# Improved Data Storage and Sharing in Cloud Computing Using User Partioning Method

 <sup>1</sup>Ruksana Bano, <sup>2</sup>Nagendra Kumar
<sup>1</sup>Research Scholar Dept of CSE
<sup>2</sup> Assistant Professor Dept of CSE
<sup>1, 2</sup> Shri Ram Institute of Science & Technology, Jabalpur, Madhya Pradesh, India.

Abstract- Many software industries use Cloud for data storage and computation at Cloud service provider (CSP). Because it's on demand self service, rapid resource elasticity, and relief of burden for storage management, universal data access with independent geographical locations, cheap, doesn't require installation/maintenance. Even mobile phone users also store the data in Cloud. Cloud platform can be deployed as private cloud, public, community and hybrid cloud. Insecure Interfaces and APIs, malicious insiders, sharing of infrastructure, account hijacking can lead to threats to cloud.

How to maximally use and efficiently manage cloud storage resources has become one of major problem in cloud computing area. If cloud storage capacity is limited or restrictions on a single file size, user may hire more cloud computing storage services. Data partitioning and distributed storage technologies are focused on to solve the problem that size of large dataset is limited in cloud storage. A model of user partition based data storage in cloud is put forward and its availability is discussed. Fragment partition description is given and three algorithms are designed on the basis of model. The first algorithm is about user partition and storage of data, the second algorithm is about data confidentiality and hashing.

*Keywords*- Key management, onsite cloud, outsourced cloud, public cloud, symmetric key, applications, Security, Integrity, Hashing.

## I. INTRODUCTION

Cloud computing has revolutionized the way computational resources are made available [1]. It has brought up new delivery models that render new possibilities of renting infrastructure, computing platforms, and software as services, and reduces the need for large-upfront investments. Particularly, this entails benefits such as payment on a perusage basis, low or no fixed costs, and short time-to-market [2]. Especially public cloud solutions often induce significant economies of scale resulting in cost advantages that cannot be achieved by in-house or mid-sized data centers [3]. On these grounds, cloud computing constitutes an attractive model for small and medium sized companies, and its resource elasticity enables to respond quickly to changing business conditions. But also larger organizations are increasingly integrating cloud services into their daily IT operations while many are already using them for almost all storage needs of company assets such as customer data, operational data, and digital archiving [4]. Cloud computing has become a key enabler for big data since it provides nearly unlimited resources, such as computational power and storage capacity, on demand [5]. Due to cost advantages, massive storage capacities and service levels, Cloud Data Stores (CDS) can also serve as inter-organizational storage centers and exchange hubs [6].

However, with more and increasingly confidential data being stored in the cloud and the shift of responsibility toward cloud services, new security and privacy threats arise. How far can a third party cloud service provider (CSP) be fully trusted in securing data from unauthorized access, secondary use, disclosure, manipulation, or even loss? Can data availability be guaranteed when relying on a single CSP? Recent reports on intelligence operations and internet surveillance as conducted by, e.g., U.S. secret agencies as well as subjection to foreign law and judicial authorities have reflected concerns about the trustworthiness and privacy of public cloud solutions. Moreover, legislative frameworks, such as the EU Directive 95/46/EC [7] or HIPAA, require strong protection of sensitive data and impose requirements on their transmission and processing. This emphasizes the need for practical solutions that increase the security and privacy of data in the cloud. In today's society, information technology has changed people's traditional way of life; people's dependence on information technology continues to increase in recent years. The users are anxious with their data storage location and by whom their data can be accessed. The user cannot easily trust the cloud service provider. Therefore trust management is the important factor which arises in this field. As data being outsourced, one client data is stores on the same resource where other client's data has been stored. So if one client's data is compromised so there are chances of the other client's data to be compromised. Encryption is one of the security solutions. Although clients still feels insecure as they don't know which other client also has the same measures in place.

Cloud computing is making everything simpler and flexible nowadays, but there is another important aspect which is "What about the security of data over the cloud?" Cloud architecture with robust security implementation is the key to cloud security. Cloud is complex and hence security measures are not simple too. Since it is new, it faces new security issues and challenges as well. Till date, most users don't trust storing their data on SASS based cloud computing providers such as Dropbox, Skydrive and Google Drive etc. Since the outburst of Cloud Computing in the year 2009, various methods are devised to increase the security of the data being stored over Cloud Servers. Some of which include Encryption, Decryption, Data Partitioning, Digital Signatures etc. In the presented work we implement a strong Security mechanism by making use of multilevel storage strategy and Key generated by using strong hash function.

## 1.2 Cloud Data Storage Challenges & Issues

The cloud computing does not provide control over the stored data in cloud data centers. The cloud service providers have full of control over the data, they can perform any malicious tasks such as copy, destroying, modifying, etc. The cloud computing ensures certain level of control over the virtual machines. Due to this lack of control over the data leads in greater security issues than the generic cloud computing model. The only encryption doesn't give full control over the stored data but it gives somewhat better than plain data. The characteristics of cloud computing are virtualization and multi tenancy also has various possibilities of attacks than in the generic cloud model.

1.2.1 Cloud Storage issues

#### A) Data privacy and Integrity

Even though cloud computing provide less cost and less resource management, it has some security threats. As we discussed earlier cloud computing has to ensure integrity, confidentiality, privacy and availability of data in generic cloud computing model but the cloud computing model is more vulnerable to security threats in terms of above conditions. Because of simplicity cloud users are increasing exponentially and applications are hosted in cloud is very high. These situations lead to greater security threats to cloud clients. If any attack is successful on data entity will leads to data breach and takes an unauthorized access to data of all cloud users. Because of this integrity violation cloud data lost multi-tenant nature. Especially SaaS providers may also lost their technical data and they have great risk over data storage. Apart from these risks, data processing also has great risk while data being transformed among multiple tenants. Because of virtualization multiple

Page | 1193

physical resources are shared among the users. This leads to launch attacks by malicious insiders of the CSP and/or organization. These situations may allow the malicious user to perform attacks on stored data of other customer while processing their data. Other major risk is when data is outsourced to third party storage by the CSP [5]. The key generation and key management in cryptography for cloud computing is not standardized up to the mark. But without standard and secure key management for the cloud doesn't allow the standard cryptography algorithms to perform well in generic cloud computing model. Such that cryptography may also ensures the potential risks to cloud computing obtain the data of previous users. The authors in [8] were able to recover Amazon machine images files 98 % of the times.

## **II. LITERATURE SURVEY**

Although cloud computing itself has many open problems, researchers in the field have already made the leap to envision Cloud computing Data Security. Their goal is to achieve better overall Confidentiality, reliability, and availability by utilising multiple cloud security parameters. Cloud Data security research is still in its infancy and the body of knowledge in the area has not been well defined yet. In this chapter, we propose and motivate taxonomies for Cloud Data Storage architectures and Advance Encryption mechanisms. We present a detailed survey of the most prominent developments from both academia and industry and we fit each project onto the proposed taxonomies. We discuss how the current Cloud Data Security environments facilitate the uses of Cloud applications across worldwide. Finally, we analyse the existing works and identify open challenges and trends in the area of Cloud Data Security. To access cloud server user has to create a profile in the cloud service provider by providing the basic information like username, Name, password, address, mobile number and e-mail id. After user registration in service provider users can create, edit, modify, delete, upload, download and save the file in the cloud server. If the hackers hack the password of the user they can access user files. So our presentation in this paper is using high security password to ensure data storage in cloud computing for distributed systems. After user login if they want to do any operation like create, edit, modify, delete, upload, download and save the file they have to enter the high security password that was received in the user mobile. The session time is also fixing to the high security password within that time user has to enter that high security password. If the user did not receive the high security password through sms, users need to press refresh button to get new high security password. This chapter gives details of different Cloud Data Security techniques implemented in cloud computing. Along with it, we have the previous and yet emerging Security implementation

algorithm in support that provides guidelines in investigation and pursuit of our advanced proposed scheme. Eventually, we come across the latest trends and surveys performed on advanced Cloud Data security techniques in cloud computing by leading organization. We hereby, describe the study performed prior to developing the proposed system.

2.2 Multi-Level Cloud Data Storage

## A. End Users

There is a push by governments to collaborate on the cloud with projects such as Helix Nebula [9] to encourage collaboration via sustainable elastic resources. It is the responsibility of the end user of cloud resources must consider if the service is supplied at a reasonable cost. If the service is underutilized it may become a liability. It the service is underutilized it may be assigned a low priority for monitoring and maintenance purposes. This can lead to difficulties as software is not maintained. Security becomes weakened through lack of patching, etc. Thus the IT staff of the end-user company must pay particular attention to the service level agreement to which the users have subscribed. The end users utilizing resources such as data storage on the cloud need to be educated as to the risks of storing the data on the cloud. Issues such as data location, monitoring & data provenance are essential particularly in view of current European laws. Where EU collaboration in a cloud is required [9] it is necessary for each partner to assess that each partner evaluate the system for their functionality, legal and security requirements. Seemly mundane issues such as the locality of the data stored on the server are becoming increasingly relevant to the end-user due to updated European laws.

## B. Data Location

An important issue is whether or not their data is kept separate from others. It is not appealing to an end to an end user with the idea of their important and confidential business files being stored on a multi-tenant server, which is only divided through the use of hypervisor software and virtualization [10]. Reasonable security measures would have to be implemented just to make sure that the files stored by a business on a multitenant server do not somehow come into the hands of another business on the server. Unless some form of guarantee by the provider is established for the safety of their files, the end user may not be able to accept this arrangement. Service Level Agreements (SLA's) may be applied to such arrangements. However, the monitoring of the is often difficult to govern.

C. Maintaining the Integrity of the Specifications

Another issue that the end user must address is the exit strategies available should anything untoward occur. An event such as a provider going out of business could result in the requirement that an end user must relocate their data and applications elsewhere with another provider. This may involve the creation of some software for the sole purpose of retrieving their data and applications, followed by moving it to an internal location within the business or an external location, such as an alternate provider. This movement between providers can be problematic as it may involve the recoding of software to work with the new provider which can take some time to fully implement.

## 2.3 Security Challenges in the CIA Triad

Confidentiality, Integrity and Availability (CIA) losses can make a big impact in the business of the cloud computing because the data is the core component for any business. Data integrity is the assurance given to the digital information is uncorrupted and only be accessed by those authorized users. Thus, integrity involves maintaining the accuracy, consistency and trustworthiness of data over its entire life cycle [11]. Maintaining CIA is easier in enterprise computing but in cloud computing it is more complicated because of the multi-tenant architecture and the distributed nature of the infrastructure. The following steps can be used to maintain a proper CIA in cloud computing:

> Once the data are created, classify the data, identify the sensitive data, define policies, and create access methods for different types of data. Also, create policies for data archive and data destroy.

 $\succ$  Store data with proper physical and logical security protection, including the backup and recovery plan.

> Identify which type of data can be shared, whom and how it can be shared and define data sharing policies. In cloud computing, many such policies are collectively called as Service Level Agreements (SLA).

➤ Create a corrective action plan in case data is corrupted or hacked due to network or communication devices, security flaws while data is in transit.

According to Aldossary and Allen, integrity should be checked not only at data, but also at the competition level [12]. Computation integrity refers to only the authorized applications are allowed to access the data and use it for computation. Any abnormality from normal computing should be avoided. An effective Identity and Access Management (IAM) can avoid loss of confidentiality and integrity. Loss of availability can happen through loss of data and data inaccessibility. Cloud computing employs few techniques like scalability and high availability at the architecture level. There are different methods and procedures are followed to improve data security related to the CIA triad at different stages of the data lifecycle. Some of the important methods are listed below:

Apply data encryption when the data is at rest and also when the data is in transit. Apply strong encryption algorithms like Advanced Encryption Standard (AES) and Rivest Shamir Adelman (RSA) algorithms. Different types of encryption methods are described in [12]. Amazon S3 uses one of the strongest encryption algorithms, 256-bit AES.

◆Encryption methods are generally to provide confidentiality against attacks from a cloud provider, but it cannot protect data against configuration errors and software bugs [4]. Hash methods can be used to find out accidental and intentional data changes. But they consume more bandwidth and timeconsuming.

♦ Third Party Auditing (TPA) can be employed to check for the data integrity. Many researchers [13] insist to audit data integrity by third-party auditors because they are specialized in that.

## 2.4 Existing Data Storage Mechanisms in Cloud Computing

The SecCloud is presented by Wei et al. [14], it provides a storage security protocol for cloud customer's data and it not only secures the stored data but also provides security on computational data. The SecCloud protocol uses encryption for storing data in secure mode. The multiplicative groups and cyclic additive pairing is used for key generation for cloud customers, CSP, and other business partners or trusted third party. The encrypted data along with the verifiable signature is sent to cloud data center along with session key. The Diffie-Hellman algorithm is used for generation of session key for both bilinear groups. By receiving encrypted data the cloud decrypts the data, verifies the digital signature and stores the original data in specified location in cloud. The SecCloud verifies whether data is stored at specified location or not. The Merkle hash tree is used for computation security in SecCloud protocol. The verifying agency will verify the computational results that are building by using Merkle hash tree. The File Assured Deletion (FADE) protocol provides a key management with data integrity and privacy in [15]. The key management along with the data integrity and privacy are assured by File Assured Deletion protocol (FADE) proposed in [16].Because of FADE simplicity; it is a light weight protocol and uses both asymmetric and symmetric key encryption of data. The Shamir scheme protects symmetric and asymmetric keys to generous the trust in the key management. A group of key managers are used by FADE protocol, those acts as a trusted third party. The key k is used as encryption key for file F of the client and another key used for encryption of data key (k.). The policy file maintains the details that which files are accessible. So that, to upload data the user requests the key pair from the third party by sending policy file p. The key manager sends public and private keys to the user by using the policy file. The upload file encrypts with randomly generated k and k is encrypted with symmetric key. That encrypted file is decrypted with the public key of generated key pair and MAC is also generated for integrity check. The reverse process will be taken by the receiver to get back original data.

Liu et al.[15] proposed a scheme that has a time based reencryption with ABE algorithm to support secure data sharing among the group with access control. This scheme ensures that forwarded data safely reached to the group users and it maintains the user revocation. In this scheme, the time period is associated with every user and by expiration the revocation automatically by Cloud Service Provider (CSP). This time based encryption scheme allows users to share keys in prior with CSP and CSP generate re-encryption keys by taking request from user. The ABE protocol ensures an access control by examining the set of attributes rather than identity. This scheme ensures the privacy and availability of data among the group peoples but doesn't concentrate on data integrity. The probabilistic sampling is used to reduce the computational redundancy instead of rebuilding the whole tree again. The below list are key recommendations by the Computer Security Alliances (CSA) [16] for the data security and effective key management. The scope of key should be maintained by group or individual. The standard encryption algorithms should be used and weak algorithms should discard.

## **III. PROPOSED SYSTEM**

## 3.1 Proposed Method

When users put their data on the cloud, the data integrity protection is a challenge. User should be relieved from the burden of data storage & maintenance. Data in the cloud is typically in a shared environment alongside data from other customer. It will be important for data center to be secure against unauthorized access. Security is fundamentally about three goals:

- Confidentiality (C)
- Integrity (I)
  - Availability (A)

Confidentiality refers to keeping data private. Confidentiality is supported by technical tools such as encryption & access controls as well as legal protections. Integrity is a degree confidence that the data in the cloud is what is supposed to be there and is protected against accidentional or intentional alteration without authorization. Integrity is controlled by robust access control mechanism. Availability means being able to use the system as anticipated. Cloud technologies can increase availability through wide spread internet enabled access, but the client is dependent on the timely and robust provision of resources. Data among users within cloud is shared by defining security level. Encryption is used to assure that it there was a breach of communication integrity between two parties that the data remain s confidential. Authentication is used to assure that the parties communicating data are authorized. The proposed system has three different entities data owner, data retriever and cloud server.



Figure 3: Entity involved in the system.

Data owner uploads a file on cloud server and share files to other users according to security level decided by him, than only user having that level & permitted by owner can have secure access of files. Proposed system provides three level of security for user's data on the basis of Confidentiality (C), Integrity (I) & Availability (A).These three levels are basic, intermediate and high. Basic is the least secure level and high is the most secure level. Intermediate uses mid level security. We called these levels as zone.

In proposed system user may have choice to categorized there data or files on the basis of their importance & security with respect to user. In an organization there may be different type of user working on same cloud environment, which shares same data or files, between each other. In the existing environment only one security policy is applied to all data or files, so it increases complexity. To reduce this proposed system introduces multilevel security for sharing data or files between different users. That gives choice for data owner to maintain secure data sharing for different user. In proposed system security level is allotted for every user at the time of registration according to the priority. Priority is decided by the value of Confidentiality (C), Integrity (I) & Availability (A). Every user can share contents for other user that exists on same level of security. So actually user priority assignment plays role of user classification at different level of security i.e. basic, intermediate or high as depicted in figure 3.2.



Figure 3.2: Users with different security levels in the proposed system.

Every user should be registered first on the cloud. Cloud server handles user database. It also stores information about user's owned data contents. Cloud server creates directory for data owner who uploads data on cloud server. User directory contains data uploaded by him. User that play role of data retriever should login into the system first. If it is authorized than it selects the user and level of security, than cloud server display list of files that are shared by data owner to him, and sends data retrieval request to cloud server. If user is authorized and requested data with correct security level than server encrypt the file using ECC and generate key for it. Data retriever gets encrypted data. It should generate key for finding actual data. Cloud server generates key for that user and send it which should be entered in the system by data retriever to get actual data.

3.2 Zone allocation algorithm

Input: value of C, I, A.

Output: categorized data for corresponding ring. Zone is returned.

Algorithm:

Input value of C = Value of Confidentiality.

I = Value of Integrity.

A =Value of Availability.

Calculate S = (C + (1/A)\*10)/2 4.

IF S = = 1||2||3 then zone =third /\* zone 3 allotted (public data). IF S = 4||5||6 then zone =second /\* zone 2 allotted (protected data).

IF S = = 8||9||10 then zone = first /\* zone 1 allotted (private data).

#### IV Performance Evaluation

The data retrieved is in encrypted form and again the user has to go through some security checks to access the data from cloud server. The figure below shows data in encrypted form.



Figure 4.7: Data Retrieving Module(c).

Information and data security are based on factors such as authenticity, accuracy, availability, data credibility, confidentiality and no repudiation. The proposed approach has the ability to contribute to the necessary data and information security. It uses data classification techniques as to reduce the complexity and increase the security of data. In the proposed method the data owner has direct control over its data. The proposed method provides more security of owner data. Confidentiality is assured by encrypting the data.

The data availability is provided by overcoming many existing problems like Denial of Services, data leakage, user managed encryption keys etc. It also provides more flexibility and capability to overcome the problems faced by today's complex and diverse networks.

Following security parameters has been fulfilled:

1) Denial of Services: In a denial-of-service (DoS) attack, an attacker attempts to prevent legitimate users from accessing information or services. By targeting your computer and its network connection, or the computers and network of the sites you are trying to use, an attacker may be able to prevent you from accessing email, websites, online accounts (banking, etc.), or other services that rely on the affected computer. By distribution of cloud data in different zones, an attacker is not able to get into all the services offered by different zones at a time. Hence it prevents the DoS attack to some levels.

2) Data Leakage: Sensitive customer information and confidential corporate data can slip out of an organisation via email, lost laptops, USB drives and a host of other ways, But there are drawbacks to all of the options and no one technology fully addresses the problem. Data leakage gateways can help enforce policies on the network but can't stop an employee from copying confidential data onto a USB storage device or from taking a laptop home and sending confidential data via Web mail. Endpoint device protection technologies that track operating system and application operations to enforce policies at the desktop can block someone from copying data to a USB drive, but it won't be on all devices in an organisation and it can become too costly to block people from doing what they want to do. Classifying sensitive documents on the network to different zones can overcome the problems.

5.3 Evaluation on the basis of Key Decryption Parameters: The hash generated by using MD5 is evaluated by attempt to decrypt it by using different simulation parameters. First we try to decrypt the key by using Data- Dictionary attack in our simulation. The results show that Data Dictionary is unable to decrypt the key. Figure given below shows the detailed result of my simulation.



Figure 5.1: Hash Generated By using MD5.

## V. CONLUSION AND FUTURE WORK

Generally, data classifications can yield significant benefits, such as compliance efficiencies, improved ways to manage the security of user's resources, and facilitation of data management in cloud. Although data classification efforts can be a difficult undertaking and require user assessment for successful implementation, quicker and simpler efforts can also bring benefits. Any data classification efforts should endeavour to understand the needs of each user and user can be more aware on data storing, processing capabilities, and data transmission in the cloud. The suggested data classification of protection levels: protected, private and public are worth noting as a recommended data classifications guide. It is also expected to help reduce and mitigate risk with the suggested technical security solutions.

This thesis gives observations aspects of data classification functionality and Generation mechanism. It analyses the various security threats and their conceivable impact on the system.

The future work of this research will investigate the data protection levels that can be applied by CSPs in cloud storage architecture. This involves investigating security control and measures implemented by known CSPs. The adopted security model and its data classifications (if any) will also be explored and assessed as the current technology and guideline. This will later involve a development of the security framework that can be used in a cloud storage architecture addressing the identified security classification for protecting data in cloud storage. The next phase will involve validating the idea with experts and simulations of the designated security classifications on cloud storage that will be used as the proof of concept.

## REFERENCES

- D. Agrawal, S. Das, A. El Abbadi, Big Data and Cloud Computing: current state and future opportunities, in: 14th International Conference on Extending Database Technology (EDBT/ICDT'11), ACM, 2011, pp. 530–533.
- [2] Amazon Simple Storage Service. http://aws.amazon.com/s3/.
- [3] A. Rosenthal, P. Mork, M.H. Li, J. Stanford, D. Koester, P. Reynolds, Cloud computing: a new business paradigm for biomedical information sharing, J. Biomed. Inform. 43 (2) (2010) 342–353.
- [4] Gartner Inc., Gartner Says Public Cloud Services Are Simultaneously Cannibalizing and Stimulating Demand for External IT Services Spending, 2012, http://www.gartner.com/newsroom/id/2220715.
- [5] R.F. Chong, Changing the World: Big Data and the Cloud, 2012, http://www. theatlantic.com/sponsored/ibm-cloudrescue/archive/2012/09/ changing-the-world-big-data andthe-cloud/262065/.
- [6] S. Shini, T. Thomas, K. Chithraranjan, Cloud based medical image exchange security challenges, Procedia Eng. 38 (2012) 3454–3461.
- [7] The Library of Congress, Bill Text 107th Congress (2001–2002) H.R. 3162, Public Law 107-56 Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT ACT) Act of 20012001.
- [8] H. Chen, R.H. Chiang, V.C. Storey, Business intelligence and analytics: From big data to big impact, MIS Q. 36 (4) (2012) 1165–1188.
- [9] Y. Raekow, C. Simmendinger, P. Grabowski and D. Jez, "License Management in Grid and Cloud Computing", International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, IEEE, 4-6 Nov. 2010.

[10] B. Kepes, "Understanding the Cloud Computing Stack SaaS, PaaS and IaaS", CLOUD U: Understanding the Cloud Computing Stack SaaS, PaaS and IaaS, [Online]. 3, 17.

Availableat:broadcast.rackspace.com/hosting\_knowledge/ whitepapers/Understanding-the-Stack.pdf. Accessed 22 February 2012.

- [11] Y. Wang, Q. Wu, B. Qin, S. Tang, W. Susilo, and S.Member, "Online / Offline Provable Data Possession,"IEEE Trans. Inf. Forensics Secur., vol. 12, no. 5, pp.1182–1194, 2017.
- [12] I. Baciu, "Advantages and disadvantages of cloud computing services, from the employee's point of view," no. 13, pp. 95–101, 2015.
- [13] M. Sookhak, H. Talebian, E. Ahmed, A. Gani, and M. K.Khan, "A review on remote data auditing in single cloud server: Taxonomy and open issues," J. Netw. Comput Appl., vol.43, pp. 121–141, 2014.
- [14] L. Wei, H. Zhu, Z. Cao, X. Dong, W. Jia, Y. Chen, A.V. Vasilakos, Security and privacy for storage and computation in cloud computing, Inform. Sci. 258 (2014) 371–386.
- [15] Y. Tang, P.P. Lee, J.C.S. Lui, R. Perlman, Secure overlay cloud storage with access control and assured deletion, IEEE Trans. Dependable Secure Comput.9 (6) (2012) 903– 916.
- [16] Cloud security alliance, security guidelines for critical areas of focus in cloud computing v3.0, 2011.