*International Journal of Engineering & Technology,* 7 (4.5) (2018) 329-332

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*

***Research paper***

# A Review of Assured Data Deletion Mechanism in Cloud Computing

**Sultan Ahmad[1]*, Dr. Mohammad Mazhar Afzal[2]**

*[1]Research Scholar, Department of Computer Science and Engineering, Glocal University, Saharanpur, Utter Pradesh, India*
*[2] Associate Professor, Department of Computer Science and Engineering, Glocal University, Saharanpur, Utter Pradesh, India*
**Corresponding author E-mail: Sultan.14nov@gmail.com*

**Abstract**

As Cloud Computing has become a new way to store information remotely. It makes possible for many companies to become paperless. IT people and companies can outsource their huge data by using this emerging technology. The major important characteristics of cloud computing are on-demand self-service, broad network access, resource pooling, rapid elasticity and measured services. These cloud computing characteristics are helpful for organizations to become agile. Also there are few serious problems like security, scalability, availability and interoperability which are associated with the above cloud computing characteristics. Providing a secure and efficient access and retrieval of the valuable data is a main concern of cloud computing technologies. It is very important to protect data and credentials of cloud client on the internet. Apart from these security issues of data, the deletion of data on cloud storage is also a very serious issue nowadays. Once a cloud user wants to delete his data, it must be sure that data actually deleted from all cloud storage sources and there should not be remain any copy of data exist at anywhere in cloud storage anyway. In this paper, different existing data deletion and retrieval methods are studied, observed and analyzed. The importance and related challenges to assured data deletion on the cloud storage are explored.

***Keywords***: *Cloud Computing; Data Security; Data Deletion; Data sanitization; Assured Deletion; Data Retrival*.

## 1. Introduction

Cloud computing has changed the IT world. Since last decade, IT industries and IT people are shifting their data to the huge cloud storage as their data is getting bigger and bigger. These big data are needed to be accessible from any devices from anywhere. That's why moving the data to the cloud storage becomes an easy norm and practice. A share pool of computing resources, being provided by Cloud providers like servers, storage and application can be accessed on your fingertips with minimal effort and pay as you use basis. Cloud adoption is being done in all sectors such as education, government agencies, financial and research institutes. Some very trending and emerging technologies likes' mobility, Internet of Things, Big Data analysis and social media are also enforcing organization to move towards Cloud Computing.

Cloud computing providers facilitate robust data to customers and business organizations. But in some condition user want to delete permanently his data. That should be unrecoverable in very reliable manner after a particular time. Assured deletion is very important. User's data files should be permanently inaccessible after the delete request by user. It is also important and undesirable to keep data backups permanently. There is a chances of data leakage and breach if data files and backup remain undeleted in cloud storage. Cloud storage provider mismanagement or operator erroneous action may cause expose of sensitive data of cloud user. That's why all cloud users want to keep their unwanted backups assured deletion. To avoid liabilities and future risk, many government organization and enterprises keep their data for a finite numbers of years only. On the other hand, Cloud providers might have multiple copies of the data on the cloud storage in order to do the fault-tolerance and data recovery in case of disaster or failure of some part of system. Due to this reasons, Cloud users may not be aware that how many copies of their data are on cloud storage and what are the exact location of data in some cases. Many Cloud providers do not make clear their replication policies. Are they reliably clean all copies of client data on the requests of deletion by user? It is really doubtful. As users lose direct control over their data, so assured deletion of their data becomes more difficult. However, incomplete deletion is another issue that lead to exposure of sensitive data and costs associated with high financial loss sometime.

It should be also very clear that the assured deletion of data is not only important for cloud users but also for cloud providers. Cloud providers have to follow regulations and compliance of various geographical region and countries. It effects the market of Cloud Providers also.

This paper gives an initial study and review of the assured deletion of data in cloud computing. It has the comparative study of the existing solutions and finding the limitations and challenges associated in this area. This will help for further research in this area. Microsoft's Azure, Google's suite of apps (Gmail, Google Docs, and Google Calendar etc.), Amazon's EC2 (Elastic Compute Cloud), IBM blue cloud and Salesforce.com etc. are some of the popular public cloud service providers. In this study, we analyze

the assured data deletion requirements and the existing approaches and their limitations against these requirements. It concludes the further scope of research. The encryption method which is being used in data deletion needs more improvement, innovations and research, along with other issues. The remainder of the paper proceeds as follows. We provide the necessary background on cloud storage systems, the requirement for assured data deletion, and review of the existing approaches and their limitations. Finally, last section concludes and presents future work to be done.

## 2. Background of Cloud Computing Storage

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. servers, storage, networks, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [1]. It is based on virtualization, distributed computing, utility computing and service-oriented architecture. The Cloud customer/user requests over network and cloud provider computes according to request to response cloud user. The cloud user creates his data and store in cloud storage provided by Cloud provider. Cloud provider provides elastic storage services as a pay-as-you-go manner. NIST has explained the Cloud Computing reference model as in the given Figure 1.

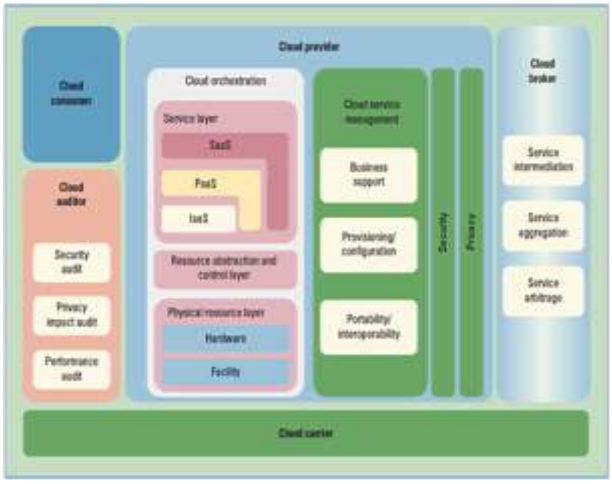


**Fig. 1:** NIST Cloud Computing Reference Model

Data created by individuals, businesses, and applications need to be stored so that it can be retrieved when required for processing or analysis. A storage system is the repository for saving and retrieving electronic data and it is integral part of any cloud infrastructure. This can be shown in Figure 2.

A storage system has storage devices that enable the persistent storage and the retrieval of data. Storage capacity is typically offered to consumers along with compute systems. Apart from providing storage along with compute systems, a provider may also offer storage capacity as a service (Storage as a Service), which enables consumers to store their data on the provider's storage systems in the cloud. This enables the consumers to leverage cloud storage resources for purposes such as data backup and long-term data retention. A cloud storage infrastructure is typically created by logically aggregating and pooling the storage resources from one or more data centers to provide virtual storage resources. Cloud storage provides massive scalability and rapid elasticity of storage resources. The cloud storage infrastructure is typically shared by multiple tenants or consumers which improves the utilization of storage resources.

Disaster recovery, Cloud backup and archiving infrequently accessed data are most commonly used cases. Many customers also use cloud-storage services for DevOps to minimize their capital cost. During the project, they just spin up the compute and storage resources and spin them down at end of the project.

Some of the major cloud-based storage providers dominating the global market are Microsoft Azure, Google and Amazon Web Services. Some traditional storage providers such as EMC, Hitachi Data Systems, Hewlett Packard Enterprise, IBM and NetApp etc. are operating with products for both small business as well as big enterprises. They have their self-service portals to monitor and provision the uses

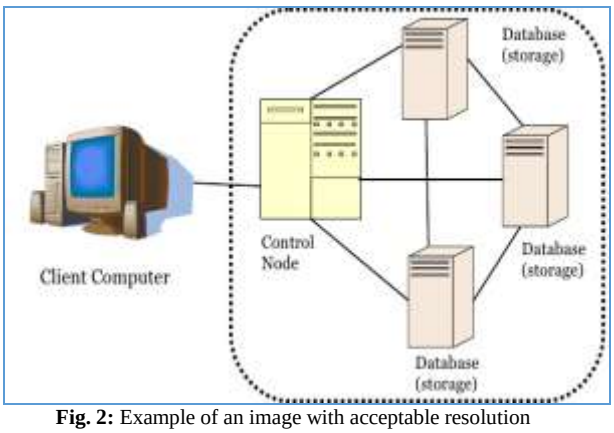


**Fig. 2:** Example of an image with acceptable resolution

## 3. Requirements for Assured Data Deletion

To solve the data security and privacy issue, many significant efforts have been done earlier. But assured deletion of data is also an important aspect of managing and protecting data. It comes a time when cloud user needs data to be deleted. In non-Cloud context, assured deletion of data is very known and understood [2,3]. But in cloud context, it becomes very complex where many features likes service delivery models, multi-tenancy, virtualization, elasticity, high availability and huge data storage are associated. The features challenges and make hard the assured data deletion in Cloud computing.

These are the below requirements for assured Deletion.

**Fine-grained :** Only targeted data should be deleted without disturbing the remaining data. Remaining data should safe and accessible to client user. Fine-grained deletion is comparatively costs less as it provide user control on data.

**Service Availability :** Assured deletion should be done without disturbing working pattern. The daily work and use of data should not be effected. Service should not be disturbed for all cloud users even for the user who requested deletion.

**Cloud Computations :** Work should not disturbed by assured deletion. It should also ensure computation of data without any problem.

**Complete Deletion :** Assured deletion should reflect everywhere data or the copies of data exist. It should affect all copies and also reflect to metadata. Data also should be cleared and cleaned from buffer, RAM or any service layer that uses the data.

**Timeliness :** Deletion should be completed on time and deleted data should not be assessable from immediate action.

**Error Handling :** Assured deletion should be completed without any error or problem in reasonable timeframe.

**Acknowledgement of deletion:** Cloud user should be acknowledgement and assured after deletion. It should be informed to cloud user that data deletion has been completed successfully. should be completed without any error or problem in reasonable timeframe.

For Cloud user prospective, it is always challenging to assure data deletion as cloud users do not have access of geographical location of the cloud storage and infrastructure. Cloud user cannot do any

modification in cloud infrastructure. Generally, the deletion method is based on cryptography. Mainly encryption techniques are used to delete the data from cloud storage.

## 4. Existing Approaches and Limitations

There are many approaches to achieve assured deletion. **Secure overwriting** [4] is way to overwrite on existing data to be deleted by new data. But overwriting a huge data is very time consuming and not in practical. It is very rarely adopted method. The major problem is that it requires internal modification. In Cloud Computing prospective, data is outsourced and storage infrastructure is managed by third party. So this technique does not make sure for replicated data also.

**Disk Scrubbing**[5] is another approach of multiple overwriting. But there is also guarantee that all backup copies of data will be removed simultaneously.

**Cryptography** is a general process that is being used in many proposed methods. The technique behind this cryptography method is to encrypt the data before storing to the cloud storage. In order to delete the data, the encryption key is being destroyed. This makes the data unrecoverable. In this process, with the help of a randomly generated data key, the data file is first encrypted by data owner. The data key is then encrypted by ephemerizer (Third party key manager) with a control key for that file. This is called key escrowing. The control key is time-based. If the data-owner requests delete operation or the designated time for data file expires, then the control key will be automatically destroyed. According to the methods of key escrowing, the existing schemes can be classified into three categories: Trusted Third party(TTP) based methods, Distributed Hash table(DHT) based methods and Key Modulation Function(KMF) based methods.

Cryptography concept is also being used in the method of assured deletion given by Perlman[6] FADE. It uses policy based assured deletion. Each file is having his own access policy and related with control keys. Before sending to cloud, every data file is being encrypted with data key and control key. In order to delete the data file, the encryption key is being destroyed. This makes the data file unrecoverable. This approach follows fine-grained deletion. It only allows to the user to revoke policy associated with data or data files to be deleted. But in case of multiple versions of a data file, it does not work. Using Perlman concept many researchers have given their proposal for assured data deletion.[7, 8, 9, 10, 11] with minors modifications. Many methods like policy-based assured deletion[7,8,9], self-destructing based on distributed hash table(DHT) [10] and two-party assured deletion based on key modulation function(KMF) [11] are being used and supported by these schemes.

Since all these schemes are based on cryptography, so these have the below major limitation.

**Computation and resource overhead** – Cloud users have to encrypt data before storing to the cloud storage, so it is overhead for user.

**Encrypted data Leakage possibilities** – After deletion of the data in these schemes, encrypted data remain exist in the server. Encrypted data may be leaked due to the disclosure of the encryption keys or any other attack.

**Difficulty in Encryption Computation** - Cloud data encryption makes computation difficult.

**Key Management** – Cloud users have to encrypt data before storing to the cloud storage. This requires key management. Protecting the keys is also a task to do as overhead in encryption.

**Limited Data Usage** – Encryption results the limitation of data usages. For example, on encrypted data, user cannot do sorting or searching operation as it is in form of unrecognized cipher text. A naïve solution need to use to perform these operations on encrypted data.

**Data Sharing** – Data sharing and working on same data becomes difficult on shared Encrypted Cloud data. That effects one of the key benefits of Cloud Computing i.e. high availability of data regardless of location and time.

## 5. Conclusion

Assured data deletion is a main obstacle and concern for potential cloud customer in public cloud adoption. It is significant for both the Cloud users and the provider. The issues and challenges related with encryption method for assured deletion should be more explored and solved. Researchers have an area and possibility to improve the current encryption schemes. Trusted Computing concept with Encryption could be an alternative. This study gives us a research agenda for me and also for research community. It has still more scope for improvement and innovations. Any further development will accelerate companies to achieve more efficient use of Cloud computing and go for the adoption of cloud computing.

## Acknowledgement

This paper is made possible through the help and support of Head of CSE department, Glocal University, Saharanpur, Utter Pradesh, India and my research colleagues. The product of paper is not possible without them.

## References

[1] P. Mell and T. Grance, “The NIST definition of cloud computing,” 2011; http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

[2] J. Reardon, D. Basin, and S. Capkun, “Sok: Secure data deletion,” in Security and Privacy (SP), IEEE Symposium on pages 301-315, 2013.

[3] J. Reardon, H. Ritzdorf, D. Basin, and S. Capkun, “Secure data deletion from persistent media,” in Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security, pages 271 - 284, 2013.

[4] P. Gutmann, “Secure deletion of data from magnetic and solid state memory,” in Proc. Of USENIX security Symposium, 1996.

[5] Junping Liu, ke Zhou, Liping Pang, Zhilum wang, Yuhui Deng and Den Feng, “ A novel cost effective scrubbing scheme,” in Fifth International joint conference on INC, ISM and ISC 2009.

[6] R. Perlman, “ File system design with assured delete,” in Security in Storage Workshop, 2005. SISW '05, Third IEEE International, 2005.

[7] Y. Tang, P. P. Lee, J. C. Lui, and R. Perlman, “ Fade:Secure overlay cloud storage with file assured deletion,” in Security and Privacy in Communication Networks, pages 380-397. Springer, 2010.

[8] Rahumed, H. Chen, Y. Tang, P. Lee, and J. Lui, “ A secure cloud backup system with assured deletion and version control”, in Parallel Processing Workshops (ICPPW), 2011, 40th International Conference on, pages 160–167, 2011.

[9] Cachin, K. Haralambiev, H.-C. Hsiao, and A. Sorniotti, “ Policy-based secure deletion,” in Proceedings of the 2013, ACM SIGSAC Conference on Computer & Communications Security, CCS ’2013, pages 259–270, New York, NY, USA, 2013.

[10] R. Geambasu, T. Kohno, A. A. Levy, and H. M. Levy, “Vanish: Increasing data privacy with self-destructing data,” in USENIX Security Symposium, pages 299–316, 2009.

[11] Z. Mo, Y. Qiao, and S. Chen, “Two-party fine-grained assured deletion of outsourced data in cloud system,” in Distributed Computing Systems (ICDCS), 2014, IEEE 34th International Conference on, pages 308–317, 2014.

[12] M. Barhamgi, A. K. Bandara, Y. Yu, K. Belhajjame,and B. Nuseibeh, "Protecting privacy in the cloud: Current practices, future directions" in IEEE Computer, 2016.

[13] Digital Ocean. Data leakage, 2013 (accessed Dec 19, 2017).

[14] https://www.digitalocean.com/company/blog/ resolved-lvm-data-issue/

[15] OpenStack. Data privacy concerns, 2017 (accessed December, 2017)

[16] https://docs.openstack.org/security-guide/tenant-data/data-privacy-concerns.html.
[17] Amazon Simple Storage Service (S3) (accessed Dec, 2017)
[18] https://aws.amazon.com/s3/faqs/#storage-management (accessed December, 2017)
[19] The National Cyber Security Centre, Guidance on Secure sanitization of storage media (accessed Dec, 2017) https://www.ncsc.gov.uk/guidance/secure-sanitisation-storage-media.