An Improved Auditing Model for Data Remanence in Cloud Computing

By

Rabia Saleem

Supervisor

Assoc. Prof. Dr. Tauseef Ahmed Rana

A thesis submitted to the Department of Computer Software Engineering.
Military College of Signals (MCS), National University of Sciences and Technology.
Islamabad, Pakistan, in partial fulfillment of the requirements for the degree of MS in Software Engineering

September 2022

An Improved Auditing Model for Data Remanence in Cloud Computing

by

Rabia Saleem

Supervisor

Assoc. Prof. Dr. Tauseef Ahmed Rana

A thesis submitted in conformance with the requirements for

the degree of *Masters of Science* in

Computer Software Engineering.

Department of Computer Software Engineering

Military College of Signals (MCS)

National University of Sciences and Technology (NUST)

Islamabad, Pakistan.

August 2022

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Ms. **Rabia Saleem,** Registration No. **00000275138** of **Military College of Signals** has been vetted by undersigned, found complete in all respect as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial, fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as2 pointed out by GEC members of the student have been also incorporated in the said thesis.

Signature: ___________________________

Name of Supervisor: **Assoc Prof Dr. Tauseef Ahmed Rana**

Date: _______________________________

Signature (HoD): ______________________

Date: _______________________________

Signature (Dean/Principal): _____________

Date: _______________________________

Declaration

I, *Rabia Saleem*, declare that this thesis titled "An Improved Auditing Model for Data Remanence in Cloud Computing", and the work presented in it are my own and has been created by me as a result of my own original research.

I confirm that;

1. This work was done wholly or mainly while in candidature for a Master of Science Degree MS at NUST.
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated
3. Where I have consulted the published work of others, this is always clearly attributed
4. Where I have quoted from the work of others, the source is always given. With the exceptions of such quotations, this thesis is entirely my own work.
5. I have acknowledged all main sources of help.
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

NS Rabia Saleem,

NUST MCS MSSE 25

Copyright Notice

Dedication

“In the name of Allah, the most Beneficent, and the most Merciful"

I dedicate this thesis to my parents, siblings, husband, and supervisor who supported me in each step.

Acknowledgments

All praises to Allah for His blessing and motivation when led me to the completion of the thesis. I want to acknowledge and pay my gratitude to my supervisor Asst. Prof. Dr. Tauseef Ahmed Rana who was extremely supportive and encouraging through this period. His high accessibility and qualitative feedbacks aided a lot in bringing my thesis to its final stages. Also, I am very grateful to my committee members; Assoc. Prof. Dr. Ihtesham-ul-Islam, Asst. Prof. Dr. Ikram Syed, and Asst. Prof. Dr. Imran Qureshi for their constant inputs and remarks along the course of the research. In the end I would like to take the opportunity of thanking my parents, my siblings, and my husband for believing in me and supporting me. Last but not the least, I would like to acknowledge my son for being patient with me and bringing a smile on my face when I was going through the stress and exhaustion that came along with this thesis.

Abstract

The Cloud Computing concept is still under development and is being hampered by problems that are preventing people from adopting it and advancing it. The biggest worry has been a lack of security in terms of data and information storage. A cloud provider server's integrity and confidentiality must be guaranteed by organizations and other entities, among other things. Solutions to improve security models have been looked at (strong authentication, encryption and storage fragmentation before storage, access control policies, etc.).

One of the desired of issue is data remanence which poses a serious risk. How can we be certain that data is actually and properly erased from remote systems when requested? Due to the virtualization technology used in the cloud infrastructure, traditional methods for diminishing data remanence are not enough. Since the data is not stored only on a tangible device like a hard disk at one location/physical machine, but on multiple virtual machines located across the globe. Our work here includes the background and introduction to cloud as well as how data remanence is prevalent in it enough, to raise concern for data security. We propose an improved auditing model that works on seven aspects of the cloud to help prevent or remove data remanence in the cloud computing paradigm. These aspects range from the usage of traditional methods of sanitizing devices in conjunction with techniques, like encryption and ensured data deletion. Our model provides solution for the data security and confidentiality concerns when it comes to the residual data being present in a public cloud. We also analyze the proposed model in light of an eHealth case study. The security concerns in an eHealth cloud are mentioned and how those concerns can be tackled with a performance evaluation using our model. Our future work will be to create a public cloud and visualize our auditing model in a real-time environment to get better results.

Table of Contents

Page

List of Figures

List of Tables

Chapter 1

1. INTRODUCTION

This chapter gives a brief overview of the cloud computing environment on which our work of thesis is based. It also highlights the main concerns and how data remanence is one of the least addressed of them all. The objectives and goals of this work are mentioned as well as the structure of the write up is also detailed.

1.1. Overview

In the present era, there is an increasing need of cloud computing infrastructure amongst home users, organizations, and enterprises. Each customer wants the cloud for several reasons including data storage, remote network access, application software, and servers etc., over the internet which can be easily provided and managed with a minimum of interaction with the cloud service provider (CSP). However, various aspects of security concerns have risen with the ubiquity of cloud computing that are: data-in-motion, data-at-rest, data procession, data lineage, data provenance and data remanence [1 – 5].

Data-in-motion is at risk because of the encryption technologies and network standards that are less feasible when applied in a cloud infrastructure. Protocols like vanilla, File Transfer (FTP) and Hypertext Transfer Protocol (HTTP) have also been tested in the cloud environment but have failed to provide data integrity and a weaker confidentiality level for data-in-transit [2].

In the data-at-rest aspect, the cloud security equals the security of its weakest link. This means that a malicious activity on secured data and a non-malicious entity with an unsecured data, both will result in the security breach of stored data. Additionally, due to the multi-tenant infrastructure on the SaaS and PaaS models, a violation of integrity may also occur due to any unauthorized access on the platforms. There have been solutions and tools created by third-party data security providers, but since the data is not stored on a single platform of an enterprise, such solutions have been partially helpful for the service providers [2].

Data lineage and provenance shows the data trail through applications from its inception and how it could be useful for the auditors for ensuring the data integrity. However, the dynamic and public nature of the cloud infrastructure makes it difficult to keep track of each data entity that is stored on it [3].

Lastly, amongst all the aspects the least addressed one by the service providers is data remanence. Data remanence is the remnants of data that is left behind after a deletion operation. There has been little to no attention given to this aspect because of the distributed cloud environment. How can a data owner be sure that their data is completely eradicated from the cloud after they have deleted it? And that there is no footprint or image of the data entity still available in the cloud memory? Serious concerns like confidential or sensitive information that is stored in cloud has been disclosed to unauthorized personnel because of the residual data still present on the cloud. Therefore, it is significant to give due consideration to the issue of data remanence removal so that to guarantee data security and confidentiality [2, 4].

1.2. Motivation and Problem Statement

Among other security concerns related to data, data remanence has garnered much less attention from the clients and CSP. For example: big cloud service providers like Amazon, IBM and Microsoft have put on claims on the internet that how they sanitize their physical devices before they provide the tenancy to any other client. They state that they are in accordance with the standards like NIST, DoD 5220.22 M and others, however, none of these standards define media sanitization in a cloud infrastructure [5].

It is essential for the client to know that the data they deleted is securely erased from the cloud network and cannot be retrieved through any means. Also, it is significant for the CSP to provide proof of erasure which, then, can act as a critical service differentiator. Data can be exposed and even retrieved from the VM memory either after its termination and reallocation [6]. Authors in [7, 8] have done some work on proofs of secure erasure in mobile embedded devices, however, their underlying assumptions are only feasible for traditional IT infrastructure and not for a dynamic infrastructure like the cloud.

Therefore, an auditing framework or model is necessary to ensure that the data has been securely deleted from the cloud network and that no mirror images or remnants of data are still available.

1.3. Objectives

Following are the main objectives of this thesis:

- To propose an improved auditing model that can work in conjunction of the conventional methods to diminish data remanence in a cloud infrastructure. This includes:
  - Guideline on how to trace and log data for detecting residuals in cloud
  - Standardize the auditing procedure in a cloud environment

- To evaluate the proposed model with the help of an eHealth data security case study to determine the effectiveness of the model.

1.4. Thesis Contributions

Consequently, there is a risk that information that has been erased can still be retrieved from memory; this could be a major threat regarding the confidential nature of presumed deleted files (passwords, encryption keys, private account information, financial or health data…). Therefore, the proposed research work will help in diminishing that problem and can help the different sectors of a society to store their data in the cloud with enough trust that their data is secure.

Also, in a dynamic infrastructure like cloud, it is difficult to eliminate data remanence. But with the help of this research, the data remanence issue can be diminished to quite an extent. Since there is no standard auditing protocol for the cloud environment, this research work can be used as a starting point to define a proper standardize structure for the same. With the use of an auditing model, the CSPs will have a differentiating factor for the clients who are looking for a trustworthy cloud platform.

1.5. Thesis Organization

This thesis is organized as follows:

- **Chapter 2** contains the background and general information about the cloud computing paradigm. It also sheds light on the current threats pertaining to the infrastructure.

- **Chapter 3** contains the literature review for data remanence in cloud computing and what are the existing solutions that have been used for the prevention or removal of data remanence in conventional/traditional computing environment.

- **Chapter 4** presents our proposed auditing model that can be used for handling the data remanence issue in the cloud environment. This chapter has seven aspects of the model laid out. These seven aspects are traditional methods, transparency, guaranteed data deletion, encryption, service level agreement (SLA), certification and isolated environments.

- **Chapter 5** evaluates the proposed model by using an eHealth cloud case study. In this chapter the main security concerns are described and how our model can help in providing a solution to those concerns.

- **Chapter 6** concludes the thesis write-up along with an insight to the future work directions.

Chapter 2

2. CLOUD COMPUTING ("THE CLOUD")

This chapter summarizes the cloud computing working architecture which include its characteristics, service delivery models, and deployment models. Furthermore, there are threats mentioned that are currently prevailing in this architecture.

2.1. Background

Over the past decade, the cloud computing paradigm has been ubiquitous in almost all aspects of the society especially in multinational industries and academia. According to the definition provided by the National Institute of Science and Technology (NIST), "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction…" [9].

It has also listed the five main characteristic in a cloud infrastructure, as follows:

1. **On-demand self-service:** Services are available every time on user's disposal.

2. **Expanded network authorization:** Services are available on the internet and can be accessed by the users having a stable network connection.

3. **Shared resource pooling:** Resource pooling allows for cloud to be location independent and helpful in providing the same service to multiple users at the same time.

4. **Flexible scalability:** Each cloud entity (such as storage, servers, computing power, etc.) is flexible in terms of architecture and therefore, it can be available anywhere to be used and released the same

5. **Measured resources:** All resources are calculated based on their usage/purchasing accordingly, for each user.

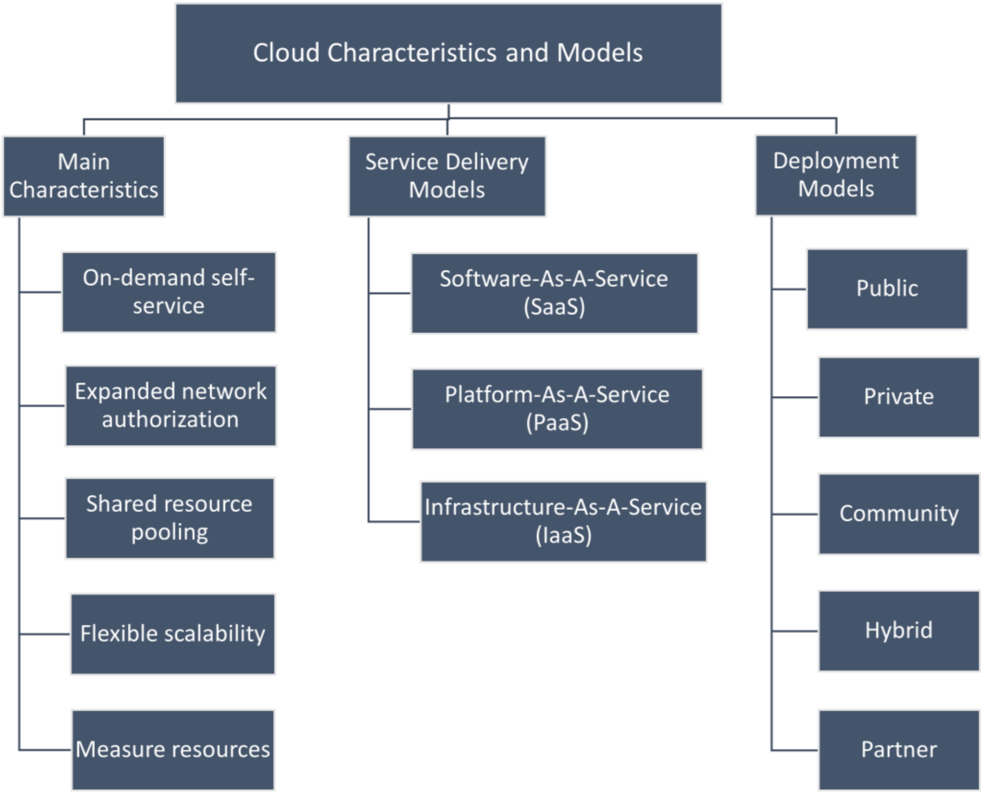


Figure 1 - Cloud characteristics, service delivery and deployment models

2.2. Cloud Computing Service Delivery Models

Cloud delivery models are also known as the SPI Model that is short for Software, Platform, and Infrastructure. Figure 1 summarizes the cloud characteristics, service and deployment models. NIST has defined the following three models at different layers of a business model for cloud computing [4, 10, 11, 12]:

2.2.1. Software-As-A-Service (SaaS)

SaaS is the topmost layer which serves the user with the availability of software or applications over the cloud infrastructure through the Internet. The user rents the services instead of purchasing it (pay-as-you-go) which means that there is no need for any installation, maintenance or initial purchasing cost when buying any software or application. The web browser can be used to view and access these applications. The real-world applications include Microsoft Office 365, Google Apps like Gmail, Salesforce, Taleo, Dropbox, etc.

2.2.2. Platform-As-A-Service (PaaS)

PaaS is the middle layer which provides services for developers by giving them accesses to the different development platforms and resources (IDE, Database Management Systems, Operating systems, toolkits…), that are also offered by the cloud vendor itself. The platform can be used for development, maintenance, and deployment of application software. The examples of PaaS products are WordPress, Google App Engine, Azure, GoDaddy, etc.

2.2.3. Infrastructure-As-A-Service (IaaS)

IaaS is the lowest layer which offers the user an infrastructure as a product to run their solutions. The infrastructure provides virtual access to hardware entities like server, routers, connectors, etc. as well as to devices like networks, data storage, web servers, etc. The model works on a pay-per-use fees and is useful in diminishing the initial cost of computing hardware. The IaaS products that are popular in the market today are EC2 bluecloud, Amazon Web Services, Google Compute Engine, CISCO Meta-cloud, etc.

2.3. Cloud Computing Deployment Models

Cloud computing has various types of deployment models amongst which, the main ones are as follows [10, 12]:

2.3.1. Public

A public cloud is accessible for general users who use all the resources made available by the vendor through the Internet. However, the ownership is retained by the provider and hence, manages the cloud itself also.

2.3.2. Private

A private cloud is owned by an individual user or organization which can only be accessible and managed in a private networking space. A private cloud can be managed either by the provider or a third party, on-premises, or off-premises.

2.3.3. Community

A community cloud serves those enterprises/organizations that have shared interests (business statements, policies, conformance to standards, etc.). This type of cloud can be managed by either the clients themselves or the service provider, on-premises, or off-premises.

2.3.4. Hybrid

A hybrid cloud, as the name suggests, is a hybrid of different cloud models (public, private or community) which is mostly needed to cater to the enterprise's needs. Although, these models remain segregated but there is an application and data scalability amongst them.

2.3.5. Partner

The cloud provider extends its resource to a well-defined partner. This partner cloud provides resources to the clients through their own console. They also manage the fees and other administrative help related to billing on their own.

2.4. Threats to Cloud Computing

Due to the distributed nature of cloud computing, along with the multi-tenancy and virtualizations technologies applied to achieve the same, several concerns have risen that has stopped a lot of organizations to completely switch to the infrastructure. The Computer Security Alliance (CSA) group has identified the threats that come with the cloud infrastructure, they are mentioned as follows [11]:

- Unethical use of the cloud
- Shared resources issues
- Hardware failure
- Data Loss and Leakage
- Zero-day risks
- Malicious activities from inside the cloud
- Cloud-related Malware
- Natural calamities
- Cloud service unavailability

- Hardware crash
- Faulty integrated devices and interfaces

Amongst the above-mentioned threats, data loss and leakage has found to be the second most common threat when it comes to the cloud computing paradigm. This threat occurs because of the transmission of data between various data centers and to/from the clients' system. This transmission also includes the type of execution mode set for the cloud [10]. The same goes for data storage in the cloud as well as where the data is being stored in current mode but since, the cloud uses virtualization technology, the data can be stored in a virtual machine (VM) in one instant and transferred to the other VM in the next instant. These raise issues for privacy, integrity, and security for the data in transit and in motion.

2.4.1. Aspects of Data Security in Cloud

For data security, following factors are taken into consideration to minimize the related risks [13]:

2.4.1.1. Data-in-motion

This refers to the data that is transferred either from the service provider to the client or vice versa, over the Internet, or in public/private network. Data security is needed because the data is often considered less protected when it is in motion because of the already defined protocols (HTTP, FTP, Vanilla, etc.) at web-level, which can provide confidentiality but not integrity.

2.4.1.2. Data-at-rest

This aspect refers to the data that is being stored on a tangible device. The major concern with static data is the inactive data stored that is liable to be attacked by malicious user to avoid detection. In a SaaS/PaaS architecture, multi-tenancy occurs which can lead to unauthorized access of the data like the New Jersey data center breach [14].

2.4.1.3. Data lineage

Data lineage provides the basic history of the data from its origin to the destination and in-between. This aspect is critical for auditing and challenging in a dynamic infrastructure like the cloud. The three main concerns of data lineage are its origin, its destination and where it is transmuted within an organization.

2.4.1.4. Data Provenance

Data provenance is to maintain the accuracy of the data and assure that through computing evaluation. The integrity of the stored data in a cloud can be verified by the clients or customers. However, data provenance is much more challenging than data lineage because it provides a detailed history of the data. These include operations like insert, edit and deletion of a data at a minute level.

2.4.1.5. Data remanence

Data remanence refers to the residual or leaked data after a data deletion or removal operation. This leaked data occurs due to the incomplete deletion operation or a physical device with fuller storge capacity. This aspect is more considerate in the public cloud as compared to the private network where it poses minimal threats. Data footprints are available in the cloud that can be misused once the attackers get their hands on them without any knowledge to the service provider or the data owner. Data remanence has found to be the most under researched and sometimes ignored aspect by the cloud service provider [1, 2, 3, 4, 10, 11, 13].

Chapter 3

3. DATA REMANENCE IN THE CLOUD

This chapter presents the existing solutions of media sanitization for conventional and traditional computing environments. These include the standardized methods as well as the other methods that can be used in conjunction for better efficiency and performance. Also there are certain gaps identified in this chapter.

3.1. Introduction

Sensitive and confidential information needs to be secured from any unauthorized access by an intruder (internal or external) as well as data sanitization should be performed accurately to prevent data recovery. Data remanence is defined as the existence of residual data after its secure erasure, reformation, or reallocation to another party. This residual data is prone to data recovery through several data recovery methods and technical forensic methodologies. This is a significant risk for some confidential presumedly deleted files (like military, healthcare data, financial and government-based information, passwords, users' IDs, etc.). For example, if a user deletes or moves the file to the recycle bin, it is not actually deleted from the memory but is just removed from the system so that the user cannot see. However, the digital footprints are available to be manually recovered hence, causing the concern for data remanence [2, 4, 10].

According to the authors in [5, 6, 15], there are three main reasons that cause the data remanence liability in the cloud:

1. Inappropriate sterilization of data
2. Multi-tenancy
3. Negligent administration of the CSP

Data remanence in flash memory devices is relevant to the computer forensics. It was investigated that the transistor in these devices do not restore completely to the initial state even after the deletion/removal process and therefore, the malicious intruder can retrieve the information from these devices. Data remanence has been known to be prevalent in RAM, ROM, EEPROM, EPROM, flash drives, etc., which gives room for memory vulnerability to attacks [4 – 6], [15 – 17].

In their study, Albelooshi et. al [6] have demonstrated the presence of remanent data on both the hard disks and virtual memory. Since the virtualization technology is a significant factor to cloud computing as it makes the large provisioning of on-demand resource, possible. A virtual infrastructure allows for multiple operating systems to be run on one system by sharing and pooling of resources that results in cost reductions as well as makes sure the service availability according to the users' requirements. Methods like file swapping, ballooning and load balancing are used to assure good measure performance and efficiency [4]. However, the threats of the same are as real as they can be when it come to the confidentiality of the data stored on a virtual and mutually shared memory device.

One of the many concerns are related to the complete deletion of data and making sure that the data once erased cannot be recovered through any means. The failure to properly remove data, however, could result in inadvertent disclosures and costly fines as well as reputational harm. The problems with partial deletion are widely known in non-cloud environments. To the best of our knowledge, no systematic investigation of assured deletion problems in public clouds has yet been conducted. Sensitive information about tenants may be exposed accidentally or due to premeditated erasure of data. Financial losses (for both customers and providers, for example, through regulatory fines) and reputational damage are two of the significant costs involved with such disclosures.

On the other hand, it's crucial to keep in mind that cloud providers value certain deletion promises just as much as renters do. It is crucial to get guarantees that data will be handled and destroyed in accordance with the agreement from the tenant's point of view. Such assurances, seen from the perspective of a cloud provider, are required to satisfy the needs and expectations of tenants while also adhering to the data protection laws of various nations and regions. Additionally, a cloud provider's deletion guarantees may start to set them apart from competitors. Some of the existing solutions related to data remanence are mentioned below, given the fact that there are limitations when the cloud is considered.

3.2. Existing solutions for Data Remanence

There are some existing approaches and techniques that have been incorporated for preventing or removing the remnant data, all of which are being briefly discussed as below. However, these solutions cater less to the cloud computing paradigm as we will see the amount of relevance they produce to the infrastructure [18 – 22].

3.2.1 Sanitization

According to the NIST 800-88 Media Sanitization Guideline, sanitization is a process that is carried out on a target data so that its recovery is made impractical with a given amount of computing labor [22]. This process is essential to each step in an information system life cycle in which data disposition decisions are carried out. There are three types of software and/or hardware data sanitization methods known as *clear*, *purge* and *destroy* as data remanence removal techniques (Figure 2).

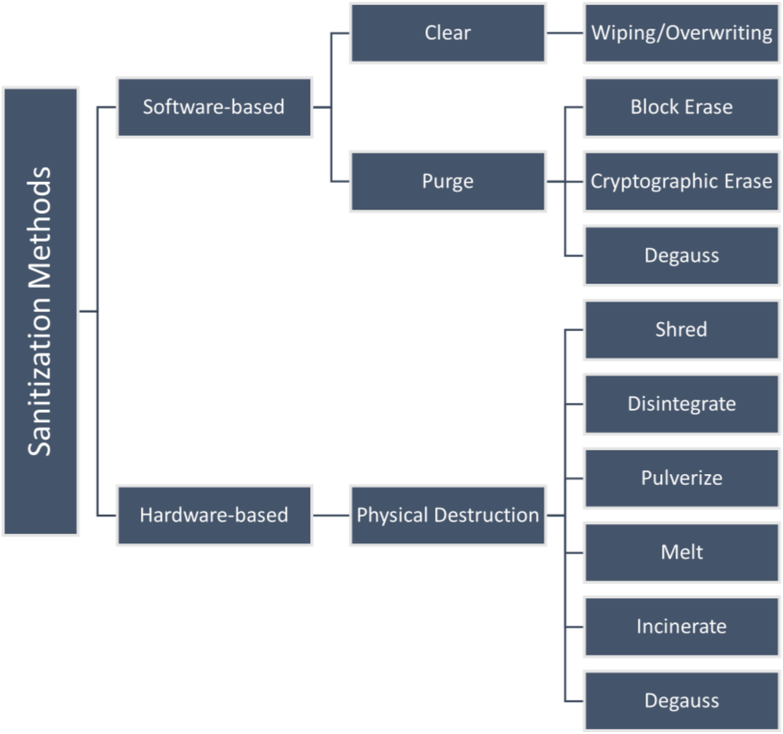


Figure 2 - Sanitization Methods from NIST Guidelines for Media Sanitization

3.2.1.1. Clear

The clear method comes under the software-based mechanism where the data storages are either reset or overwritten/wiped. A number of different standards by multiple organizations have been defined for the overwriting process. The most used ones include the Gutmann method in which the data is overwritten over a series of 1-35 passes with successive bit pattern of 0's, 1's or random number rewrites [17]. Similarly, Bruce Schneir presents a 7-pass wiping mechanism. However, some of the segments on the medium are unreachable and therefore complete sanitization is not possible. Also, various overwriting methods like the Gutmann, Schneir-7, file system delete, are susceptible to data recovery even on storages like flash drives or system hard disks.

While in the cloud environment, the overwrite methods have the limitation of longer processing time as well as it requires the complete information of the type of storage devices used keeping in mind the virtualization of these devices, and the CSP's protocols of sanitization.

3.2.1.2. Purge

Purge is also a software-based method that uses physical and logical mechanisms to make the stored data recovery impractical. The purge method is additionally classified as Block Erase and Cryptographic Erase which use some industry-related standards and state-of-the-art techniques that are each specified according to the target media. The Cryptographic erase sanitizes the encryption key for encrypted data so that even if the data is recovered, it cannot be decrypted because of the absence of the encryption key. Degaussing is a hardware-based purging method that uses a magnetic field for root-deep sanitization of the data storage.

However, the purging methods pose great threats in the cloud environment because of their requirement of media at hand, encryption keys or the hard storage itself. Since the Data owner is not in possession of either of these, therefore the CSP will perform these tasks. Alarmingly, these points should be a part of the service level agreement (SLA) with the users or at least it should be noted that the CSPs follow some standards or guidelines [4].

3.2.1.3. Destroy

According to the NIST guideline, the physical destruction of the storage media is classified in several methods like shredding, disintegration, pulverize, melt, incinerate, etc. The main aim of physical destruction is to prevent data recovery through means like device interface, magnetic segments, etc., but, on the other hand, this method limits its reuse or reselling of the target medium. Furthermore, physical destruction is costly and damaging to the environment atmosphere.

As far as cloud is concerned, it is impractical because the physical presence of the device is mandatory, and that the data owner does not have any access to them. Therefore, the CSP will be responsible for the process which is question of trust on the data owner's end.

3.2.2 Encryption

Traditional computing architecture uses the encryption method for their data loss or leakage. A survey in [10] has done a detailed analysis of encryption mechanisms like Transport Layer Security (TLS), Advanced Encryption Standard (AES), and Secure Hash Algorithms (SHA) that are incorporated for data protection against loss or leakage [24].

Unfortunately, encrypted data is infeasible in the cloud paradigm because of its overheads like processing time, cost, need of resources, etc. For private cloud encryption/decryption occurs at the private server where the encryption keys are securely stored and hence, safe from any outsider attacks. However, for other types of cloud the encrypted data cannot be processed in the cloud as it must be decrypted first, because of which there is an increase in processing time. Gentry in [23] tried to solve this problem with a Fully Homomorphic Encryption (FHE) which is an encryption algorithm that allows for any processing to take place on the encrypted data and produce the same results as they will present on plaintext. But this method was still insufficient for the vastness of the cloud.

Furthermore, the keys are stored in the Random Access Memory (RAM) which can easily be manipulated and once they are acquired, the data is not protected anymore. This leads to the problem of difficulty in key management [2, 4]. Lastly, it has been noticed that the encryption technologies are still very immature and incapable of protecting in a cloud infrastructure due to its multi-tenancy and distributed nature.

3.2.3 Data Remanence Standards

Table 1 shows some of the standards available for removal or prevention of data remanence in a computing environment [22, 25 – 32]:

Unfortunately, none of these standards provides mechanisms for the removal or prevention of data remanence in a cloud environment. In short, a cloud standard for data remanence is yet to be delivered. As well as the auditing standards too, come short for the cloud computing environment due to its physical need of the devices to be available. This is, however, not possible because of the virtualization technology in the cloud.

Sr. No.	**Standard Name**	**Purpose**
1	NIST 800-88	Media Sanitization
2	DoD5220.22-M	Remanence Security
3	Navy NAVSO P5239-26	Remanence Security
4	Army AR380-19	Information Systems Security
5	RCMP B2-002	Media Sanitization and Security
6	ADISA	IT Assets Security and Sanitization
7	ASD ISM 2014	Information Security
8	AFSSI8580	Remanence Security
9	GCSB NZISM 2017	Information Security
10	GDPR	Data Protection and Sanitization

Table 1 - Standards for Data Remanence

3.2.4 Third-Party Auditing

As we have seen in the discussion above that the weakest link in the data security in a cloud network is the CSP itself. The data owner puts their trust in the service provider however the CSPs fail to accurately estimate the risk or even consider data remanence a legitimate risk. They work on assumptions where they believe that the remnant data will be of little to no use for a malicious outsider.

There have been solutions suggested in the studies above, amongst which the notable one was Third-party Auditing (TPA). For example, in study [4] they suggested that the virtual machines created should be sanitized remotely by the CSP as well as an auditor should be available to supervise on how the CSP plans the process. However, they also mention that it is a rather difficult job to perform auditing in the cloud infrastructure. On the other hand, [10] recommend some characteristics that must be available during the third-party auditing process. These include the non-retrieval of the copy of the data, maintaining data confidentiality through encryption, and data verification for integrity upon the data owner’s request.

Additionally, [3] argues that data remanence remains an open issue for the data auditing needs of the clients requests as well as amongst the CSPs. The survey provides information on the number of service providers who give complete details of their compliance to standards however, unfortunately, none of the service providers offer any answers to the user data security auditing of the service users, especially, in the notion of data remanence. It was further discussed that how providing this feature to its users regarding the data security could prove a critical differentiator for CSPs amongst the other providers.

.

Chapter 4

4. PROPOSED AUDITING MODEL FOR DATA REMANENCE

Our proposed model in this chapter covers the seven aspects of the cloud computing paradigm that can ensure the little to no presence of residual data in the public cloud. These aspects include the traditional sanitization methods, transparency, guaranteed data deletion, encryption, service level agreements (SLAs), certifications, and isolated environments. SLAs along with the Trusted computing architecture using transparency and certifications can help build the trust of the users on their service providers. Additionally, this can also help the CSPs by differentiating them from the others in form of the Trust-As-A-Service (TaaS) models for their customers. On the other hand, encryption and guaranteed data deletion will help in eliminating the remnants as well as keeping the data private, even from a TPA. Lastly, traditional methods can be used whenever the tangible devices need sanitizations or if the CSP has cluttered the to-be-destroyed data to one device for one final destruction.

In a cloud computing infrastructure, there is a CSP that owns and manages cloud computing systems and has large resources and expertise in creating and managing cloud servers. Customers of the CSP are those who need data to be kept in the cloud and who use the cloud for processing and computing. A pay-per-use billing approach is used since the service charge will be disclosed before adopting the cloud computing service, and the customer has a long-term contract with the CSP. The customer may choose to request verification from a third-party auditor if they are uncertain about whether the data processing was carried out honestly and correctly.

The TPA can be compared to a highly regarded, completely trustworthy, honest, and uncorruptible regulatory agency of the government or an auditing firm. It not only has the power to judge whether the service fee levied by the CSP is appropriate, but it also possesses the knowledge and skills that the customers might lack to evaluate and record the data on cloud resource utilization on their behalf. The auditor has the authority to order a fine from the CSP in the form of a compensation to customers if they "find" that the CSP had lacked in any way regarding the security of data, including data remanence.

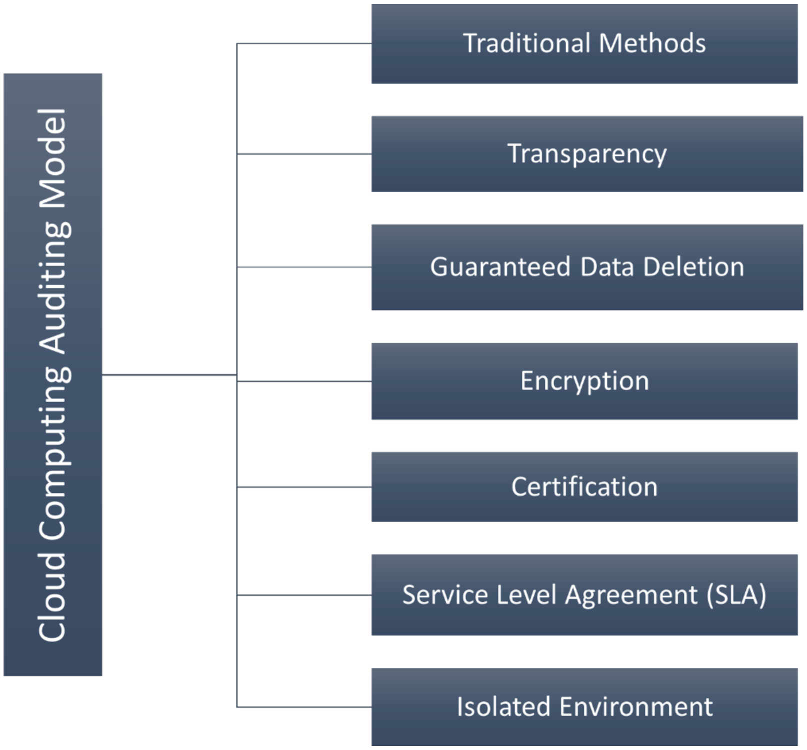


Figure 3 - The proposed Cloud Computing Auditing Model for Data Remanence

Since there is an unavailability of a proper cloud specific standards or auditing mechanisms for prevention/removal of data remanence, we propose an improved auditing model for the same. This model will aid into the auditor to formally audit the cloud infrastructure for the necessary safety and security requirements as well as its ability to trace down the residual data still present in the cloud data storage. The audit process will further increase the trust between the service provider and the consumer with respect to data confidentiality and privacy. There are multiple aspects in the proposed model which can be used all at once or according to the needs at hand. Figure 3 gives an overview of the model with its aspects mentioned.

4.1. Traditional Methods

According to NIST Guidelines for Media Sanitization [22], there are three types of sanitization methods named *clear*, *purge* and *destroy*. Although these three methods have been defined in Section 3.2.1 of this thesis, here these will be discussed considering the cloud environment with the auditing process involved.

Since we know that cloud uses the virtualization technology for maximum support of services and scalability, at the very root it uses tangible hardware devices to keep the system up and running. Therefore, traditional methods of media sanitization are necessary for the erasure of data remnants residing in those devices. While most devices offer some sort of Clear support, not all of them have a trustworthy Purge mechanism. The owner of the media may decide to take the risk of using Clear methods on media containing moderately sensitive information even though it is possible for someone with the necessary time, resources, and expertise to retrieve some of the data.

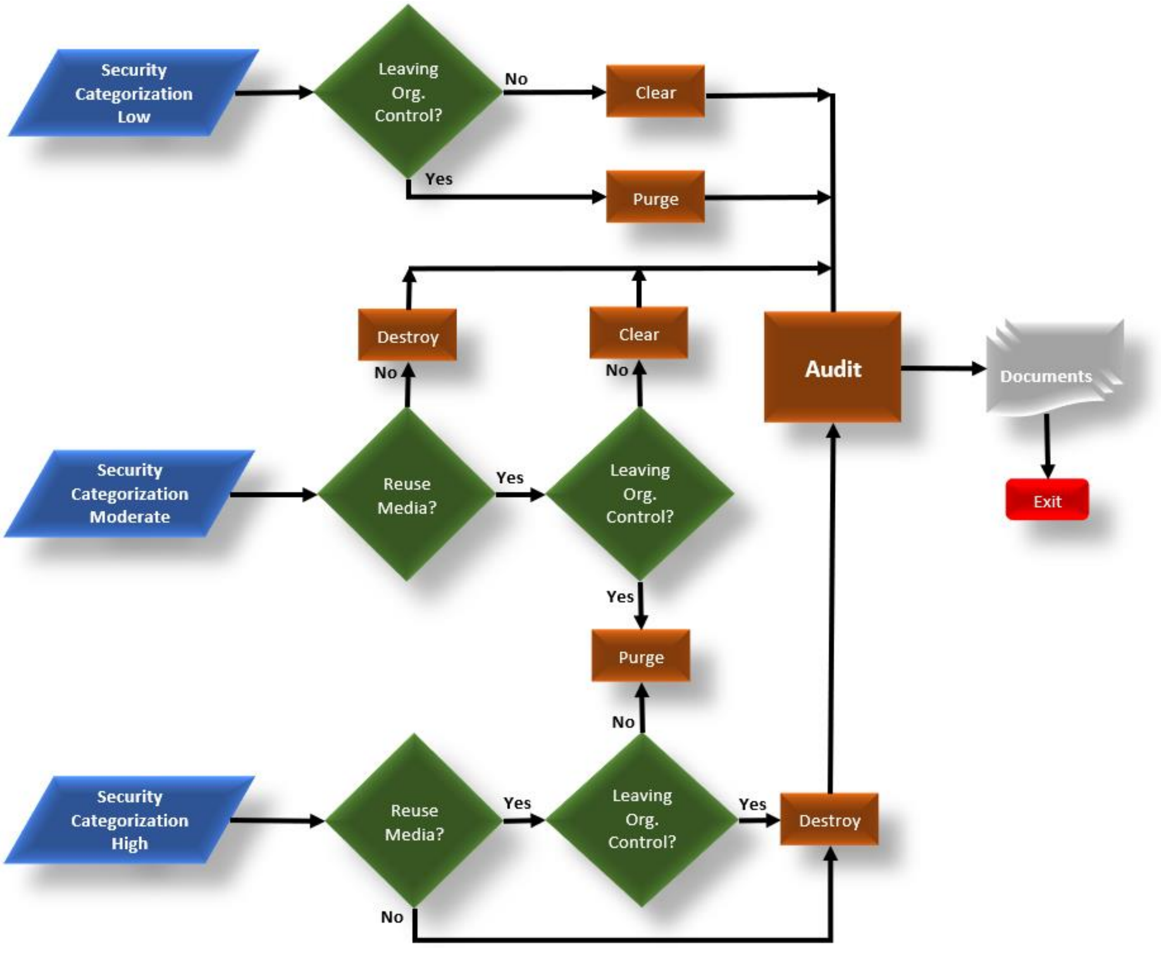


Figure 4 - Sanitization and Disposition Decision Flow with Auditing

When focusing on environmental issues, the desire to reuse the media (either within the organization or by marketing or donating the media), the cost of a media or media device, or challenges in physically destroying some types of media, Purge (and Clear, where applicable) may be more appropriate than Destroy. The risk choice should take into account the potential repercussions of disclosing information that can be obtained through the media, the cost and effectiveness of information retrieval, and the cost and effectiveness of sanitization. The period of time the data would be critical should also be taken into account. Various environments may have various values for these things.

The specifications in Figure 4 regarding the decision making, can be used by organizations to help them decide which media should be sanitized in accordance with the security classification of the information's confidentiality. The decision-making process is based on the information's confidentiality rather than the medium. The type of media will then have an impact on the technique utilized to accomplish this sanitization goal once companies have decided which type of sanitization is ideal for their specific scenario. Each of the three mechanisms conjoin to the auditing process in the end where the auditors (internal or external) will inspect the carried process and determine their efficiency. For example:

1. **Clear** will have audits will be used to ensure that whether the CSP has overwritten all the data or used factory reset (where rewriting is not possible), once the data was released by the user.

2. **Purge** applies physical or logical techniques that render Target Data recovery infeasible using state of the art laboratory techniques (like Cryptographic erase, block erase and overwriting). Auditors will be responsible for inspecting the laboratory devices and their techniques along with the efficiency they provide.

3. **Destroy** renders Target Data recovery infeasible and prevents further reuse of the media device for storage, using state of the art laboratory techniques. Auditors will be needed to oversee these processes while they are being executed.

4.2. Transparency

Clients can make an informed decision about a cloud's reliability based on profiles and security guarantees by using transparency. Consumers can learn about a cloud provider's strengths and shortcomings through the reflection mechanisms of their security profile, as well as how their organizational security rules will be handled. Then, businesses can decide if they need further protection to address any flaws they discover in the cloud. Customers could locate the geographic footprint of various networks in the cloud computing infrastructure using this capacity in conjunction with an automatic tracing facility, allowing them to understand where their data is managed and stored.

It has been determined that transparency is necessary to build trust in a cloud service. The disclosure of security and privacy mechanisms that are employed to safeguard the data of consumers is one aspect of this. More crucially, it describes disclosing how one customer's actual data is handled.

This may, for instance, imply that a consumer could at any time check to see exactly where physically their information is secured. Log files (for keeping geographical location addresses and data protection profiles) are another intriguing method to promote transparency because they might also provide information about when and how data was securely erased. Further information on logging and auditing are discussed in detail in the next section.

To establish the visualization between virtual and actual resources, data could once more be traced up to the physical layer utilizing provenance. Additionally, it would demonstrate to cloud users how virtual locations and actual, physical server locations are connected or mapped together. Additionally, this mapping will enable physical confirmation of deletions made in the virtual world, providing deletion proof.

4.3. Guaranteed Data Deletion

From the standpoint of the user, guaranteed data erasure is a crucial component of reliable cloud services and significant to remove or diminish data remanence. This is a challenging undertaking from the CSP's point of view because of the distributed nature and intended redundancy, especially if ensured deletion or physical destruction must be used.

The storage provider might aggregate data with comparable deletion dates on a single physical device if it knew in advance when data should be removed (for example, the user demanding deletion after 30 days) however, multiple copies of the data would require this process for more than one device. The entire device would then be safely removed at the appropriate time using any of the techniques mentioned in section 4.1. Data segregation methods could be designed to separate data that would need certain erasure to ensure deletion without affecting service. Sensitive data can be subject to limitations. Sensitive data (like military, health data, encryption keys, personal data, etc.), for instance, might only be permitted in specific cloud locations.

4.3.1. Audit Trails and Logging

In the context of computers, a log is the automatically created and time-stamped record of occurrences pertinent to a specific system. While a set of records of computer activities, about an operating system, a program, or user behavior, is known as an audit trail.

Logs are sought after by many real-world applications for forensic analysis. However, logging can be time-consuming and vulnerable to data-tampering assaults. Crosby et. al [34] addresses the situation of an unreliable logger that is used by a number of clients who want to save their events in the log and is held accountable by various auditors who will ask the logger to provide evidence of its proper operation.

They also offer the log server a versatile method of presenting authorized and tamper-evident query results for all incidents that match a predicate. This can essentially allow log servers to carefully remove old events in a way that is mutually acceptable while producing effective evidence that no unwarranted events were removed. These events can also be used to populate to form an audit trail. This audit trail can provide proofs regarding various instances like deletion requests by the user and whether these requests were carried as well as to what extent.

The log is not tamper-evident if an untrusted logger is aware that a recently added event or returned commitment won't be audited. As a result, any tampering with the newly added event or the events corrected by that commitment will go undetected. A tamper-evident log necessitates regular audits to prevent this. Additionally, they suggest a tree-based historical data structure that is progressive for all auditing and query operations in order to achieve this. Events can be entered into the log at any time, commitments can be created, and audits can be carried out independently of one another. Batching is not applied. Loggers can effectively demonstrate that the order of individual logs committed to over time make coherent claims about the former events using the history tree.

The study focuses on how auditing will be used to find tampering and how to reduce the expenses associated with these audits.

4.3.2. Auditing Strategies

In a variety of circumstances, it's possible that each logged may or may not be of any significance for the auditor. Auditing events added to, or commitments obtained from other clients may not be of interest to clients. It is simple to envision circumstances in which a single logger is shared by several businesses, with each only motivated to check the accuracy of its own data. These companies might manage their own auditors, concentrating on customer commitments and occasionally trading promises with other companies to make sure no forking has taken place.

It is also possible to envision situations in which independent accounting companies run auditing programs against the log servers of their corporate clients. If customers leak the logger's claims to at least one trustworthy auditor who utilizes the log when asking for an updated proof, the log will still be tamper evident.

It is very time-consuming and costly for the auditors to inspect each logged event but on the other hand, too risky to skip nodes which requires for some type of trade-off between the two. There can be time-security trade-offs when skipping nodes. Audits may be performed probabilistically by auditors, who only choose a portion of incoming commitments for auditing. The likelihood of a logger getting away with it becoming unnoticed would become infinitesimally small if they constantly tampered with the log.

4.4. Encryption

The issue of data privacy presents the biggest difficulty in the design of data storage auditing mechanism i.e., the auditing mechanism should protect the data privacy against the auditor as well. This is due to:

- For public data, the auditor can retrieve information by extracting the data packets from the data proof.

- For encrypted data, the auditor might be able to acquire the content keys necessary to decrypt it through specific unique channels.

We suggest using the method in [35] to resolve the data privacy issue that involves creating an encrypted proof with the challenge stamp utilizing the bilinearity property of the bilinear pairing, which the auditor is unable to decrypt. However, the auditor can check the accuracy of the proof without having to decode it. In order to avoid using the hardware token, the user's fuzzy private key in this approach is created using biometric data (such as an iris scan or fingerprint). The plan can still successfully finish the auditing of the data privacy in the meantime. They use a linear design with coding and error-correction procedures to verify the user's identification.

When a user wants to use a cloud storage service, their biometric information, such as their fingerprint, is taken during the user registration step. In uploading data to the cloud, a data owner must first extract biometric information to use as his fuzzy private key and generate a signature key at random. Then, using his signature key, this data owner evaluates credentials for data blocks. Finally, he deletes these communications from the local storage and uploads these data blocks and the credentials set to the cloud.

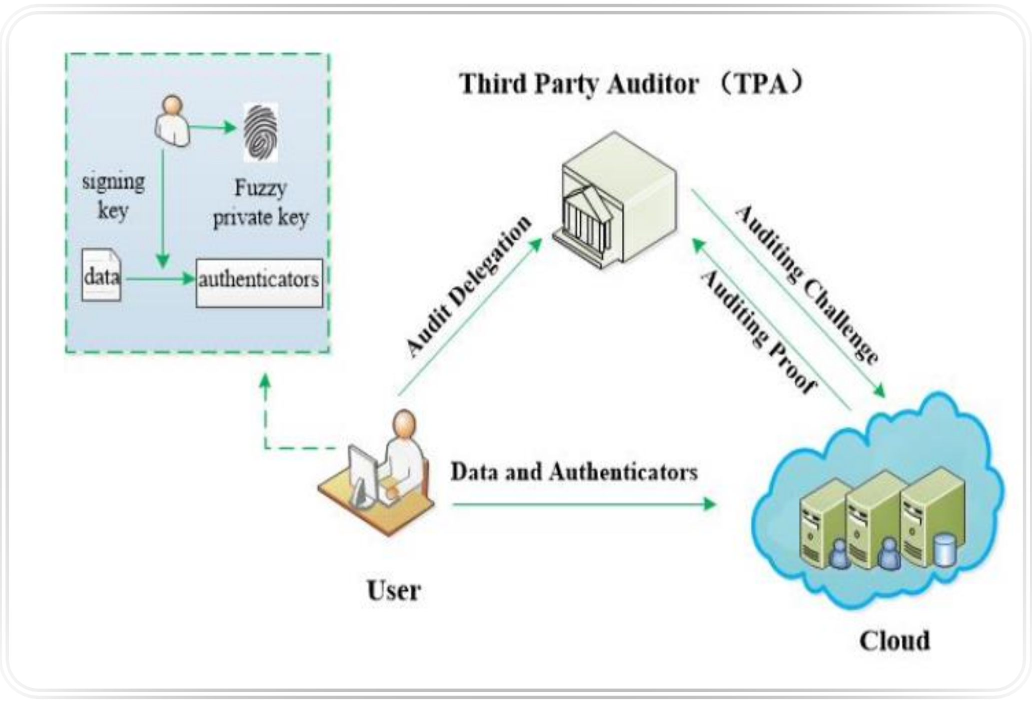


Figure 5 - Encryption model for data auditing

The TPA executes the challenge-response protocol with the cloud during the data integrity auditing phase to determine whether the cloud actually maintains the user's data intact or not. To enhance and speed up the process, encryption and sanitization may be combined by deleting only the portions of the disc that contain keys.

Additionally, data will have encrypted tag that will remain within till the end of data’s lifetime. This tag can only be decrypted by the data owner and can replicate itself if there are more copies of data being created on other locations (physical or virtual). This tag will also be saved in the log files which can then help the auditors to learn the paths of the data through audit trails as to where the data has been transferred. Also, whether the data still resides in a particular location or that it has been permanently deleted. These tags will also have a timer set by the user if they want the data to be deleted after a certain period. This tag will be updated throughout the cloud and must be verified by the auditor through the log files and data trails.

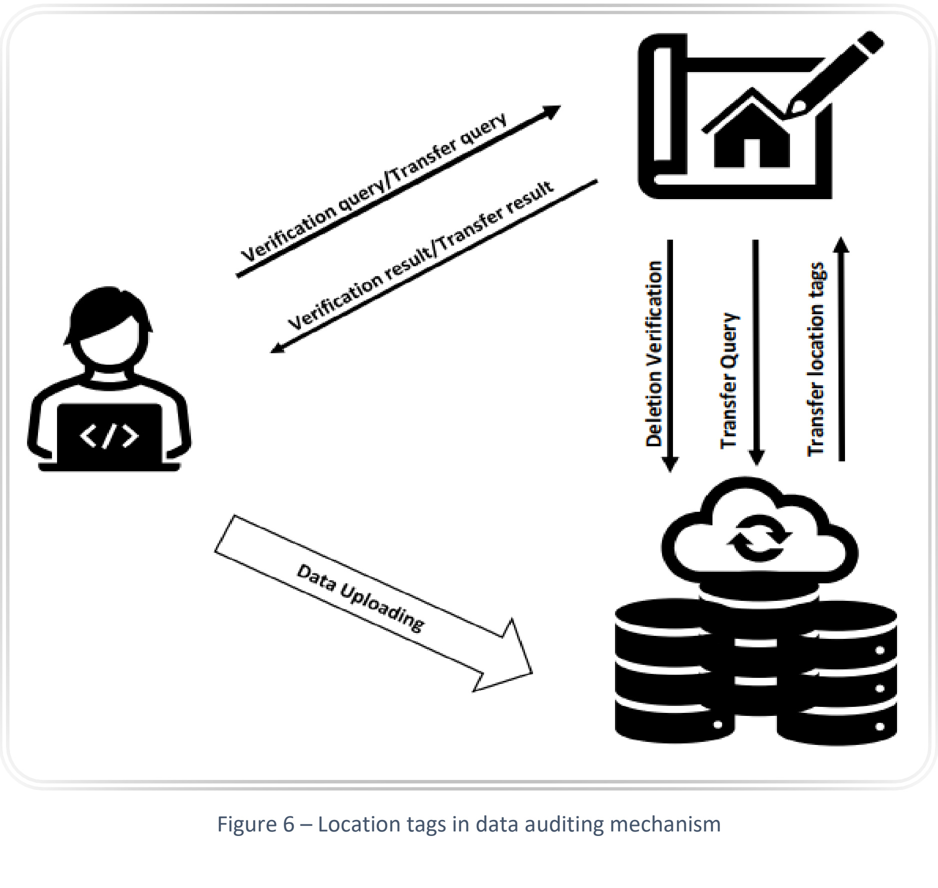


Figure 6 – Location tags in data auditing mechanism

4.5. Service Level Agreements (SLAs)

The Cloud Service Provider agrees to rent data storage on a cost per gigabyte stored and cost per data transfer basis in exchange for the user, or his company, signing a service level agreement (SLA) with the CSP. The data of the company would then be automatically transferred over the storage provider's private wide area network (WAN) or the Internet at the designated time. The user will routinely access and update the data using the several application interfaces he employs under the same preamble, for e.g., web services.

The number of concurrent storages that will be employed to store our data in accordance with the SLA is a source of concern. The user must make sure that the provider includes in the SLA the specific servers and locations in which their data will be delivered simultaneously. Typically, in a cloud, a predetermined number of copies of the same data are made, routinely updated, and stored at different geographical places to ensure that a disaster in one of the sites won't impair any client applications' ability to access data continuously. The exposure window time (time that is needed for a complete updating cycle across various copies of the data) specified in the SLA must also be checked.

Assume that the user decides they no longer need to store their data in the cloud at the end of a certain amount of time. Or perhaps the customer has discovered a superior supplier offering the storage space at a significantly lower cost. The user wishes to remove the data from the vendor at this point or make sure it is no longer there. Even if the data is kept at the current cloud storage provider and is encrypted, it can still be made worthless to prevent privacy leaks even if it falls into the wrong hands. An auditor will examine this SLA to check for any informational omissions on the part of the user and to determine how the data will be permanently deleted from the cloud when the user deletes it.

4.6. Certifications

Making sure security compliance in the dynamic and fluid environment of cloud computing can be challenging. The cloud is opaque, and different cloud providers may offer varying levels of security guarantees. An independent security certification body could authenticate cloud support in terms of their security features and properties in order to fully materialize a trusted cloud model. The certificate would serve as a quality seal, ensuring secure services with a specific level of assurance. It could guarantee that the service security implementation adhered to the published security profiles. The certificate might serve as a model of trust to increase consumers' trust in cloud services.

Verifying the machine-readable security policy statement should be a part of the cloud provider's auditing and certification process. For instance, this would involve verifying claims about the placement of infrastructure. Obeying data protection regulations or having the option to erase files safely.

4.6.1. NIST SP 800-92

A list of steps that can be taken to assure the authenticity, availability, and administration of audit data is provided in NIST SP 800-92, Guide to Computer Security Log Management. Despite the fact that this document is not explicitly focused on cloud services, it does identify difficulties including the necessity of correlating cloud audit incidents with intra-organizational Security Information and Event Management (SIEM), priority, application-specific audit, and storage hierarchy challenges.

4.6.2. ISO/IEC 27000-series

A few standards also deal particularly with the regulation and operation of information security, which includes identifying risks and putting security measures in place to deal with them. The ISO/IEC 27000-series of standards is most likely the series of guidelines for ICT (Information and Communication Technology) system security that is most commonly known and applied. The two main standards are 27001 and 27002, where 27002 describes a set of controls that deal with particular facets of the information security management system and 27001 specifies the requirements for an information security management system.

4.6.2.1. ISO/IEC 27001

An advisory standard, ISO/IEC 27001 is supposed to be construed and implemented to businesses of different shapes and sizes depending on the specific information security threats they confront. Users are given a great deal of freedom to implement the specific information security policies that make sense to them, but this flexibility might make compliance testing more difficult than with some other official certification programs.

4.6.2.2. ISO/IEC 27002

The security standard ISO/IEC 27002 is a series of security requirements (sometimes referred to as benchmarks). It can be argued that a cloud service provider's environment complies with the standard if the design and/or operation of their information security management systems are compatible with the standard (that is, there are no obvious gaps).

4.6.3. ISO/IEC 27017

Additionally, ISO provides standards for security (ISO/IEC 27017) and separately for the protection of Personally Identifiable Information (PII) in Public Clouds (ISO/IEC 27018) that provide particular recommendations for cloud service providers and cloud service clients. Guidelines for information security controls relevant to the provision and use of cloud services are provided by ISO/IEC 27017, Code of practice for information security controls based on ISO/IEC 27002 for cloud services [20]. Given the division of duties that occurs when using and delivering cloud services, ISO/IEC 27017 has specific recommendations to explain cloud service consumer and cloud service provider obligations.

4.6.4. ISO/IEC 27018

The ISO/IEC 27018, Code of Practice for Protection of Personally Identifiable Information (PII) in Public Clouds Acting as PII Processors, contains standards, control goals, and controls targeted at safeguarding PII that is stored or processed by public cloud services. Additional instructions are provided by ISO/IEC 27018 to guarantee that PII is sufficiently safeguarded in accordance with the principles outlined in ISO/IEC 29100 Privacy Framework. The frequently complex requirements that pertain to PII are taken into account by ISO/IEC 27018.

Both ISO/IEC 27017 and ISO/IEC 27018 start with the set of security measures outlined in ISO/IEC 27002 as a base and then add to them as needed for the management of personally identifiable information (PII) stored in a public cloud and the cloud service environment.

Customers of cloud services are urged to search for cloud service providers who adhere to the ISO/IEC 27001 and 27002 information systems security standards. Although this isn't specifically related to cloud computing, its general ideas can nevertheless be used to the delivery of cloud services. A cloud service provider may claim on its own behalf that it complies with a standard, but having the compliance confirmed by an impartial and knowledgeable third party is a noticeably stronger type of attestation. Many cloud service providers, many of them through third-party certifications, currently assert compliance with ISO/IEC 27001 standards.

Additionally, customers want to look for cloud service providers who have ISO/IEC 27017 accreditation. Currently, several major cloud service providers are certified, and the standard is spreading throughout the sector. When PII is involved, clients should also search for ISO/IEC 27018 certification. As was previously said, clients of cloud services shouldn't just rely on certifications; they should also try to grasp the security-related policies and procedures of the cloud service providers (e.g., log retention policy, privileged access policy, change management process, etc.).

4.7. Isolated Environments

As is common in the defense sector, cloud computing providers might create a secure zone for their clients. Cloud enclaving—also known as a network enclave or secure enclave—is a method of workload protection that avoids the complexity and security flaws of network segmentation. The enclave is isolated from the rest of the network, and standard security policies govern access. The conventional capabilities of event detection and mitigation, boundary defense, and monitoring could be offered by enclaves. They could be exclusive to a particular business or to a group of related services that are used by numerous businesses. Providers could compartmentalize users' data at the same time to prevent cross-contamination with that of other users.

The issue of data remanence can be solved for confidential data related to military, economy, health sector, etc., through enclaving. By refusing to reveal the physical architecture of the cloud data center for a service or user, cloud providers can stop attackers from building a cloud cartography of the enclave. Because you're only managing the portion of the cloud that is connected to the user information or processes in an enclave rather than the full cloud, it's simpler to implement the enterprise's auditing policy.

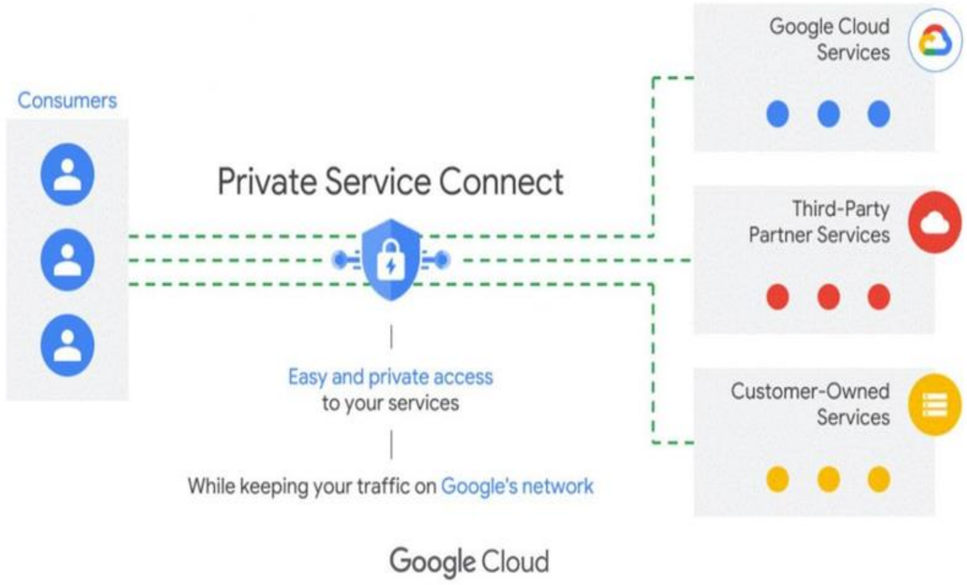


Figure 7 - Google Private Service Connect

In Figure 7 (Source: https://cloud.google.com/blog/products/networking/private-service-connect-is-now-generally-available), Google presents an example of enclaving in cloud known as the Privacy Service Connect. Users can establish a secure and safe connection from a Virtual Private Cloud (VPC) to Google Cloud, other parties, or proprietary services using Private Service Connect. The new service transfers traffic from the customer's VPC to the provider's VPC network via endpoints and service attachments.

Chapter 5

5. E-HEALTH CASE STUDY EVALUATION

This chapter is an evaluation of our proposed model. We use the eHealth cloud case study where its data security threats and limitations are discussed. As well as, how our auditing model can help in mitigating these threats and overcoming the limitations, so that the customer that want to use cloud for health care sector can do so without any hesitation.

5.1. Introduction

Al Issa et. al in their study [40] talk about a comparatively new technology that will significantly change our lives which is cloud computing. This technology enables access to computing services and capabilities from any location, at any time. The healthcare sector is constantly changing, and it is projected that the next healthcare model would be information focused. The cloud technology can help every industry handle complexity and change. This promising technology can assist in fostering collaboration, coordination, and communication among various healthcare practitioners.

The healthcare sector can benefit from using the cloud to maximize every dollar spent. It can provide infrastructure and applications that are quick, adaptable, scalable, and affordable. Electronic health records (EHRs), laboratory information systems, pharmaceutical information systems, and medical photographs can all be stored, managed, protected, shared, and archived using the cloud. Overall, patients will receive better care as a result of current health records and ongoing communication between various healthcare professionals. The biggest barriers to the widespread adoption of the cloud by healthcare providers include security, confidentiality, and trust challenges, in addition to the absence of protocols, regulations, and operational requirements.

Computer security is a rapidly expanding area of computer science that focuses on securing computer systems and digital data against risks and exposures such backdoors, DoS assaults, hardware theft, and data manipulation. The goal of implementing computer security measures is to protect important data and system resources; protecting system resources includes safeguarding the hardware and software of a computer system, whereas data security is more concerned with safeguarding data that is stored or transmitted between computer systems as well as cloud systems.

On the other hand, privacy is one of the primary goals of security; it enforces laws and guidelines that limit how much information about specific people or groups can be viewed, collected, or given to a second or third party. Data privacy and security are more closely tied to data ownership. When using information systems, privacy might be argued to be a moral right for both persons and groups, although computer security is not in and of itself a moral right. There are definitely places where computer security and privacy overlap, making the distinction between the two more difficult. This type of data transmission is an example of a secure implementation when healthcare providers use secure technologies to interact with patients about their health rather than transferring health data via personal e-mail accounts. However, privacy will only make an effort to restrict authorized hospital staff members' access to patient health records.

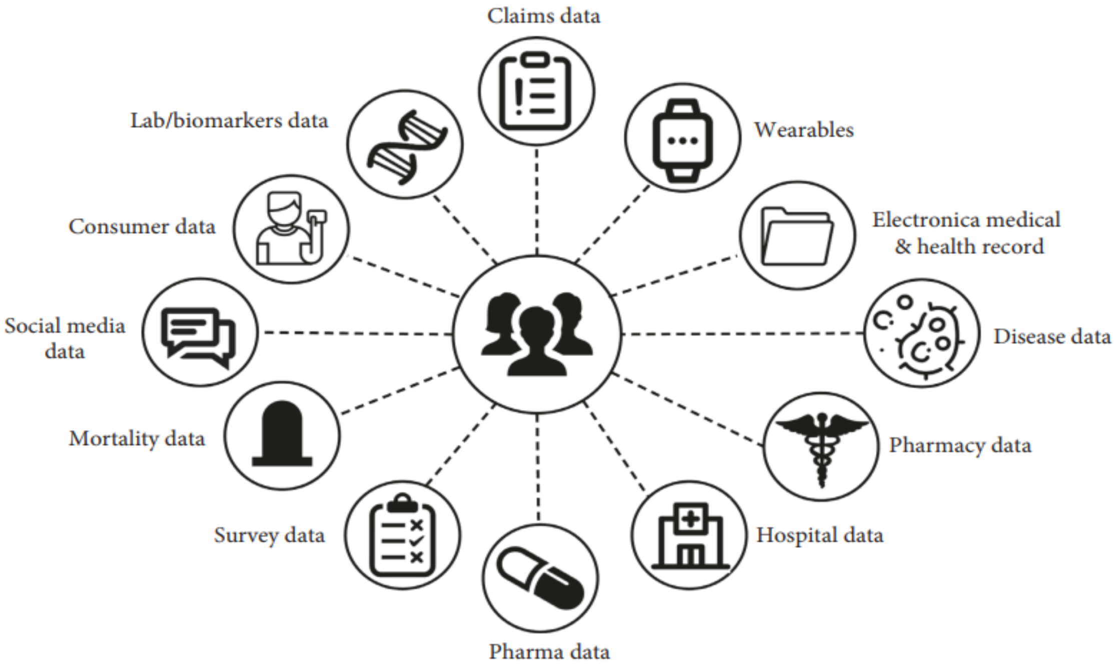


Figure 8 - Information-centric healthcare model

Both benefits and difficulties come with cloud computing. Cloud security is a worry, just like it is for every other IT application. It is susceptible to data loss, theft, and malicious assaults because it typically operates in a shared and open environment. One of the major issues preventing the widespread adoption of the cloud in the healthcare sector is lax cloud security.

There are several reasons why healthcare professionals should not trust the cloud, including the fact that they cannot cede ownership of their patients' medical records. The majority of cloud service providers keep their data in many data centers spread across the globe. -is clearly advantageous because data storage in the cloud will be redundant and several data centers will aid in disaster recovery in the event of force majeure. However, this same benefit can also provide a security risk because data kept across numerous places will be more vulnerable to loss and theft. In general, using the cloud comes with a number of security problems, including the failure to isolate virtual users, unauthorized access, privilege abuse, and inadequate encryption.

5.1.1. eHealth Cloud Advantages

The cloud offers several advantages.

1. Better patient care as a result of constant patient involvement with many healthcare stakeholders. Doctors may study and diagnose patient data at anytime, anyplace.

2. Cost savings: Expensive gear and software are not required. Savings cover both the upfront expenses of buying on-premises hardware and software as well as the costs of support and maintenance.

3. Energy savings: By eliminating the requirement for on-site data centers and the associated costly cooling, the energy expenditure will be reduced.

4. Strong disaster recovery: Nearly all cloud vendors provide redundant systems and services in case of emergency.

5. Research: The cloud serves as a central database server for monitoring epidemics, disease control, and national medical research.

6. Overcoming a lack of resources: remote doctors can conduct consultations via telemedicine.

7. Rapid deployment enables the usage of hardware and software systems virtually instantly.
8. Data accessibility: All healthcare stakeholders, including doctors, pharmacies, hospitals, and insurance firms, have access to data.

5.1.2. eHealth Cloud Limitations

eHealth cloud has a lot of restrictions:

1. *Availability and dependability:* Depending on the speed of the Internet connection, the service may be slow, interrupted, or unavailable. is going to have a big impact on user experiences.

2. *Interoperability:* To provide effective communication, integration, and interaction between the platforms of various healthcare providers, standards are required.

3. *Security and privacy:* environments that are open and shared are vulnerable to data loss and theft.

4. *Laws and regulations:* For cloud computing to be widely used, there must be laws, rules, and ethical and legal frameworks.

5. *Limited flexibility and control:* Because of centralization, it has little control over who owns the data. Generic e-cloud apps are frequently available and renting specialized software may be challenging.

6. *Attack susceptibility:* The cloud is susceptible to many security assaults.

5.2. eHealth Cloud Security Concerns

Accessing medical records from any location at any time is a key component of modern healthcare. Medical record exchange and integration are made easier by the cloud computing paradigm in healthcare. Although the cloud computing paradigm has many advantages, it also puts the privacy and security of patient data at risk. In order to increase confidence between consumers and healthcare professionals, cloud service providers should address security issues in the cloud. Cloud computing applications must meet a number of security standards in order to increase user confidence in this still-evolving technology. The crucial security and privacy requirements for cloud-based healthcare applications are outlined below along with how our proposed model can help in serving these requirements.

5.2.1. Audit

A security mechanism that guarantees the security of a healthcare system is auditing. The term "audit" refers to the process of chronologically logging all user interactions with the healthcare system, such as keeping track of each time data is accessed or modified.

Users within the healthcare provider's company must be held responsible for their conduct while handling patients' protected health information, according to both Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health (HITECH) Act. We can use our model for keeping audit controls for this data; for instance, Integrating the Healthcare Enterprise (IHE) sets a profile for the Audit Trail that has enough details to respond to queries like: "For some user: Which patient records was accessed? Who accessed the patient record for a certain patient? What failed user authentication attempts were noted? By assuring the detection of unauthorized access and unlawful disclosure of medical records, such strategies could assist administrators in reducing insider threats. Auditing can also assist administrators identify potential flaws in the system and track down hacker attempts to access a cloud-based public healthcare system.

5.2.2. Data Remanence and Freshness

An inadvertent data confidentiality assault could result from data remanence. If data freshness is not taken into account in the healthcare system, data confidentiality and integrity are insufficient. The patient's medical records must be current and up to date in order for data to be considered fresh. Data discrepancy, especially in urgent circumstances, is caused by storage delays and the transmission of out-of-date notifications. As mentioned in section 4.3, guaranteed data deletion will be carried out with the help of logs and audit trails to diminish data remanence, as well as these will also help in maintaining data in its recent form.

5.2.3. Cloud Multitenancy

Shared processing, shared memory, and shared storage were some of the main motivations behind the development of clouds. Multitenancy is a regular practice used by cloud providers to maximize resource utilization and cut costs. In order to secure data sharing and integration, security threats to data access and management are prevalent. Data about patients should be isolated to deliver secure multitenancy. Section 4.7 talks about enclaving and secure isolated environments which can be used to isolate critical information from the public cloud for better security and data confidentiality.

5.2.4. Authenticity and Non-repudiation

The accuracy of origins, attributions, promises, and intents are all examples of what is meant by authenticity. It confirms the legitimacy of the entity making the access request. The Authentication Act must be used in healthcare systems to verify the identity of the entities using the information and the information provided by healthcare providers. Information authentication can provide unique challenges, such as man-in-the-middle attacks, which are frequently avoided by using a login and password combination. To guard against man-in-the-middle attacks, the majority of cryptographic protocols incorporate some type of endpoint authentication. In a healthcare system, both consumer identities and healthcare information provided by providers should be validated at every point of access.

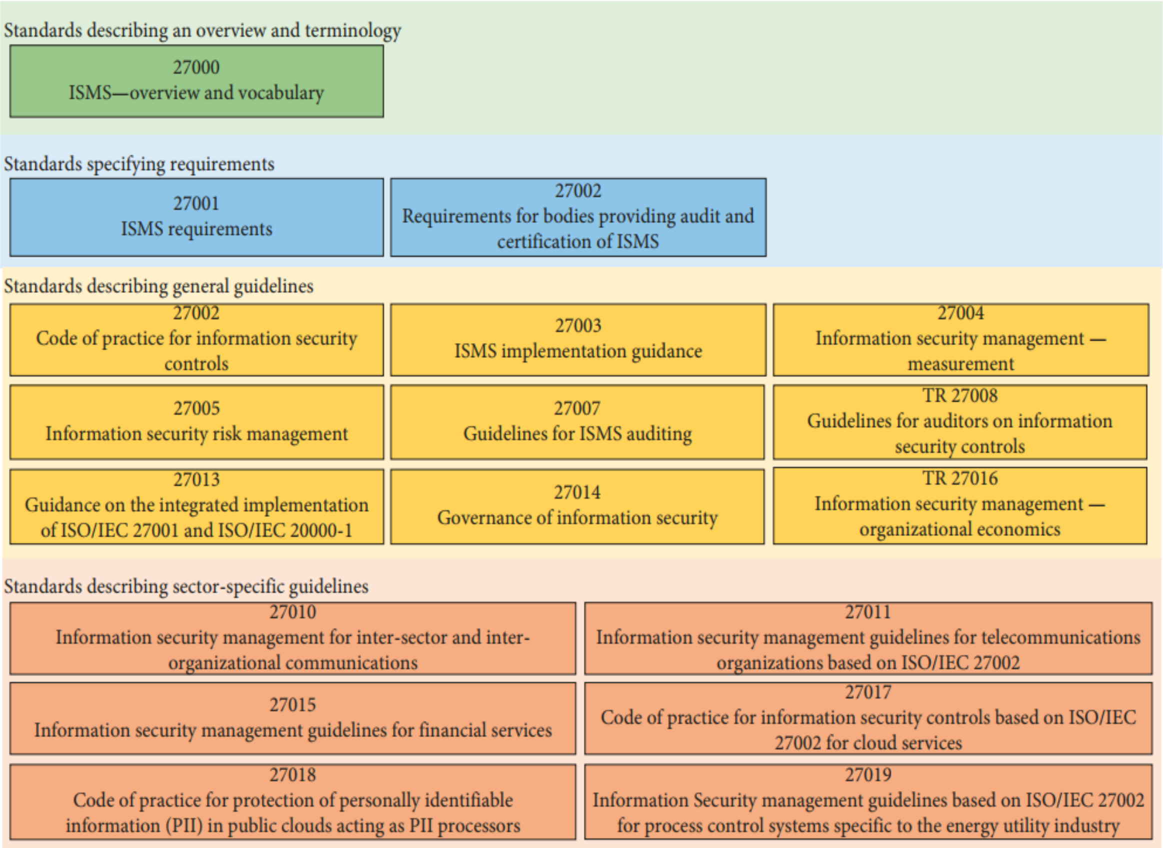
Threats of repudiation are raised by users who dispute the validity of their signatures after viewing health information. For instance, in a hospital setting, neither the patients nor the doctors can contest the validity of their signatures after stealing patient information.

Identity-based encryption can be used for both the authentication and non-repudiation issues in the eHealth cloud infrastructure (Section 4.4). We can use fuzzy private keys like an iris scan or fingerprint which are unique for a particular user and no one else can replicate it.

5.2.5. Regulatory Aspects

Standards are typically developed by professionals from organizations and scientific institutes to specify the acceptable attributes of a service or product. To indicate a consensus on characteristics like quality, security, and dependability that should be applicable for a long time, these standards have been established and published. The standards' objective is to assist people and businesses in their pursuit of goods and services. By upholding standards, cloud service providers can improve their reputation. Multiple standards were developed by various nations to ensure cloud security and privacy.

Figure 8 summarizes the ISO/IEC 27000 series that should be used for regulating the cloud service providers.



Standards describing an overview and terminology

27000 ISMS—overview and vocabulary

Standards specifying requirements

27001 ISMS requirements

27002 Requirements for bodies providing audit and certification of ISMS

Standards describing general guidelines

27002 Code of practice for information security controls

27003 ISMS implementation guidance

27004 Information security management — measurement

27005 Information security risk management

27007 Guidelines for ISMS auditing

TR 27008 Guidelines for auditors on information security controls

27013 Guidance on the integrated implementation of ISO/IEC 27001 and ISO/IEC 20000-1

27014 Governance of information security

TR 27016 Information security management — organizational economics

Standards describing sector-specific guidelines

27010 Information security management for inter-sector and inter-organizational communications

27011 Information security management guidelines for telecommunications organizations based on ISO/IEC 27002

27015 Information security management guidelines for financial services

27017 Code of practice for information security controls based on ISO/IEC 27002 for cloud services

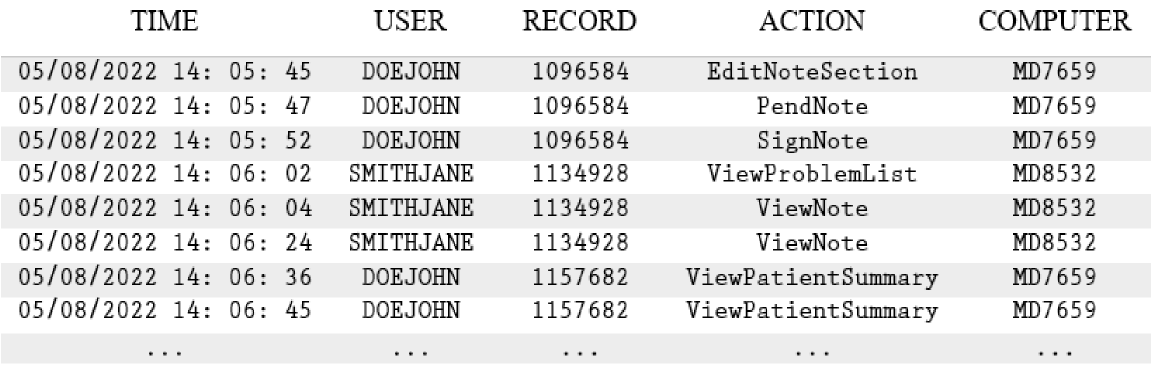
27018 Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors

27019 Information Security management guidelines based on ISO/IEC 27002 for process control systems specific to the energy utility industry

Figure 9 - ISO/IEC 27000 series standards categories

5.3. Model Performance Evaluation for eHealth Cloud

This section assesses how well the suggested mechanism works. Table 2 shows how an eHealth Record (EHR) audit log is populated. This log will be used for data traceability which in turn helps in keeping track of when and where the data was stored. In addition, Table 3 provides a list of all the variables which include the Traceability Time as well as the Auditing Time. This table makes it abundantly evident that a public verifier requires additional communication overhead to complete the auditing operation in order to achieve high detection accuracy.



TIME	USER	RECORD	ACTION	COMPUTER
05/08/2022 14: 05: 45	DOEJOHN	1096584	EditNoteSection	MD7659
05/08/2022 14: 05: 47	DOEJOHN	1096584	PendNote	MD7659
05/08/2022 14: 05: 52	DOEJOHN	1096584	SignNote	MD7659
05/08/2022 14: 06: 02	SMITHJANE	1134928	ViewProblemList	MD8532
05/08/2022 14: 06: 04	SMITHJANE	1134928	ViewNote	MD8532
05/08/2022 14: 06: 24	SMITHJANE	1134928	ViewNote	MD8532
05/08/2022 14: 06: 36	DOEJOHN	1157682	ViewPatientSummary	MD7659
05/08/2022 14: 06: 45	DOEJOHN	1157682	ViewPatientSummary	MD7659
...	...	...	...	...

Table 2 - Example of an EHR audit log

However, the data owner incurs a little communication burden in order to complete the traceability test. There is a tradeoff needed in terms of tracing the remnants of data versus the communication overhead.

S. No.	System Variables	Size
1	Data Storage	15MB
2	Data blocks	20
3	Communication Cost	10.45KB
4	Auditing time	3.24s
5	Traceability time	5.69s

Table 3 - Performance of the proposed model

To be more specific, the recommended mechanism's auditing time is shown in Figure 10 for clarity. It displays the linear relationship between auditing and traceability time, and the number of users. The data owner needs 5.69 seconds to finish the traceability task relating to the data users for a health care data set with a storage capacity of 15MB. While the auditing time is under 0.8 seconds if the data storage size is 15MB and two users are sharing the cloud data like a doctor and its relevant patient.

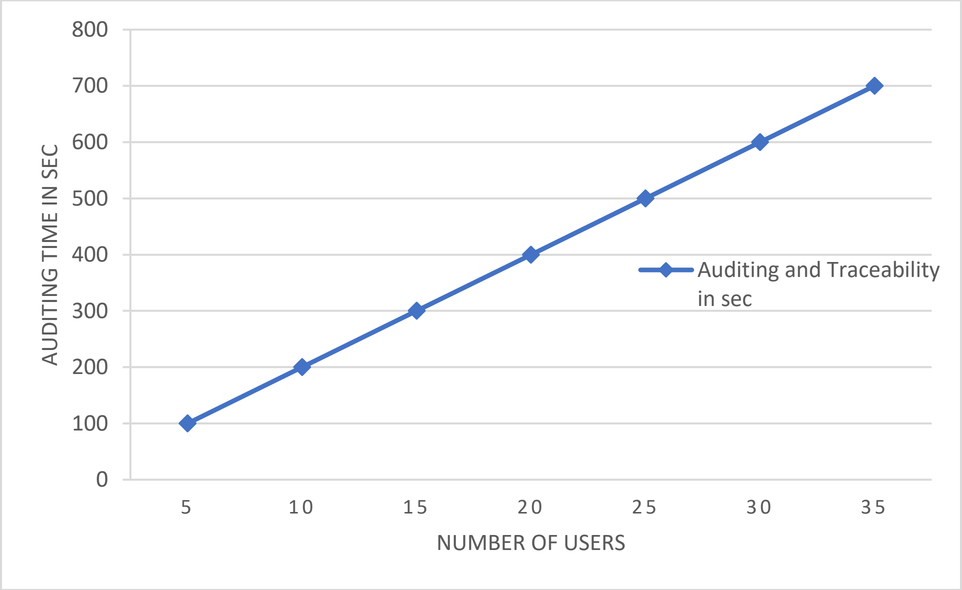


Figure 10 - Auditing and Traceability time of the proposed model

Secondly, we used Cloudsim (Beloglazov and Buyya, 2012) for a simulation where a CSP has a data center of 50 servers that are used to host 50 virtual machines. This simulation is executed for the SLA compliance between a hospital and its CSP. The VMs are owned by various patients and doctors. In our scenario, host servers were divided into two groups. In contrast to the subsequent 25 servers, the first 25 are premium servers that uphold the SLA. Table 4 shows a summary of the simulation while the variables used are based on the Cloudsim (Calheiros et al., 2011) simulation scenarios.

The simulation's goal is to demonstrate the cost savings that dishonest CSPs obtain by taking advantage of an absent auditing mechanism. This also results in the non-compliance to any delete operation or complete data erasure as required by the user. On the other hand, if there is a regular monitoring of the CSPs actions and how it carried out the user demands as mentioned in the SLAs, there has been seen a lesser amount of violation of SLAs by the CSPs. It can also be seen in Figure 11 that with using data auditing, which asks the CSPs to follow the SLAs, has decreased the amount of data remnants being present in the cloud.

Simulation Variables		
Entity	*Parameter*	*Value*
CSP	Servers	50
	Server specifications	Xeon Processor 2 GHz, Dual Core, 1 GB Ram, 1 TB storage, 1 Gbs BW
	VMs	50
	VM specifications	1 Processor Element, 2 GHz, 1 GB Ram, 100 Mbs BW
TPA	Monitoring period	24 hours
	Rate of monitoring	1 per half hour
Simulation Results		
Entity	*Parameter*	*Value*
CSP	Est. number of data remnants per batch	282
	SLA violations	11%

Table 4 - Simulation specifications and results

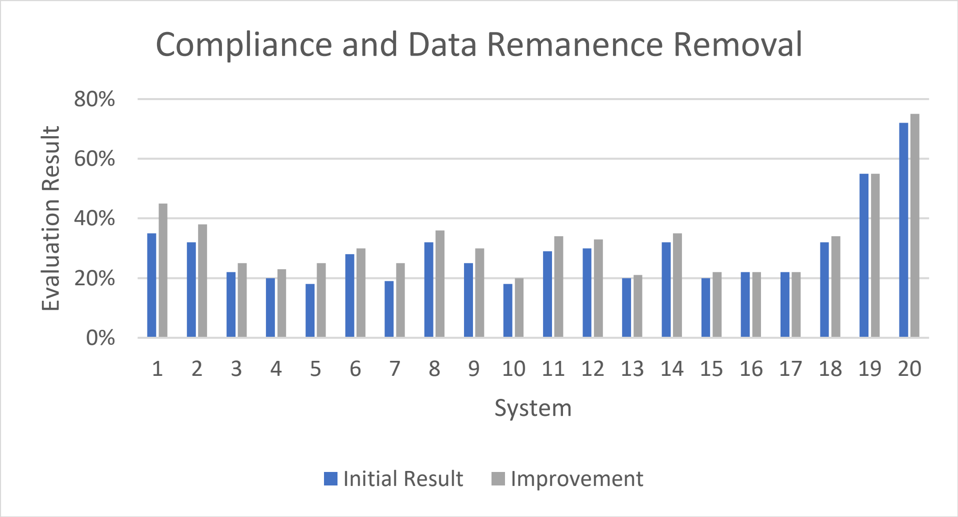


Figure 11 - Assessment results after auditing implementation

Chapter 6

6. CONCLUSION AND FUTURE WORK

This chapter provides a conclusion to the thesis write-up by summarizing all that has been discussed in the previous chapters. Future work will also be mentioned in this chapter.

6.1. Conclusion

In the present era, there is an increasing need of cloud computing infrastructure amongst home users, organizations, and enterprises. Each customer wants the cloud for several reasons which can be easily provided and managed with a minimum of interaction with the cloud service provider (CSP). However, various aspects of security concerns have risen with the ubiquity of cloud computing and amongst all the aspects, the least addressed one by the service providers is data remanence. It has been known to be prevalent in RAM, ROM, EEPROM, EPROM, flash drives, etc., which gives room for memory vulnerability to attacks. One of the many concerns are related to the complete deletion of data and making sure that the data once erased cannot be recovered through any means. The failure to properly remove data, however, could result in inadvertent disclosures and costly fines as well as reputational harm. The problems with partial deletion are widely known in non-cloud environments. To the best of our knowledge, no systematic investigation of assured deletion problems in public clouds has yet been conducted. Existing solutions like Sanitization methods including clear, purge and destroy, encryption, data remanence standards and third-party auditing. Also, data remanence remains an open issue for the data auditing needs of the clients requests as well as amongst the CSPs. Our proposed model tried to bridge the gap by presenting a data auditing model for remanence in the public cloud. There are seven aspects each of which have a separate necessity and significance that can be used in conjunction to the other for better results. These aspects are traditional sanitization methods (for end-level device destruction after the media has been rendered useless), transparency (trusted computing), guaranteed data deletion (secure data erasure), encryption (data privacy), service level agreements (user-provider contract), certification (compliance to standards) and isolated environment (private enclaves for critical data). In the end we provide eHealth case study for the evaluation of our model.

6.2. Future Work

As we have mentioned before that data remanence is a very less research and discovered aspect of data security. Our work here is very new and primary which can be extended to provide more security and safety for user data in the public cloud. The most basic aspect that can be worked upon is a realization of a standard solely dedicated to data remanence in a cloud environment. This standard should include the structured deletion and encryption methods which bind the CSPs to comply with them. Also, deletion methods could be made more advance by creating solutions to track the redundancy of data and geolocating the backup copies of them. Transparency and SLAs are an advanced topic that can be worked upon which require the necessary trust between the consumer and provider. Encryption methods could be made more lightweight so that it does not produce an overhead when being processed in the cloud.

REFERENCES

[1] Ali, Mazhar, Samee U. Khan, and Athanasios V. Vasilakos. "Security in cloud computing: Opportunities and challenges." *Information sciences* 305 (2015): 357-383.

[2] Joshi, Seema B. "Standards and techniques to remove data remanence in cloud storage." *2018 IEEE Punecon*. IEEE, 2018.

[3] Rasheed, Hassan. "Data and infrastructure security auditing in cloud computing environments." *International Journal of Information Management* 34.3 (2014): 364-368.

[4] Aissaoui, Khalid, Hicham Belhadaoui, and Mounir Rifi. "Survey on data remanence in Cloud Computing environment." *2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*. IEEE, 2017.

[5] Mather, Tim, Subra Kumaraswamy, and Shahed Latif. *Cloud security and privacy: an enterprise perspective on risks and compliance*. " O'Reilly Media, Inc.", 2009.

[6] Albelooshi, Bushra, et al. "Experimental proof: Data remanence in cloud vms." *2015 IEEE 8th International Conference on Cloud Computing*. IEEE, 2015.

[7] Karvelas, Nikolaos P., and Aggelos Kiayias. "Efficient proofs of secure erasure." *International Conference on Security and Cryptography for Networks*. Springer, Cham, 2014.

[8] Perito, Daniele, and Gene Tsudik. "Secure code update for embedded devices via proofs of secure erasure." *European Symposium on Research in Computer Security*. Springer, Berlin, Heidelberg, 2010.

[9] Peter Mell and Tim Grance. “The NIST Definition of Cloud Computing. NIST Special Publication 800-145 (Draft)”. Retrieved September 10, 2011.

[10] Alghofaili, Yara, et al. "Secure Cloud Infrastructure: A Survey on Issues, Current Solutions, and Open Challenges." *Applied Sciences* 11.19 (2021): 9005.

[11] Jayalekshmi, M. B., and S. H. Krishnaveni. "A study of data storage security issues in cloud computing." *Indian Journal of Science and Technology* 8.24 (2015): 1 – 5.

[12] Pearson, Siani. "Privacy, security and trust in cloud computing." *Privacy and security for cloud computing*. Springer, London, 2013. 3-42.

[13] Sarkar, Mrinal Kanti, and Sanjay Kumar. "A survey on data storage security issues in cloud computing." *Int. J. Appl. Eng. Res* 13.10 (2018): 8390-8406.

[14] Modi, Chirag, et al. "A survey on security issues and solutions at different layers of Cloud computing." *The journal of supercomputing* 63.2 (2013): 561-592.

[15] Skorobogatov, Sergei. "Data remanence in flash memory devices." *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, Berlin, Heidelberg, 2005.

[16] Gallagher, Patrick R. "A guide to understanding data remanence in automated information systems." (1991).

[17] Gutmann, Peter. "Secure deletion of data from magnetic and solid-state memory." *Proceedings of the Sixth USENIX Security Symposium, San Jose, CA*. Vol. 14. 1996.

[18] Wei, Michael, et al. "Reliably Erasing Data from {Flash-Based} Solid State Drives." *9th USENIX Conference on File and Storage Technologies (FAST 11)*. 2011.

[19] Halderman, J. Alex, et al. "Lest we remember: cold-boot attacks on encryption keys." *Communications of the ACM* 52.5 (2009): 91-98.

[20] Fera, M. Arun, and M. Saravana Priya. "A Survey on Trusted Platform Module for Data Remanence in Cloud." *Proceedings of the International Conference on Soft Computing Systems*. Springer, New Delhi, 2016.

[21] Siemons, F., 2016. *Data sanitization for cloud storage*. [online] Infosec Resources. Available at: <https://resources.infosecinstitute.com/topic/data-sanitization-for-cloud-storage/#gref> [Accessed 2 July 2022].

[22] Regenscheid, Andrew R., Larry Feldman, and Gregory A. Witte. "NIST Special Publication 800-88, Revision 1: Guidelines for Media Sanitization." (2015).

[23] Gentry, Craig. "Fully homomorphic encryption using ideal lattices." *Proceedings of the forty-first annual ACM symposium on Theory of computing*. 2009.

[24] Darwazeh, Nour S., Raad S. Al-Qassas, and Fahd AlDosari. "A secure cloud computing model based on data classification." *Procedia Computer Science* 52 (2015): 1153-1158.

[25] U. S. Department of the Navy Staff, NAVSO P-5239-08, Network Security Officer Guidebook, USA, 1996.

[26] U. S. Army, Army Regulation 380-19, Information Systems Security, USA, 1998.

[27] U. S. Department of Defense Staff, 5220.22-M National Industrial Security Program Operating Manual, USA, 1995.

[28] Royal Canadian Mounted Police, “Royal Canadian Mounted Police Apology,” R. Can. Mounted Police, 2009.

[29] Australian Government Staff, Department of Defense, Australian Government Information Security Manual Controls, Australia, 2014.

[30] Secretary of Air Force, Air Force Manual 17-1301, Computer Security (COMPUSEC), 2017.

[31] The Government Communications Security Bureau, NZISM New Zealand Information Security Manual, version 2.7, Government Communication Security Bureau, Wellington, 2017.

[32] Information Commissioner's Office Staff, Guide to the General Data Protection Regulation, Data Protection Act-2018, pp. 1–240, United Kingdom, 2018.

[33] Basu, Srijita, et al. "Cloud computing security challenges & solutions - A survey." *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2018.

[34] Crosby, Scott A., and Dan S. Wallach. "Efficient data structures for tamper-evident logging." *USENIX security symposium*. 2009.

[35] Shen, Wenting, et al. "Data integrity auditing without private key storage for secure cloud storage." *IEEE Transactions on Cloud Computing* 9.4 (2019): 1408-1421.

[36] ISO/IEC 27000-series: http://www.27000.org/

[37] ISO/IEC 27017, *Code of Practice for Information Security Controls Based on ISO/IEC 27002 for Cloud Services* http://www.iso.org/iso/catalogue_detail?csnumber=43757

[38] ISO/IEC 27018 (2014). *Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors* http://www.iso.org/iso/catalogue_detail.htm?csnumber=61498

[39] ISO/IEC 29100 (2011). *Privacy Framework* http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45123

[40] Al-Issa, Yazan, Mohammad Ashraf Ottom, and Ahmed Tamrawi. "eHealth cloud security challenges: a survey." *Journal of healthcare engineering* 2019 (2019).

[41] Calheiros, Rodrigo N., et al. "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms." *Software: Practice and experience* 41.1 (2011): 23-50.

[42] Beloglazov, Anton, and Rajkumar Buyya. "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers." *Concurrency and Computation: Practice and Experience* 24.13 (2012): 1397-1420.