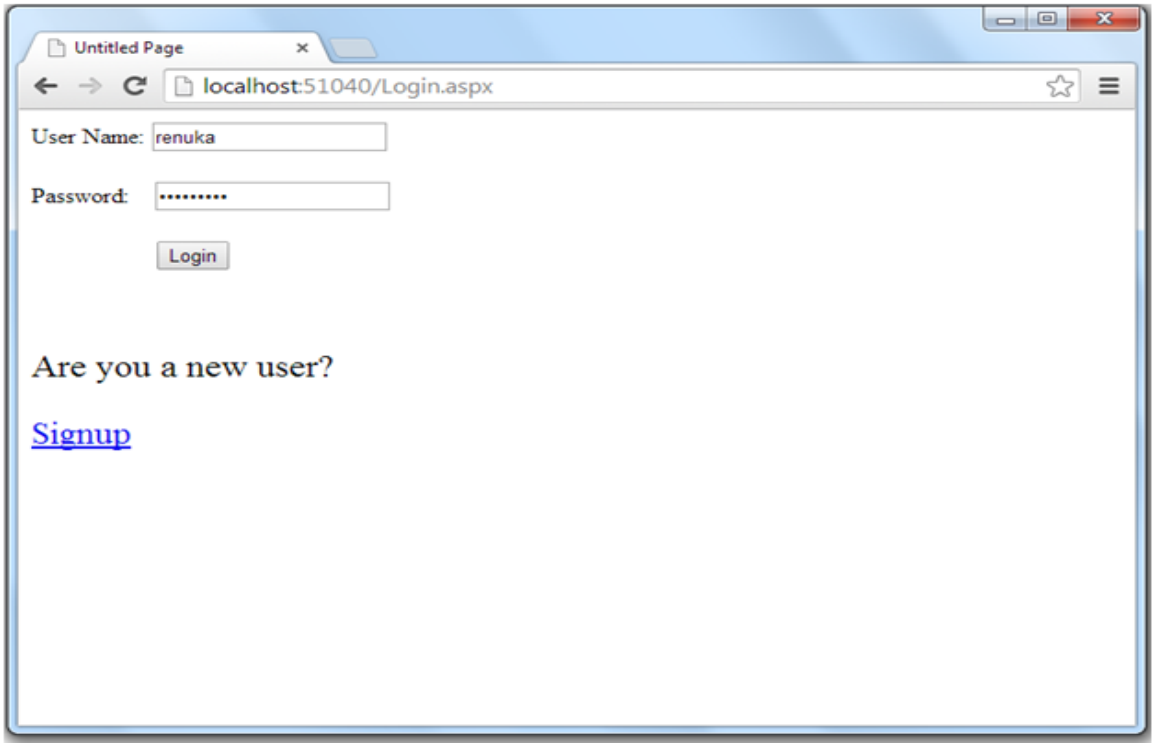

Fig. 4.4: User Login Page

4.5. Home page of Cloud User

Fig 4.5 is the home page of the user. After the user login to the system he goes to his home page. The home page of user contains the list of files that he stored on the cloud server. Each file has a status as authenticated or unauthenicated based on its integrity as a check by the Third Party Auditor. When Third Party Auditor checks the integrity of a file if it is correct, then status of the file becomes authenticated, otherwise if the file get corrupted then status becomes unauthenticated.

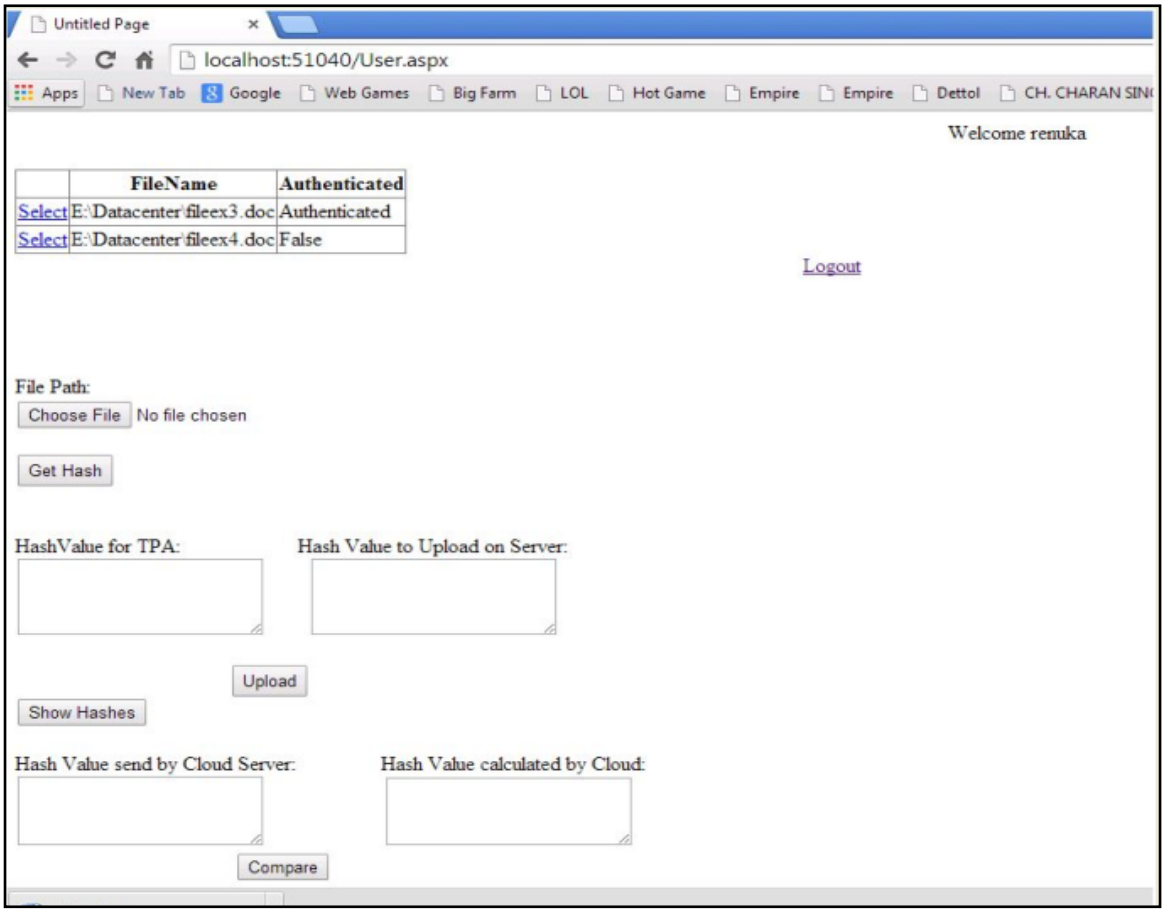

Fig 4.5: User Home Page

4.6. Home page of Cloud User showing the Hash values

Fig.4.6 is also the home page of the user. Here the hash values calculated by user are shown in the textboxes. The user first chooses the file that he wants to store on the cloud server. Then user calculates the hash of the file using the “Get Hash” button before uploading it to the cloud server. The user calculates the two different hash values of the file. One hash value is calculated to store it on cloud

server along with the original file. Another hash value is calculated to stored on the database of Third Party Auditor (TPA) for the data integrity verification purpose. After the hash values get calculated, user encrypts both the hash values with his private key. When the user clicks on the “upload” button the original file is getting stored into the data center of the cloud server. The data center of cloud server is created in E drive of the computer. A message “File Uploaded successfully” is displayed.

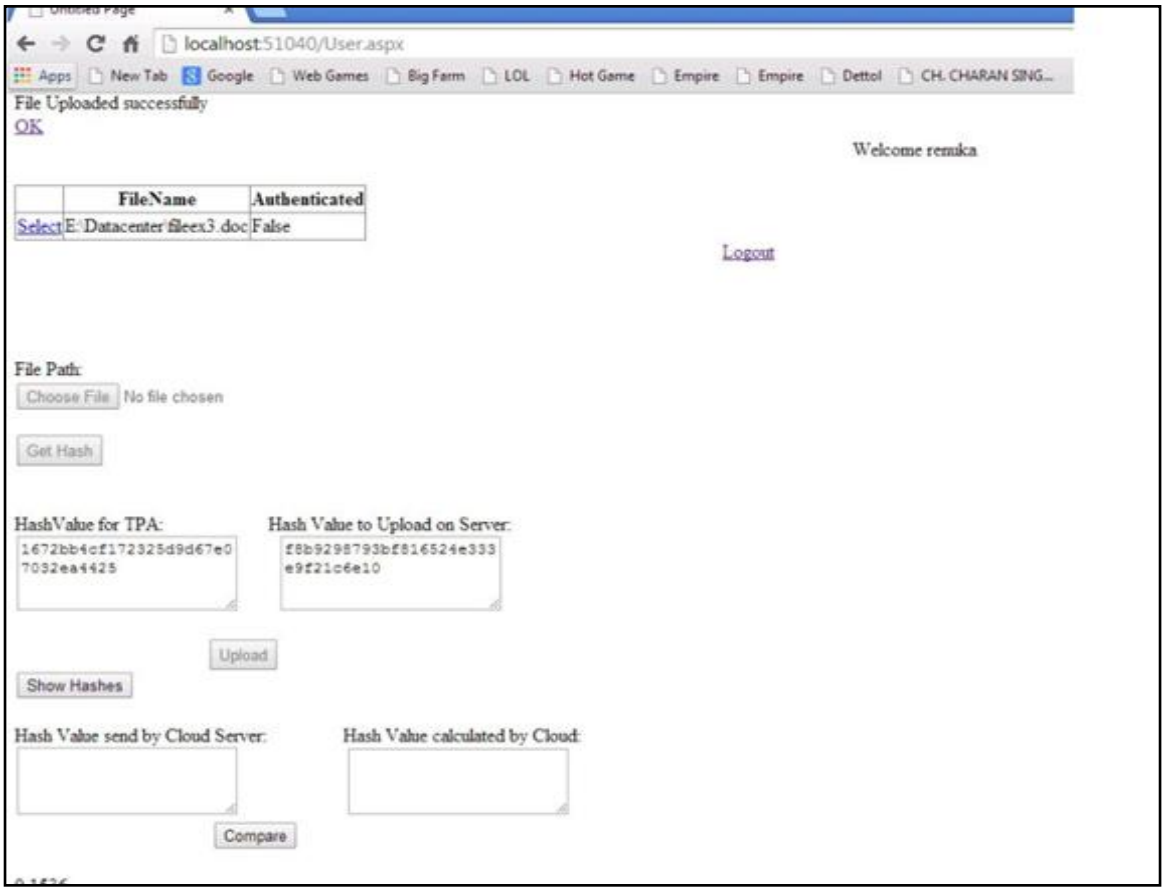

Fig. 4.6: User Home Page showing Hash values

4.7. Login page of Third Party Auditor

Fig 4.7 is the Third Party Auditor login page. Third Party Auditor (TPA) who can periodically check the integrity of all the data stored in the cloud on behalf of user and releases his audit reports which help the user in checking the status of his files whether the file is corrupted or file is not corrupted. The Third Party Auditors login to the cloud by using the username and password which he specifies earlier.

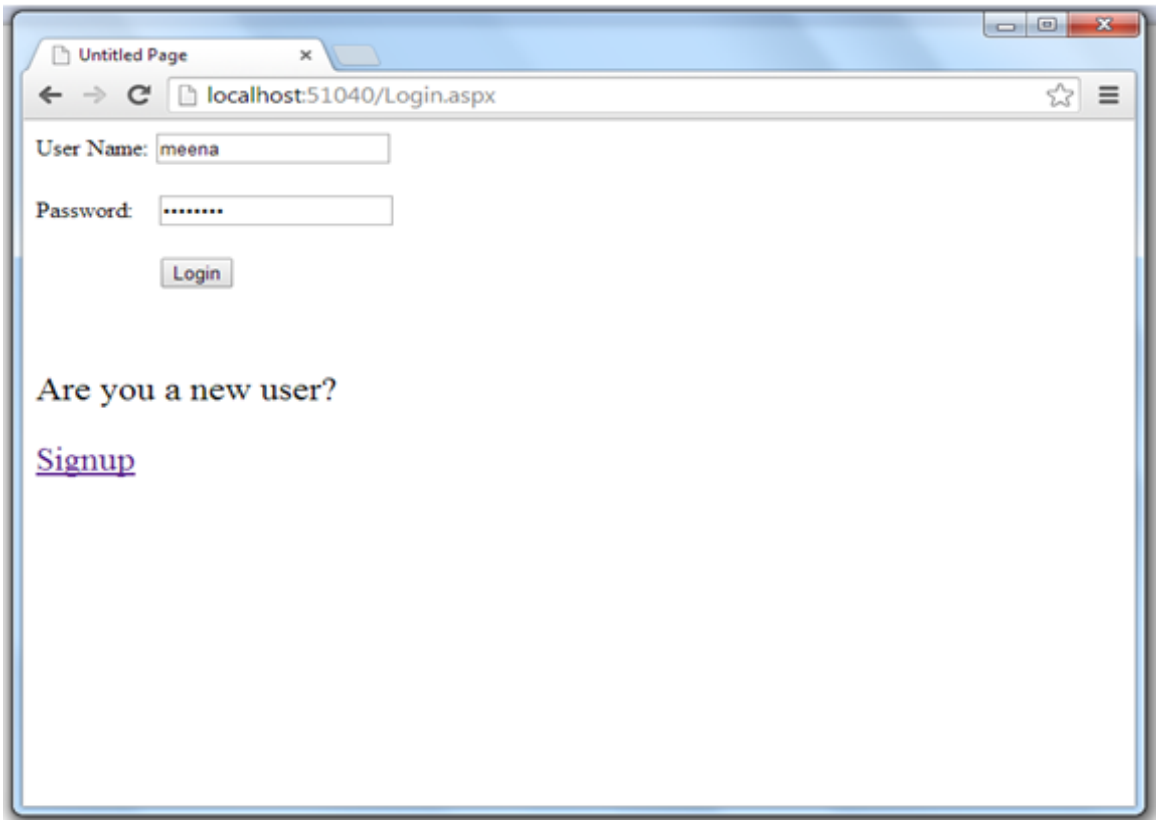

Fig. 4.7: TPA Login Page

4.8. Home page of Third Party Auditor

Fig.4.8 is home page of Third Party Auditor. When the user wants to check the integrity of any data he asks the Third Party Auditor to check the integrity of that file. The Third Party Auditor then asks the cloud service provider to calculate the hash of that file. The home page of Third party Auditor contains the list of files of users whose integrity have to check by the user. First the Third Party Auditor selects the File to check its integrity by clicking on the select.

As the Third Party Auditor clicks on the select the cloud service provider calculated the hash of selected file and send to the Third Party Auditor which is displayed in the textbox “Hash from CSP”. Then the Third Party Auditor click on the button “Decrypt the hash sent by user” the hash value stored in the Third Party Auditor database after decryption gets displayed in the textbox “Decrypted Hash”. The Third Party Auditor clicks on the “compare” button to check the integrity. The compare button compares the two hash values obtained. The Third Party Auditor displays whether the file is corrupted or file is not corrupted on the basis of hash values.

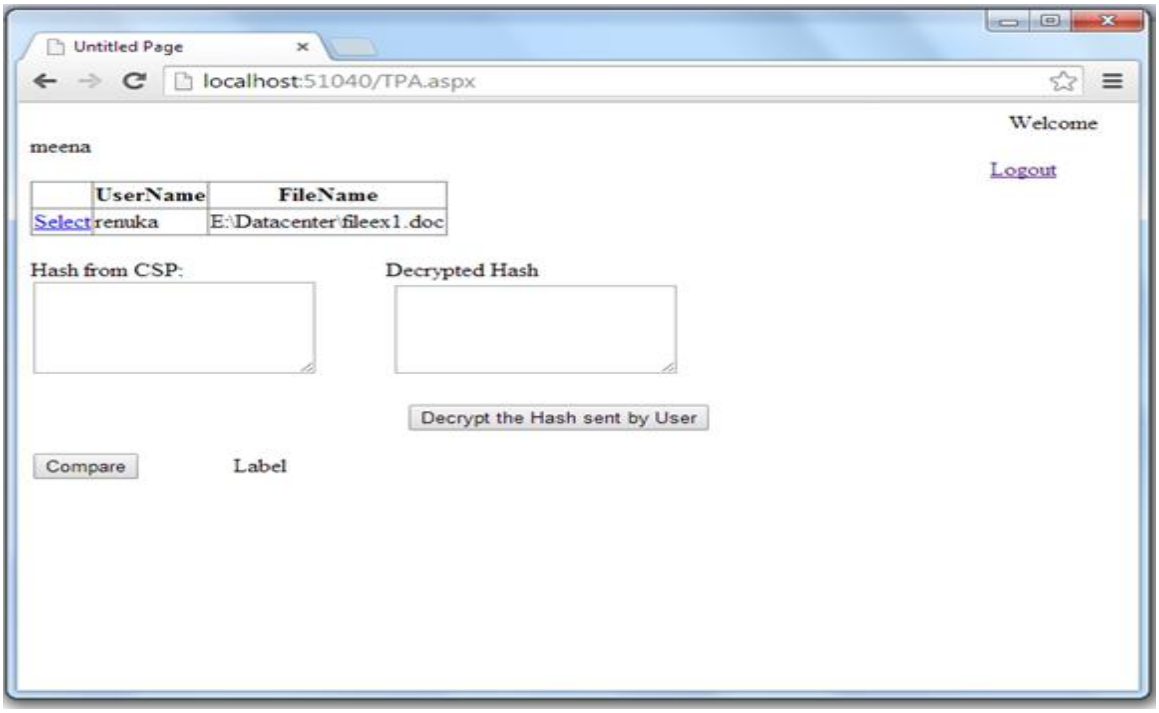

Fig 4.8: TPA Home Page

4.9. Home page of TPA showing file is not corrupted

Fig 4.9 shows the comparison result when hash values match. When TPA clicks on “compare” button, “The file is not corrupted” is displayed on the label and also authenticated is marked against the filename in user page.

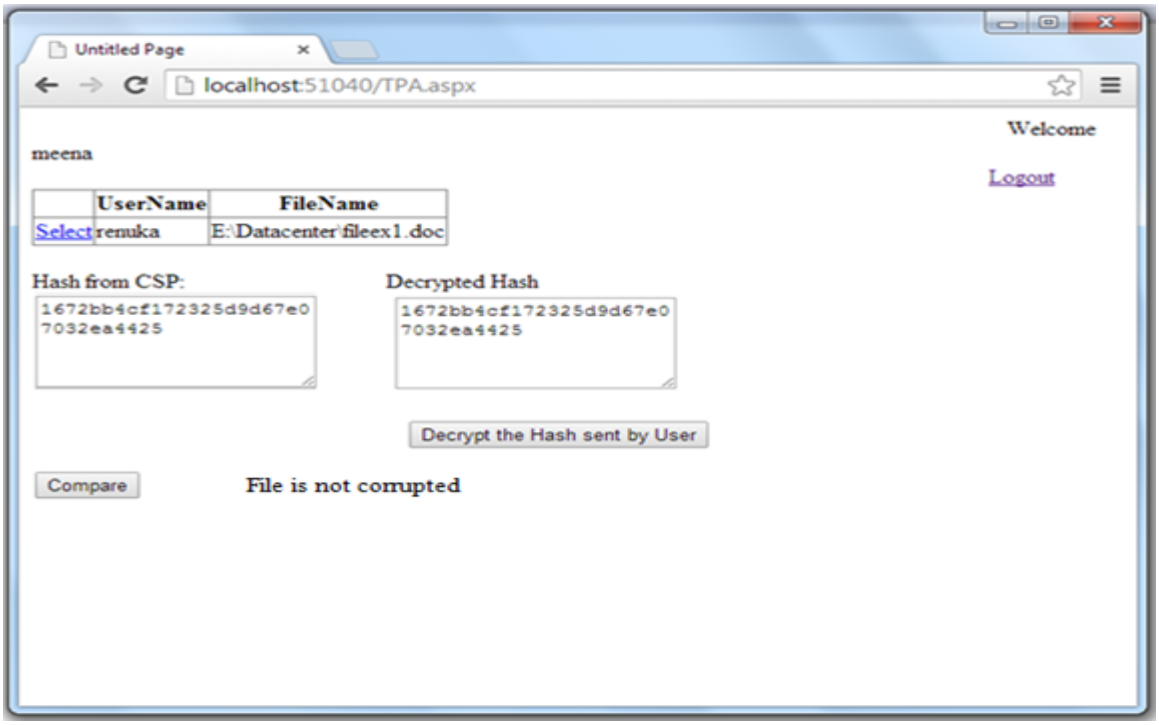

Fig. 4.9: TPA Home Page showing file not corrupted

4.10. Home page of TPA showing file is corrupted

Fig 4.10 shows the comparison result when hash values do not match. When TPA clicks on “compare” button, “The file is corrupted” is displayed on the label and also authenticated is marked against the filename in user page.

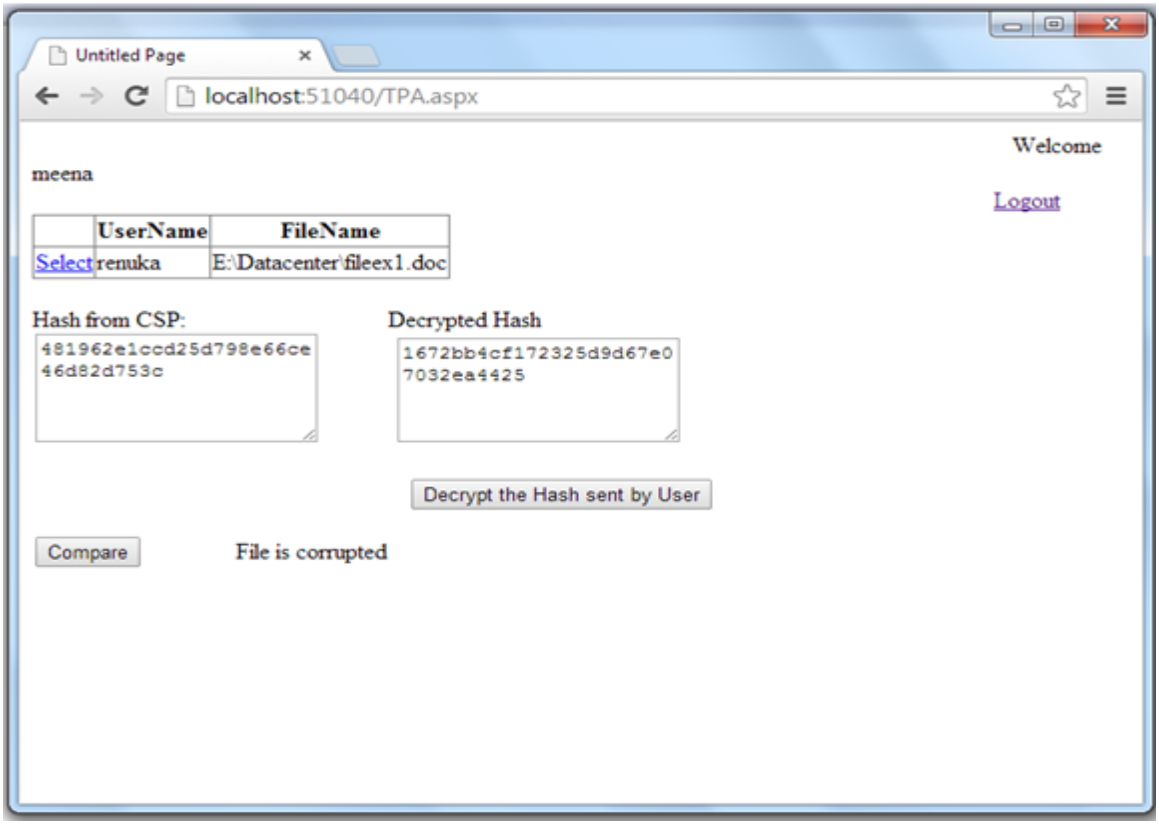

Fig 4.10: TPA Home Page showing file corrupted

4.11. Home page of user showing integrity verification process

Fig. 4.10 shows the integrity checking process. To verify data integrity and Third Party Auditor user compares the hash values. Firstly user asks the cloud server to calculate the hash of the file whose integrity has to check. The cloud server provider calculates the hash of the file and sent it to the cloud user that is shown in the textbox “HashValue calculated by the Cloud”. The user then asks the cloud server provider to send the hash which is stored on it by the user of that file that is shown in the textbox “HashValue send by Cloud server”. User click on the compare buttons to check the integrity. If the hash values get matched then the file is not corrupted and if the hash values do not get matched the file on the server get corrupted.

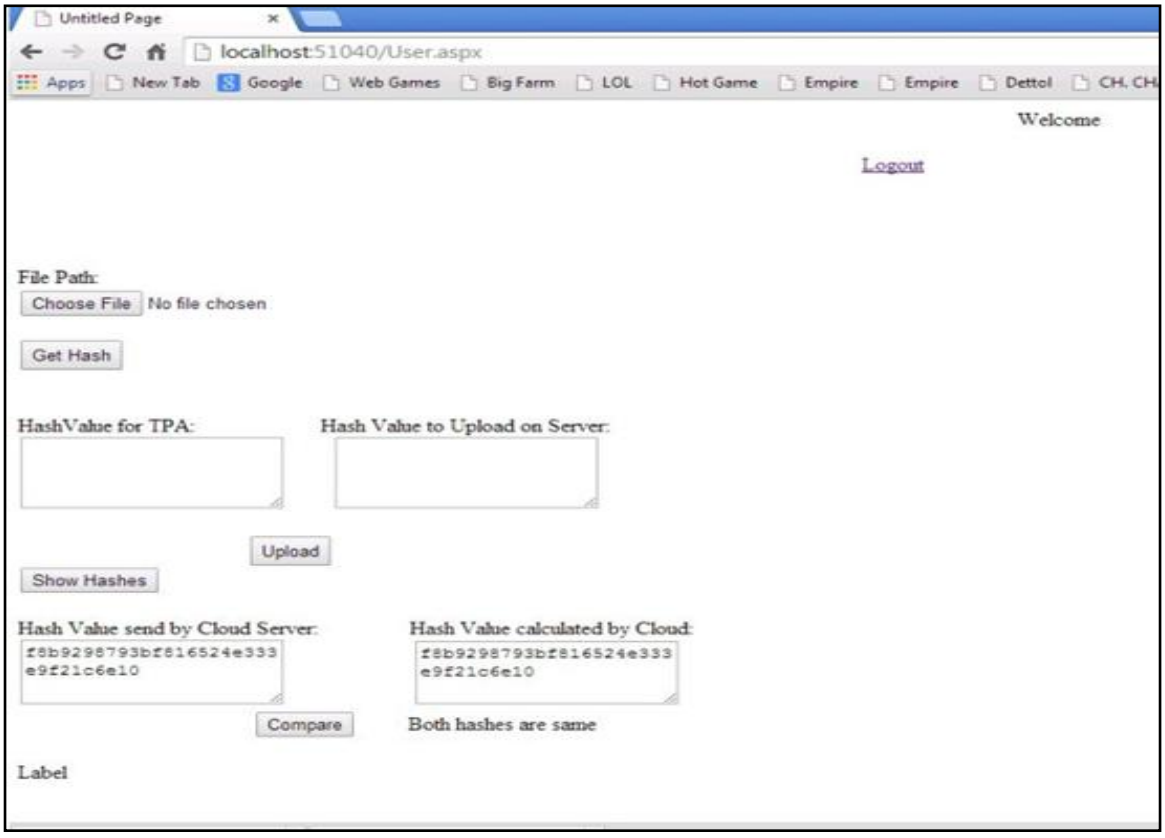

Fig. 4.7: User Home Page showing integrity verification process

The system designed to give a good result based on the different situations. It provides correct Integrity checking results depending upon the user data that is stored in the cloud server. With the pages which have been created, we make the system easy to use by the general user having background in any field.

CHAPTER 5

CONCLUSIONS AND FUTURE SCOPE

5.1. Conclusions

Cloud Computing is emerging as a big and beneficial technology of the present day and future. It provides many benefits to his customers. But there are many security issues associated with the cloud computing. The main security issue is the threat to data integrity. The user store their data to cloud they do not know the exact location of their data because of which risk to data integrity increases.

Cloud service provider and other users can manipulate the data stored on the cloud for their own benefit. So to check the integrity of data user takes the assistance of a third party auditor who checks the integrity of data on behalf of the user. The auditing report generated helps the user to know the status of its data that is stored on the cloud. It helps the user to verify and examine the data from unauthorized people that manipulate the data. But there is a possibility that the Third Party Auditor may also cheat the cloud user.

The research work focus on the verification of data integrity and Third Party Auditor at client side. First the model for Third Party Auditor is implemented to check the integrity. The proposed model is implemented to verify the data integrity and Third Party Auditor. The system is developed using Visual Studio 2010 in ASP.NET with C# language. It is found that the system implemented provides the best result based on the data stored on the cloud. It can be concluded that threat to data integrity can be decreased by checking it on user side.

5.2. Future Scope

A number of open problems must be solved to minimize the threat to data integrity. These problems suggest a variety of research directions that need to be pursued to make such a system feasible.

- In the implementation we have used only the text data stored by the user. We can improve the implementation by making the user enable to enter other kinds of data like audio and video.

- It would be preferable to check over the ability to edit or delete the data in the cloud.
- The verification process could be done by further techniques rather than the using hash values to improve the efficiency and security.

REFERENCES

Ateniese, G., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., and Song, D. (2007, October). Provable data possession at untrusted stores. In Proceedings of the 14th ACM conference on Computer and communications security. pp.598-609.

Attas, D., and Batrafi, O. (2011). Efficient integrity checking technique for securing client data in cloud computing. International journal of electrical & computer sciences. 11**(05):** 43-48.

Balakrishnan, S., Saranya, G., Shobana, S., and Karthikeyan, S. (2011, June). Introducing Effective Third Party Auditing (TPA) for Data Storage Security in Cloud. International Journal of Computer Science and Technology. 2**(2):** 397-400.

Behl, A. (2011, December). Emerging security challenges in cloud computing: An insight to cloud security challenges and their mitigation. In Proceedings of IEEE World Congress on Information and Communication Technologies. pp.217-222. Mumbai.

Behl, A., and Behl, K. (2012, October). An analysis of cloud computing security issues. In Proceedings of World Congress on Information and Communication Technologies. pp.109-114.Trivandrum.

Bhagat, A., and Sahu, R.K. (2013, March). Using Third Party Auditor for Cloud Data Security: A Review. International Journal of Advanced Research in Computer Science and Software Engineering. 3**(3):** 34-39.

Cao, Y., and Fu, C. (2008, October). An Efficient Implementation of RSA Digital Signature Algorithm. In Proceedings of IEEE International Conference on Computer Science and Electronics Engineering. pp.647-651. Hangzhou.

Chen, D., and Zhao, H. (2012, March). Data security and privacy protection issues in cloud computing. In Proceedings of IEEE International Conference on Intelligent Computation Technology and Automation. pp.100-103. Hunan.

Dillon, T., Wu, C., and Chang, E. (2010, April).Cloud computing: Issues and challenges. In Proceedings of 24th IEEE International Conference on Advanced Information Networking and Applications. pp.27-33. Perth,WA.

Govinda, K., Gurunathaprasad, V., and Sathishkumar, H. (August, 2012). Third party auditing for secure data storage in cloud through digital signature using

RSA. International journal of advanced scientific and technical research. 4**(2):** 525-530.

Gowrigolla, B., Sivaji, S., and Masillamani, M. R. (2010, December). Design and auditing of cloud computing security. In Proceedings of 5th International Conference on Information and Automation for Sustainability. pp.292-297.

Han, S., and Xing, J. (2011, September). Ensuring data storage security through a novel third party auditor scheme in cloud computing. In Proceedings of IEEE International Conference on Cloud Computing and Intelligence Systems. pp.264-268. Beijing.

Idrissi, H. K., Kartit, A., and El Marram, M. (2013, April). A taxonomy and survey of Cloud computing. In Proceedings of National Security days. pp.1-5.Rabat.

Itani, W., Kayssi, A., and Chehab, A. (2009, December). Privacy as a service: Privacy-aware data storage and processing in cloud computing architectures. In Proceedings of Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing. pp.711-716. Chengdu.

Jadeja, Y., and Modi, K. (2012, March). Cloud computing-concepts, architecture and challenges. In Proceedings of IEEE International Conference on Computing, Electronics and Electrical Technologies. pp.877-880. Kumaracoil.

Juels, A., and Kaliski Jr, B. S. (2007, October). PORs: Proofs of retrievability for large files. In Proceedings of the 14th ACM conference on Computer and communications security. pp.584-597.

Kuyoro, S. O., Ibikunle, F., and Awodele, O. (2011). Cloud Computing Security Issues and Challenges. International Journal of Computer Networks. 3**(5):** 247-254.

Luo, W., and Bai, G. (2011, September). Ensuring the data integrity in cloud data storage. In Proceedings of IEEE International Conference on Cloud Computing and Intelligence Systems. pp. 240-243. Beijing.

Mahmood, Z. (2011, September). Data Location and Security Issues in Cloud Computing. In proceedings of IEEE International Conference on Emerging Intelligent Data and Web Technologies. pp.49-54. Tirana.

Makhija, B., Gupta, V.K., and Rajput, I. (2013, February). Enhanced Data Security in Cloud Computing with Third Party Auditor. International Journal of Advanced Research in Computer Science and Software Engineering. 3**(2):** 341-345.

Mohta, A., Sahu, R.K., and Awasthi, L.K. (2012, February). Robust Data Security for Cloud while using Third Party Auditor. International Journal of Advanced Research in Computer Science and Software Engineering. 2**(2):** 1-5.

Mollah, M. B., Islam, K. R., and Islam, S. S. (2012, April). Next generation of computing through cloud computing technology. In Proceedings of 25th IEEE Canadian Conference on Electrical & Computer Engineering. pp.1-6. Montreal.

Muneshwara, M.S., and Chandarl, A.T. (2012, June). Monitoring the integrity of Dynamic Data Stored in Cloud Computing. International Journal of Engineering Research & Technology. 1**(4):** 1-8.

Paigude, T., and Chavan, T. A. (2013). A survey on Privacy Preserving Public Auditing for Data Storage Security. International Journal of Computer Trends and Technology. 4**(3):**412-417.

Patel, H., and Patel, D., (2012). A Review of Approaches to Achieve Data Storage Correctness in Cloud Computing using Trusted Third Party Auditor. International symposium on Cloud and Services Computing.84-27.

Piplode, R., and Singh, U.M. (2012, September). An Overview and Study of Security Issues & Challenges in Cloud Computing. International Journal of Advanced Research in Computer Science and Software Engineering. 2**(9):** 115-120.

Shah, M. A., Baker, M., Mogul, J. C., and Swaminathan, R. (2007, May). Auditing to Keep Online Storage Services Honest. *In proc of HotOS'07*. 1-6.

Syam Kumar, P., Subramanian, R., and Thamizh Selvam, D. (2010, October). Ensuring data storage security in cloud computing using Sobol sequence. In Proceedings of IEEE International Conference on Parallel Distributed and Grid Computing. pp. 217-222. Solan.

Tian, J., and Wu, Z. (2012, March). A Trusted Control Model of Cloud Storage. In proceedings of International Conference on Computer Distributed Control and Intelligent Environmental Monitoring. pp.78-81. Hunan.

Tianfield, H. (2012, October). Security issues in cloud computing. In proceedings of IEEE International Conference on Systems, Man, and Cybernetics. pp.1082-1089. Seoul.

Vaquero, L. M., Rodero-Merino, L., Caceres, J., and Lindner, M. (2009, January). A break in the clouds: towards a cloud definition. ACM SIGCOMM Computer Communication Review. 39**(1):** 50-55.

Vinaya, V. and Sumathi, P. (2013, May). Implementation of Effective Third Party Auditing for Data Security in Cloud. International Journal of Advanced Research in Computer Science and Software Engineering. 3**(5):** 382-387.

Wang, C., Ren, K., Lou, W., and Li, J. (2010). Toward publicly auditable secure cloud data storage services. IEEE Network. 24**(4):** 19-24.

Wang, C., Wang, Q., Chow, S. M., Ren, K., and Lou, W. (2013, Februry). Privacy-preserving public auditing for secure cloud storage. IEEE transaction on computers . 2**(2):** 362-374.

Wang, C., Wang, Q., Ren, K., and Lou, W. (2010, March). Privacy-preserving public auditing for data storage security in cloud computing. In Proceedings of IEEE INFOCOM. pp.1-9. San Diego.

Wang, C., Wang, Q., Ren, K., Cao, N., and Lou, W. (2012). Toward secure and dependable storage services in cloud computing. IEEE Transactions on Services Computing. 5**(2):** 220-232.

Wang, J. J., and Mu, S. (2011, September). Security issues and countermeasures in cloud computing. In Proceedings of IEEE International Conference on Grey Systems and Intelligent Services. pp. 843-846. Nanjing.

Wang, Q., Wang, C., Ren, K., Lou, W., and Li, J. (2011). Enabling public auditability and data dynamics for storage security in cloud computing. IEEE Transactions on Parallel and Distributed Systems. 22**(5):** 847-859.

Wang, Z. (2011, October). Security and privacy issues within the Cloud Computing. In Proceedings of IEEE International Conference on Computational and Information Sciences. pp.175-178.Chengdu.

Wei, L., Zhu, H., Cao, Z., Jia, W., and Vasilakos, A. V. (2010).SecCloud: Bridging Secure Storage and Computation in Cloud. In Proceedings of International Conference on Distributed Computing Systems Workshops. pp.52-61,.

Xiao, Z., and Xiao, Y. (2013). Security and Privacy in Cloud Computing. IEEE Communications Surveys & Tutorials. 15**(2):** 843-856.

Yang, K., and Jia, X. (2012). An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing. IEEE Transactions on Parallel & Distributed Systems. 24**(9):** 1717-1726.

You, P., Peng, Y., Liu, W., and Xue, S. (2012, June). Security issues and solutions in cloud computing. In Proceedings of IEEE 32nd International

Conference on Distributed Computing Systems Workshops. pp.573-577. Macau.

Zhang, S., Zhang, S., Chen, X., and Huo, X. (2010, January). Cloud computing research and development trend. In Proceedings of Second International Conference on Future Networks. pp.93-97. Sanya.

Zhou, X., and Tang, X. (2011, August) Research and Implementation of RSA algorithm for encryption and Decryption. In Proceedings of sixth International Forum on Strategic Technology. pp.1118-1121. Harbin.

Zhu, Y., Hu, H., Ahn, G. J., and Yau, S. S. (2012). Efficient audit service outsourcing for data integrity in clouds. Journal of Systems and Software. 85**(5):** 1083-1095.