CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION TO CLOUD COMPUTING

In recent days, Cloud storage has become a good entrant for organizations that suffer from resource limitation. Cloud computing is a technique that follows internet-based computing. The cloud computing technique is used to reduce data management cost and time. In addition, cloud computing is used to store data that can be accessed remotely. The most challenging task in the cloud is to ensure availability, integrity, and secure file transfer.

The motivation for cloud computing was needed for complex-intensive applications run by large-scale organizations like governments. Those organizations require more computational, network, and storage resources than a single computer. Using cloud computing, data owners transmit the data through cloud servers to the respective users. The use of cloud computing technique affects the security of the transmission of data. The encryption and decryption techniques to transmit data safely through the cloud servers (Brandenburger et al. 2015). Data owners encrypt the data using the encryption algorithms and forward the data to the cloud servers. After encryption, the data is transmitted to the cloud servers where the data cannot be accessed directly and transmitted to the respective users using specific searching methods.
The most famous definition of cloud computing is proposed by the USA National Institute of Standards and Technology (NIST). NIST tells us that cloud computing is a model for enabling convenient, on demand network access to a shared pool of alterable computing resources which can be highly provisioned and released with minimal management effort or service provider interaction (Bhosale et al. 2012). The recent development on cloud computing has a value added concern on data security due to the factors such as unreliability of the service and malicious attacks from hackers. Recently a lot of events on server corruption are reported with major cloud infrastructure providers. Breaching of data from important cloud services also appear from time to time. Besides, the cloud service provider voluntarily scrutinizes the customers data for various motivations. Therefore, the cloud is intrinsically neither secure nor reliable from the customer’s point of view. Without providing strong security, privacy and reliability guarantee, it would be difficult to expect cloud customers to provide the control of their data to cloud servers in spite of economic savings and flexibility of service.

1.1.1 Distributed vs Grid Computing

Distributed computing is the technique allows multiple computing systems to perform the complex task. Distributed Computing typically alludes to overseeing or pooling the hundreds or a large number of PC frameworks which separately are more restricted in their memory and processing efficiency.

Whereas grid computing, has some additional attributes which uses of a pool of heterogeneous frameworks with ideal workload administration using an undertaking's whole computational assets (servers, networking systems, storage, and data) acting together to make at least one vast pools of the computing resources. There is no impediment of clients, divisions or beginnings in grid computing.
Grid computing is centered around the capacity to bolster calculation over different managerial areas that separates it from conventional distributed computing. Grid offers a method for utilizing the data innovation assets ideally inside an association including virtualization of processing resources. Its idea of support for numerous regulatory arrangements and security confirmation and approval systems empowers it to be appropriated over a neighborhood, metropolitan, or wide-zone organize.

**Fog computing**

Fog computing, also known as fog networking or fogging, is a decentralized computing infrastructure in which data, compute, storage and applications are distributed in the most logical, efficient place between the data source and the cloud. Fog computing essentially extends cloud computing and services to the edge of the network, bringing the advantages and power of the cloud closer to where data is created and acted upon. The fog extends the cloud to be closer to the things that produce and act on IoT data. These devices, called fog nodes, can be deployed anywhere with a network connection. Any device with computing, storage, and network connectivity can be a fog node. Examples include industrial controllers, switches, routers, embedded servers, and video surveillance cameras.

**Fog and Cloud Computing**

In Fog computing, processing and applications are concentrated in devices at the network edge rather than transfer to cloud for processing. So all processing is done at smart devices in the network not in the cloud. In Mobile cloud computing the mobile devices and cloud computing combine to create a new infrastructure and data processing and data storage are outside of mobile devices (at the cloud). The goal of fogging is to improve efficiency and reduce the amount of data transported to the cloud for processing, analysis and storage. This is often done to improve efficiency, though it may also be
used for security and compliance reasons. Popular fog computing applications include smart grid, smart city, smart buildings, vehicle networks and software-defined networks.

1.1.2 Evolution of Cloud Computing

At present the cloud computing is the primary technology is seen as the important tool for improving IT economics. The root of the cloud computing is started from 1950s when mainframe computers came in picture. In mainframe computers multiple users can access the centralized computer through dumb terminal. Mainframe computers provides large amount of storage apace with high processing power.

In 1970, IBM introduced virtual machine, allows multiple operating systems run simultaneously to perform different operations. In virtual machine has the host operating system used to run multiple instances of guest operating system called as virtual machine instances. Each instances have their own infrastructure and memory.

In 1990 telecom operators introduced Virtual Private Network (VPN) for sharing single physical infrastructure. The VPN is mainly used in data transmission through private network lines. This VPN uses encryption to provide more security than traditional internet.

In 1999, Salesforce.com presented a concept to deliver enterprise application via website. Next important development was Amazon web service that allows computers on rent to the users to run their operations. Then Amazon launched Elastic Cloud (EC2) in 2009, provides SaaS (Software as a Service).

The main contribution to cloud computing is provided by Microsoft and Google. They introduced killer apps. Other key technologies that enabled
cloud computing as emerging are Universal high speed bandwidth, software interoperability standards and virtualization hypervisors.

Based on its use, Cloud computing performs different roles such as, cloud host that performs hosting, cloud server that acts as servers and the cloud storage that provides storage space. By using cloud computing, efficiency can be increased with lesser complexity. The main contribution of the thesis is to verify the availability and integrity of the cloud storage and to ensure the secure data transmission with minimal searching time and cost.

1.1.3 Challenges in Cloud Computing

To ensure the security of the stored and transmitted data, information is more important in the cloud environment. The idea of the cloud is putting away the data from Personal Computers and exchanging the data in a secured manner in a cloud (El-Booz et al. 2016). It is very important to understand the security measures from the cloud providers. In summary, the following challenges must be met while ensuring security in a cloud:

- Availability of data storage.
- Third party unauthorized attack.
- Encryption methods.
- File searching and sharing.
- Actual hardware where data is stored.

Cloud computing is a special type of computing technique that follows web based computing. In order to minimize the data management expenses and time, the cloud computing technique is being used (Joshi et al. 2010). Here, the data will be transmitted by the data owners through cloud
servers through the users. The security of the transmission of data gets affected in cloud computing when multiple users share the centralized cloud servers. To transmit the data safely through the cloud servers, encryption and decryption techniques are used.

A. Challenges in searching of data in cloud computing environment

The data is encrypted by the data owners using any of the encryption algorithms and then the data is forwarded to the cloud servers. The encrypted data is stored in the cloud servers from which the data will be accessed by the users with the help of the decryption techniques (Goh et al. 2013).

This encryption algorithm contains public key, private key and key generation. Public key is shared to the cloud servers to retrieve and forward the data. Private Key is shared to the users and the data will be decrypted by them using it. Keys such as private key and public key will be generated by this encryption algorithm.

The data is transmitted to the cloud servers after encryption where the data will be inaccessible but can be shared to the respective users. The users apply specific searching method to search data from the cloud. There are plenty of searching methods available to search data. They are,

- Single keyword
- Multi keyword
- Fuzzy keyword
- Synonym keyword
• Boolean keyword

All these searching methods used search keywords based on their own methods of searching. For example single keyword utilizes a single character to search the data; multi keyword utilizes multi words; synonym keyword utilizes the similar words; fuzzy keyword makes use of the fuzzy logic condition and the Boolean keyword makes use of the Boolean conditions to search their respective data.

Barde et al. (2014) suggested the fuzzy keyword search needs the exact match while searching. For an instance if the original word is cat and the searching word is cats, the result will be unmatched. Here only one character is different from the original word but still it will be left behind irrespective of the singular or plural form of the word. This method of searching will increase the efficiency of searching but in turn will definitely affect the search cost consumed. Multi keyword incorporated with ranking search makes use of the multi keywords to search the data.

B. Challenges in storing of data in cloud computing environment

The significance of cloud computing is to store the data and that data can be accessed in remote areas. Because of this data storage, securing the message is the most important challenge in cloud computing (Chen et al. 2015). The user has no control over the data which is stored in the remote servers. Hence securing the data files from the cloud is the most challenging task in cloud computing.

As all the data have to be stored in a data storage platform without any authenticity. If the data can be accessed by any of the third parties it leads to insecure communication. Data leakage is highly possible in the cloud storage systems. In order to minimize this the verifier is introduced to verify
whether data is secure or insecure. By employing the verifier, cloud users feel the secure data storage from cloud providers (Gajra et al. 2015). But the difficulty is cloud users do not feel comfortable with their personal data. If a third party verifies their data, they feel uncomfortable because of their privacy. Therefore verifiers are not suitable for the cloud storage system.

The main aim of the cloud computing is to store the customer’s data with the integrity which occupies more space. Due to the inadequacy of space, customer’s data is not stored in cloud storage system which thereby leads to performance degradation in the cloud computing. Depending upon the customer’s needs, large amount of space is required to store their data. So availability is another concern in cloud computing. These are the two major requirements in cloud computing, one is availability and another is integrity. After verifying these two requirements one can store the data in cloud storage system and thereby enhance the performance of the cloud computing. Without the help of verifier and third party access, cloud providers have to maintain the integrity and availability in cloud computing.

One pivotal service is data storage. Even though many organizations and enterprises would opt to keep their data in their own infrastructure due to trust and security reasons, the usage of cloud on data storage and sharing is highly popular (Zhu et al. 2012). System providers who are involved provide different forms of security of system and files, but they are not always considered enough as there are various requirements depending on the users. SECaaS, Security as a Service, is a service model that is rare and fairly new. It is, however, expected to be frequent in the near future (Yaar et al. 2004).

Juels et al. (2016) suggested the remote verification protocol for cloud computing. All the data is stored in the cloud storage and this data is converted into file processing module. This file processing module is used to
encode all the data files, which is stored in the cloud storage system. This system decodes the encrypted file and sends it to their respective client. Chen et al. (2016) explained several encoding techniques to encrypt the data. Wang et al. (2016) present the identity-based public key cryptography technique to avoid the Diffie-Hellman problem in cloud computing. This technique is efficient and flexible. Here the number of clients are used to store their data in cloud storage server. By using the authentication problem, data integrity and verification process is done for public cloud and remote cloud.

1.1.4 Applications

Cloud computing, as discussed earlier, is a kind of internet computing. The internet is commonly imagined as clouds. Therefore the term “cloud computing” for computation is done through the internet (Mell et al. 2011). With the help of cloud computing, users can easily access database resources via the internet from anywhere, as long as they need, without worrying too much about any maintenance or management of actual resources.

Databases in cloud are very scalable and dynamic. Cloud Computing is very much different from grid computing, utility computing, or autonomic computing (Ram Govind et al. 2010). In fact, in terms of computing it is a very independent platform. The suitable example for cloud computing is Google apps where any application can be accessed with the help of a browser and it can be deployed on thousands of computer through the internet. It also caters facilities for users to develop, deploy and manage their applications on the cloud.

The present benefits of CC usage for SC processes are insufficient on a profound theoretical basis because of the present research being at an early stage in both theory and practice which makes a negative influence on
its expansion over the business prospective (Marston et al. 2011). There is, however, a general consensus that CC has an IT scope and commercial administration relevance. Currently, confusing amount of CC options and concepts confronts with companies that need to integrate these two aspects when it comes to selectivity, usage limit, adaptivity. Zhu et al. (2010) proposed the hybrid CC to transfer a patient’s vital data to the respective hospital and so he will be able get the required assistance for the treatment of chronic diseases with the aid of cloud computing. This scheme is called as self-management scheme for the home-based healthcare conditions.

As every service does today, file and document services are also sent to Cloud; even for enterprises. Huge amount of research and solutions are being used for Cloud Security, including Cloud based sharing of file/document. However, securing the documents/files in the cloud cryptographically is rarely considered. The ones that consider, lack some aspects such as modularity, usability etc (Wale 2011). Therefore, the steps taken to protect content for enterprises must be integrated to the Cloud.

The main goal of many organizations is to originate a solution that provides sharing secure documents and files on a cloud environment. This aim contains designing and developing a state of the art solution which would offer the following aspects,

- Extensibility
- Flexibility
- Effective and efficient secure environment.

Furthermore, cloud security knowledge must be implemented on the server with full perfection. It involves isolating and separating different institutions’ files, users, settings, keys etc. both management-wise and
security-wise (Willcocks et al. 2011). The solution must also be simple to integrate within any SECaaS platform where much security functions can be shared. The shared functionalities in this level are Policy services, Certificate Authority, authentication services and IDMS and since security is the main aim of this system CIA (Confidentiality, Integrity and Availability) model considered as the basic validation context of security.

Mobile cloud computing (MCC) is also the important application of cloud computing. MCC incorporates the features of cloud computing. Mobile Cloud Computing (MCC) is the blending of two Computing Technologies:

1) Mobile Computing

2) Cloud computing.

Cloud computing portrays both the applications sent as services over the Internet. It provides the applications as the services via the internet. Mobile cloud computing enjoys the benefit of the merits and services of Cloud Computing. Mobile Cloud Computing provides infrastructure where both secure data storage and computationally intensive of mobile devices are migrated cloud servers (Manjunath et al. 2013).

1.2 CHARACTERISTICS OF CLOUD COMPUTING

Cloud computing is defined as a unique model for enabling convenient, on demand network access to a shared pool of configurable computing sources like servers, networks, applications, storage and services that can be provisioned rapidly and released with very little management efforts (Mell&Grance 2011). (Huth&Cebula 2011). Services rendered by cloud computing are listed as follows

- Hardware as a Service (HaaS)
• Platform as a Service (PaaS)

• Software as a Service (SaaS)

Amazon, Google, Microsoft, IBM are some the important companies in cloud computing. Recently a lot of users outsource their data to the websites organized by these companies. According to IDC the overall expenses on software, storage structures, and licensed business by the public cloud service providers will raise up at a 21.9% Compound Annual Growth Rate to $12.2 billion in 2016 (Jensen et al. 2009).

Another service of cloud computing is to provide data storage. It allows business organizations and individuals to transfer their data from personal data centers to cloud based data servers. Shifting data into the cloud servers gives much contentment to organizations and individuals as they need not concentrate much about the management of complex hardware systems. However, on dropping the data ownership, it brings security and privacy issues with data. Without data security, success of cloud computing is difficult.

Maintaining data integrity is one of the crucial security concerns. By outsourcing data, the data owner gives right to cloud service provider for performing any operation on data. Therefore data owner suffers from data loss regarding its possession. Data possession states the control of data. If the data is on local systems then the data owner has ultimate control over any operation performed on data including insertion, block deletion and modification. But if the data is on cloud storage server then cloud provider has the utmost rights to control any operation been done on the data. Cloud provider can hamper any operation on data, process any operation incorrectly and may also produce incorrect results.
The major problem in data possession loss is that the cloud provider can hide such mistakes from data owner for certain benefits. The cloud server may also face internal and external security problems including software bugs, components failure and administration problems which can harms the critical data. This third party controlling of data has endangered data integrity and thereby hindering successful adoption of cloud environment by organizations and individuals (Chow et al. 2009).

Scrutinizing data integrity when accessed is quite common for assuring data possession, but considering the amount of data stored at cloud, checking data integrity when accessed is not of great efficiency. The cloud providers or the data owners for auditing data integrity as there is no guarantee for neutral auditing. Also, in these complex and voluminous data storage systems, the data may be freshened up from time to time. Hence the data auditing protocols devised for static data archives is not suitable for data auditing at present. Here, an authoritarian auditing service is required for auditing data integrity in cloud periodically. In recent years scrutinizing data integrity gained a lot of attention of researchers (Kamara&lauter 2010).

1.3 CLOUD DEPLOYMENT MODEL

The cloud computing is mainly based on the concept of resource sharing which tends to decrease the capital cost, maintenance cost and increase the amount of space availability. Thus cloud environment can be deployed based on the user requirements and need for security. The cloud deployment model provides four types of clouds to the remote users in cloud, they are as follows:

A. Public cloud
B. Private cloud
C. Community cloud and
D. Hybrid cloud.

Public cloud is defined as the cloud accessible to entire world through internet based on certain standards. The private cloud is given as the cloud within an organization’s network. The community cloud is only among companies and it is owned privately. The hybrid cloud is the combination of two or more types of cloud (Hasan & Zawood 2015). Multi clouds are implemented to safeguard the data of clients. It is the cloud of clouds for improving the strength of storage security (Hu et al 2015). A novel approach is used to store, retrieve and forward data in the cloud recently. It uses secure erasure code to ensure data security and encryption mechanisms for forwarding data to other legitimate users.

Hybrid Cloud is defined as a composition of two or more different cloud infrastructures. For Example, a private cloud and a public cloud can be described as a type of multi-cloud that links a set of diversified cloud with respect to their deployment designs.

As the cloud is becoming more and more popular, there are growing security concerns now a days. The security concerns have paved to a lot of researchers have proposed protocols and techniques to cloud data security. The cloud service providers provide a complete security of cloud data. However, as the cloud is accessed through Internet, lot of research has been done on storage security in cloud. In distributed verification, protocols are invented for ensuring data storage security in cloud computing. This is achieved by implementation of a distributed auditing mechanism which ensures the data dynamics of all cloud users are tested and ensured for integrity. In a third party, auditing mechanism is implemented for securing cloud storage (Guo et al. 2015).
A provable data possession (Tenzin et al. 2013) concept ensures that cloud environment works cooperatively and secure the data. In security to cloud, data is rendered using Sobol Sequence. To avoid the security and privacy issues within the cloud, Shankarwar et al. (2015) suggested a novel cloud environment. This way of CC will work in various levels like cloud storage, hosting and sometimes also as the mainframe servers depending on the user’s needs. The proposed technique is mainly focusing on how to save the data from the third party users. Samad et al. (2015) discussed about the mobile cloud computing scheme and their categorization. This scheme overcomes the risks and related issues occurring in the mobile clouds. Ranjan et al. (2015) suggested that the computing remedy for software as well as hardware resources where the debate on the two web services such as amazon and Microsoft was being done.

1.4 CLOUD SERVICE MODEL

With the development of new technologies like Web Services and Virtualization, cloud computing became a reality. With cloud computing people can obtain three kinds of services such as

- Platform as a Service
- Software as a Service
- Infrastructure as a Service.

1.5 ISSUES IN CLOUD COMPUTING

Cloud storage has been highly prevalent because of its advantages. Currently, commercial cloud storage services like Microsoft Skydrive, Amazon S3 and Google Cloud Storage have attracted millions of users. Cloud storage provides not only the massive computing infrastructure but also the
economics of scale. Under such a trend, it becomes necessary to assure the quality of data storage services which involves two frequent concerns from both cloud users and cloud service providers. They are,

- Data integrity
- Availability
- Storage efficiency
- Secure file transfer
- Searching files from cloud storage

1.5.1 Data Integrity and Availability

Cloud users (i.e., data owners) shall always be able to check the integrity of the file at any time. For storage efficiency, it is better to duplicate both the file and the metadata (e.g., authentication tags) needed for data integrity check. Taking malicious users on cloud servers into consideration, the cloud server must verify that the user actually owns the file before creating a link to his file; the user must also confirm that the cloud actually has the file in its storage and must audit the integrity of the file throughout its lifetime.

Nowadays, a huge amount of sensitive data, such as financial data, personal information and electronic health records are produced by organizations. Consequently, the production of large amount of digital data has increased correspondingly and often the data storage capacity of many organizations has increased tremendously. The management of such a large amount of data in local storage system is very difficult and causes high expenses because of high-capacity storage systems needed and the requirement of experts to manage them. Although the expenditure of storage
hardware has tremendously decreased in recent years, about 75% of the total authoritative cost is still concerned with management data storage (Magalhaes et al. 2015).

The emerging cloud computing paradigm renders a convenient, pay-as-you-go, on demand network access to a shared pool of configurable computing resources. It requires minimum service provider interaction or management effort. Organizations now have an alternative to outsource their data to cloud storage to reduce the burden on local data storage and also to decrease maintenance.

1.5.2 Secured File Transfer

Although Clouds offer limited benefits to data owners, outsourcing data to a remote server and providing management of data to an untrusted cloud service provider, can pave way to losing of physical control over the data (Wang et al. 2009). The cloud is inherently neither secure nor reliable to the clients, and this provides new challenges to the integrity, confidentiality and data availability in cloud computing. For example, deleting less frequently received data to provide free disk space or concealing damaged or compromising data to safe guard the reputation of the organization.

Some organizations have reported corruption of data in servers of major cloud infrastructure providers, and there had been many instances of cloud service outages such as Gmail email mass deletion, Side kick Cloud Disaster, Amazon S3breakdown, and Amazon EC2 service’s outage (Armbrust et al. 2010). The Privacy Rights Clearinghouse (PRC) had reported that more than 535 data breaches happened in 2011- breach of email service providers based on cloud in Epsilon, compromise of Sony Play Station Network, Sony Online Entertainment and Sony Pictures, theft of 3.3million
patients’ medical data of Sutter Physicians Services, theft of customers’ information on EMC’s RSA.

When adversaries get control over the cloud server, they have the capacity to launch replay attack or forge attack which aims at breaking the linear independence among data which is encoded in the corrupted cloud server by replacing the stored ones with old encoded data. Therefore, the integrity of users’ data which is stored on the remote cloud server is highly vulnerable to internal and external attacks. Without a local copy of the data, traditional integrity verification techniques such as signatures and hash functions are inapplicable in the cloud storage. Also, it is impossible to download a large size file from the cloud storage. The situation gets to a worse condition when users access data using their mobile devices.

In this context, a more useful technique is required to verify the integrity of the outsourced data remotely in the cloud. Therefore to address the issue of data integrity in cloud computing the Remote Data Auditing (RDA) techniques have been developed by researchers, which can securely, frequently, and efficiently validate the proof of data possession by creating a random challenge.

RDA techniques are classified into three main categories:

A. **Integrity-based**: The auditor in this group is just permitted to check on the integrity of the outsourced data either directly or through a third party;

B. **Recovery-based**: Apart from scrutinizing the data integrity, techniques in this category support Forward Error Correcting (FEC) codes by leveraging on the Reed-Solomon erasure-correcting code
C. Deduplication-based: It ensures data integrity and efficiency by removing data repetition and increasing data storage optimization.

To design and implement an efficient remote data auditing method, the following relevant properties must be taken into consideration:

1. **Efficiency**: Audit data with liberal storage cost, computational cost, and communication cost between client and server;

2. **Public/private verifiability**: The data owner is able to check the integrity of the outsourced data only in private verifiability mode, while in the public verification mode, the complex task of verification is delegated to a third party to minimize the computational cost of data auditing for the data owner;

3. **Frequency**: Frequent Repetition of the verification process with different challenge information.

4. **Probability of detection**: It represents the chance of detecting potential data corruption.

5. **Dynamic update**: It denotes ability to perform operations like insert, delete, modify, and append on the outsourced data without the need to download all the data.

The RDA methods which are existing now require frequent auditing and involve many processes and frequent data transmission. Consequently, the RDA methods employ additional computational and communication costs on the auditor, which is a big problem for many data owners, especially when they employ mobile devices that have limited computing resources (Fernando et al. 2013). On the other hand, the main design aspect of the RDA methods is to promote dynamic data update operations for different applications because of the dynamic nature of the data.
Therefore different types of data structure like binary tree are used in such methods to achieve this aim. However, the applied data structures in the RDA methods are not able to support dynamic data update operation effectively for large scale data effectively, most importantly frequent data update (Knauthe et al. 2008). This is because a large number of data blocks must be rebalanced by the auditor within the data structure many times, and this results in high computational cost on the auditor. Therefore, it is imperative to design a new data structure to promote dynamic update for large scale data.

1.5.3 Storage Efficiency

One of the famous services offered in cloud computing is the cloud data storage, where subscribers need not store their own data in their servers, but their data will be stored on the cloud service provider’s servers. Subscribers have to pay that provides for this storage service in cloud computing. It not only provides flexibility and scalability data storage, but also provides customers with the advantage of paying only for the amount of data they have to store for a definite period of time, without any worries of efficient storage mechanisms and maintainability issues with highly quantitative data storage. Other than these benefits, customers can easily access their data from any region where the Cloud Service Provider’s network or Internet can be accessed easily (Ateniese et al. 2008).

The end of this decade is characterized by a paradigm shift of the industrial information technology towards a subscription based service business model known as cloud computing. It provides users with a long list of merits, such as

- Provision for computing capabilities
• Broad and heterogeneous network access
• Resource pooling
• High elasticity with measured services

Enormous amounts of data are being retrieved from geographically distributed data sources, and non-localized data-handling requirements and it creates such a change in technological and business model. As cloud service providers (CSP) are individual market entities, privacy and data integrity are the most crucial issues that need to be addressed in cloud computing. Even though the cloud service providers have powerful infrastructure and standard regulations to ensure the data privacy and integrity thereby provide a much better availability, the reports of breaching privacy and service outage have been apparent in last few years.

1.5.4 Searching Files from Cloud Storage

The data can be distributed among two storage clouds such that, an adversary will not be able to get back the contents of the data without the accessibility to the storage clouds (Syed & Baig 2013). Dependency upon a couple of service providers for the storage and retrieval of data might not be secure against conspiring service providers. The cloud user cannot find out that whether his information has been entirely retrieved from the service providers without his knowledge. The threat of conspiring service providers are demonstrated by Boneh (1998).

Let us assume that there are two cloud service providers for customer (C1), who wants to store his data securely. He will divide his data into two parts (D1 and D2) and these parts are distributed on the two available CSPs (CSP1 and CSP2) respectively. The two cloud service providers might conspire with each other, and the parts of data that the
customer has stored on their server is exchanged and the whole data is reconstructed without being detected by the user. The cloud computing users are provided with a decision model, a better reliability and availability is provided by distributing the data over multiple cloud service providers so that, none of the Service Provider can successfully retrieve and use it.

1.5.5 Threats in Cloud Computing

Stored data of customers at cloud service providers is vulnerable to a lot of threats. A lot of studies have proved that a cloud service provider can be a victim to denial of service attacks or its variants, here two kinds of threat model is being considered. First is the single point of failure, where the data availability is affected. If a server at the cloud service provider gets failed or crashed, it will be make it hard for the customer to get back his stored data from the server. Availability of data is also a crucial issue which could be affected, if the cloud service provider (CSP) runs out of business (Blum et al. 1994). Such worries are no longer hypothetical. So a cloud service customer cannot entirely depend upon an individual cloud service provider to ensure the storage of his precious data.

1.6 PROBLEM DEFINITION

To ensure the secure data transmission and storage at minimal cost and searching time. The main aim of cloud computing is to improve the computational capacity of the cloud system and to enhance the access levels to the services and resources of the cloud cheaply. The main challenges are

- Consistency
- Limited scalability
- Data replication
Cloud Computing defines a remote server that access through the internet which aids in business applications and functionality along with the usage of computer software. It saves money that is spent by users on annual or monthly subscription. Due to the advantage of cloud services, more amount of personal information are being centralized into the cloud servers, such as private videos and photos, personal health records, email, government documents, company finance data etc.

1.7 OVERVIEW OF RESEARCH

To protect data privacy, data which is confidential has to be encrypted before outsourcing, in order to provide thorough data confidentiality assurance in the cloud. Data encryption makes data utilization effective and a very challenging task as there could be a large amount of outsourced data files. Besides, in Cloud Computing, outsourced data may be shared by its data owners with plenty of users, who might want to only get back certain specific data files they are interested in during a certain session (Singhal 2001).

One of the easiest ways to do so is through keyword-based search. It allows users to selectively get back their files of interest and has been widely applied in plain text search situations. Unfortunately, data encryption, which hampers the users to perform keyword search and further demands the protection of keyword privacy, makes the common plain text search methods fail for encrypted cloud data. Whereas Ranked search improves system
usability by normally matching files in a ranked order pertaining to certain relevance criteria.

Di Battista et al. (2007) employed homomorphic encryption, multi cloud computing and mobile computing. They utilized multiple cloud schemes for capturing the data to avoid data lock in and used homomorphic encryption to run computations without downloading the data to and fro between cloud computing and mobile computing to avoid the communication expenses.

Cloud computing is an incipient trend in the field of technology where information, software and resource is shared. In cloud computing the tasks are moved from individual systems into the cloud where multiple systems can interact with each other at the same time (Hao et al. 2011). It provides a pay-per-use facility which means that pay only for what the customer uses. This would decrease the customer’s expenses on hardware, software and other services. There are various pros of cloud computing. They are,

- cost saving,
- scalability,
- reliability,
- maintenance

But still cloud computing has certain cons as well. As multiple user’s access information on cloud, there is a huge amount of risk in the privacy of the information stored. Cloud provides distribution of data through computers. When data to be processed is sent by the user in the cloud, the control of the data is provided to a remote party that may not raise the security concerns of the user (Heitzmann et al. 2008). Since users have no physical
access to the data, they are unaware about the location of his data and are not sure whether the integrity of their data is maintained or compromised in cloud. It is important to make sure that the information being processed on cloud is secure and no meddling of information is done when previously unknown parties may be present. A framework is proposed to render data integrity using TPA to make sure that the data of various users is unaltered.

According to Hewitt (2008), cloud computing is said to be the next generation computing model as it can be convenient and efficient for data management system. The emerging need of technology in all the fields has led to the evolution of cloud computing for efficient usage of IT resources. Storage is an important service of cloud computing, as it allows data owners to have their data safely. Frequently data owners have started to choose to host their data in the cloud. As cloud provides many merits but as every coin has two sides, it also has certain challenges. Every day latest publication, a fresh news item, blog entry enlightens the cloud computing challenges and issues. The following are some of these concerns in the cloud:

**Availability**: The data must be available all the time for the clients without having any problems that affect the storage and lead to the loss of client data.

**Access**: When there is an unauthorized access to the data, the ability of changing the client data arise.

**Integrity**: The data correctness, security and legality is the most fields that influence on the cloud and have major lay on the service provider.

**Data Location**: The client is unaware of the actual place that the data saved or centered in because it distributed over many places that led to confusion
Network Load: The over load capacity on the cloud may leave out the system out according to the huge quantity of data between the computers and the servers.

Cloud has several security related problems involving assurance of data. A user entrusting a cloud provider may lose his data accessibility temporarily or permanently due to an improbable event like a malware attack.

1.8 CHALLENGES OF A SECURED CLOUD SYSTEM

There are three important challenges for developing a secure and trustworthy cloud system. They are,

Outsourcing - Outsourcing decreases both capital cost and operational cost for cloud customers. The loss of control problem has become one of the main reasons for cloud insecurity. To address outsourcing security issues, the cloud provider shall be trustful by providing trustful computing and data storage; second, outsourced data and computation shall be verifiable to customers in terms of security services. In addition, outsourcing will potentially obtain privacy violations, as the sensitive data is out of the owners’ control.

Multi-tenancy - Multi-tenancy is defined as the cloud platform is shared and utilized by multiple customers. Moreover, in an environment which is virtualized, data belonging to various customers may be placed on the same physical machine with the help of certain resource allocation policy. Adversaries who may also be true cloud customers may exploit the co-residence problem (Parashar et al. 2013). A series of security threats like data breach, computation breach, flooding attack, etc., are found (Parashar et al. 2013). Although Multi-tenancy is a definite option for cloud vendors due to its economic efficiency, it provides new threats to the cloud platform. Without
altering the multi-tenancy paradigm, it is definite to design new security mechanisms to deal with the high risks.

**Massive data and intense computation** - Cloud computing is capable of dealing with mass data storage and high skilled computing tasks. So, traditional security mechanisms may not be enough due to unbearable computation or communication overhead. For example, to check the integrity of data that is rarely stored, it is impractical to cut the entire data set. To this end, novel strategies and protocols are expected.

Currently, cloud computing can be considered as a real solution in every aspect of computing world as far as the present World Wide Web architecture and standards are able to facilitate them. The elusive reason for this is many facilities it brings to both users and the system owners. They are feasibility, ease of use, cost etc. There are also many demerits and troubles it might bring depending on the users and providers; therefore there are many organizations and enterprises that prefer deploying distributed solutions into their own infrastructure. There exist many properties and variables in a cloud system that define the whole system and affect the above mentioned benefits and nuisances. One of the important one is architecture (Bhuvaneshvaran et al. 2012).

This cloud computing model renders different kinds of security services in various forms. It covers events of fundamental services of enterprise, or other similar systems such as Authentication, Accounting, Authorization, identity Management and so on. Integrity and confidentiality are the main concerns of a security system for files. Confidentiality ensures that data is accessible only by the intended recipient(s) and integrity ensures that data is still intact as the recipient should get back. These two fundamental and other properties can be addressed using cryptography. One of the methods of cryptography is PKCS#7, aka Cryptographic Message Syntax, standard
which is a packaging system on certificate based key management. This standard is used for data security. “Protocols for Secure Library Server” and “Network Based Secure File Sharing System with the help of smart cards” (Weinhardt et al. 2009) are the prominent client and server based tools of a client-server distributed application. These components develop the “Distributed Secure File Sharing System” as it is called commercially. It is a content security and file sharing system that is wholly depends on certificates, SAML Policies and several PKCS standards. However, these solutions boundaries do not fit into the Cloud yet, so that it shall be extended and integrated.

1.9 ORGANIZATION OF THESIS

The thesis is organized as follows

Chapter 1 describes the introduction to cloud computing, differences between distributed and grid computing, characteristics of cloud, cloud deployment model, cloud service model, issues in cloud computing, problem definition and the organization of the thesis.

Chapter 2 narrates the prime objective of the research work, intensive review of literatures pertaining to the present work and the objectives of proposed work.

Chapter 3 describes the Availability and Integrity Verification Protocol (AIVP) to evolve the performance and integrity of the cloud storage.

Chapter 4 presents the Searchable Encrypted Data File Sharing (SEDFS) scheme for efficient search for share the encrypted data files in the cloud storage.
Chapter 5 deals with the results and discussion and the performance comparison of the proposed scheme with the existing schemes.

Chapter 6 concludes the thesis by emphasizing the major conjecture of the study. A summary of the research contribution and the scope for future studies are also discussed in this chapter.