IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 25, Issue 1, Ser. I (Jan. – Feb. 2023), PP 17-28
www.iosrjournals.org

A Review on Securing Distributed Big Data Storage in Cloud Environment

Anah Hassan Bijik, Souley Boukari,
Abdulsalam Yau Gital and Mohammed Abdulhamid
Department of Mathematical Sciences
Abubakar Tafawa Balewa University , Bauchi
Bauchi State

Cloud computing security is a serious concern. There is a considerable requirement for infrastructure security at the network, host, application, and data levels. Data is related with each level, such as network, host, and application. This paper provides an overview and research on cloud computing, huge data storage, and various methods for cloud data security.

Keywords: *Big data, storage, cloud computing, security*

Date of Submission: 10-01-2023 Date of Acceptance: 26-01-2023

I. Introduction

Cloud computing is a vast and multifaceted phenomenon. Many characteristics of cloud computing may be traced back to the 1950s, when colleges and businesses rented out mainframe computer calculation time. At the time, renting was one of the only methods to gain access to computing resources because computing technology was too large and expensive for individuals to own or control. By the 1960s, computer scientists such as John McCarthy of Stanford University and J.C.R Licklider of the United States Department of Defense Advanced Research Projects Agency (ARPA) were proposing ideas that foreshadowed some of the major features of cloud computing today, such as the concept of computing as a public utility and the possibility of a network of computers that would allow people to access data and programs from anywhere in the world [1,2].

The main advantage of cloud computing is that it eliminates the need for users to be physically present in the same location where hardware, software, and storage space are physically present. The impact of cloud computing on industry and end users cannot be overstated: many aspects of daily life have been transformed by the omnipresence of software that runs on cloud networks. Start-ups and organizations can save costs and expand their offerings by embracing cloud computing instead of purchasing and managing their own hardware and software. Independent developers now have the ability to establish internationally accessible apps and online businesses and researchers may now share and analyze data on previously unimaginable scales [3,4].

Furthermore, internet users can easily access software and storage in order to produce, share, and store digital media in numbers that much exceed the computing power of their personal devices. Despite the growing popularity of cloud computing, many people are unaware of its specifics. What is the cloud, how does it work, and what are the advantages for corporations, developers, researchers, governments, healthcare practitioners, and students? The cloud allows you to save and access your data from anywhere, at any time, without having to worry about maintaining hardware, software, or storage space. There are numerous advantages of use Cloud computing to render or access computing resources. Many individuals today use Cloud computing without even knowing what it is. Users of Gmail, Yahoo mail, YouTube, and Skype, for example, are all in the Cloud. Companies and organizations are becoming more aware of the numerous advantages that Cloud computing offers [5]. All of these services are given at a cheap cost to the consumer. The user must pay based on the amount of storage space he uses. Everyone is transferring his data to the cloud as a result of this flexibility, and users are permitted to store vast amounts of data on cloud storage for future use [2]. Many cloud customers are concerned about their sensitive data being accessed by cloud operators [3,4].

Cloud computing security is a serious concern. There is a considerable requirement for infrastructure security at the network, host, application, and data levels. Data is related with each level, such as network, host, and application. Security concerns about many types of attacks involving various technologies must be addressed. Cloud monitoring, on the other hand, is utilized in a variety of contexts, including performance, SLA management, security, billing, and troubleshooting [6,7]. Customers will be able to choose specific CSPs based on their security requirements. Security must strike a balance between safety, usability, and simplicity. Security is everyone's responsibility, including cloud consumers and CSPs. Cloud computing security is a serious concern. There is a tremendous push to ensure infrastructure security at the network, host, application, and data levels. Data is related with each level, such as network, host, and application. Some cloud computing security challenges include [8-11]:

i *Availability* - data availability is an essential security issue. It must be made available to the user whenever it is necessary. Furthermore, the user must have control over its data. When service from another cloud service provider is requested, an availability issue must be addressed. Currently, there are three significant concerns to availability. The first danger is a network-based attack, the second is cloud service provider availability, and the third is cloud service provider backup of stored data. As a result, effective and efficient mechanisms for access control, authentication, and authorisation of sensitive data are required.

ii *Data remanence* - This is a problem when data is exposed to an unauthorized person after deletion. A data security lifecycle refers to the complete process of creating and destroying data. When destroying data, extreme caution must be exercised.

iii Third-Party Control- The user data is managed by the Cloud Service Provider. Third-party access may result in the disclosure of sensitive information and trade secrets. Corporate espionage is also a major concern. It should not also force users to rely solely on a single cloud service provider.

iv *Legal Concerns and Privacy*- The user is ignorant of where their data is stored on the cloud. Each country's cyber laws are unique. There is much worry regarding legality and data confidentiality. The user is likewise concerned about the privacy of his or her data.

Major challenges affecting cloud computing must be addressed, including data security, privacy, confidentiality, integrity, and authentication. Most cloud service providers maintain data in plaintext, and users must employ their own encryption technique to secure their data if necessary. When hostile acts are carried out in cyberspace, whether internally or externally, such activities and intruders can be prohibited [7]. When the data is to be processed, it must be decrypted [2]. Cryptography encompasses techniques such as microdots, word-image fusion, and other methods for concealing information in storage or transit. In today's computer-centric society, however, cryptography is most typically connected with scrambling plain text into a process known as encryption, then back again. Modern cryptography in computer science is concerned with numerous data security issues.

In general, when data is encrypted, it is not easily understood by unauthorized persons, and decryption is employed to recover the plain text. Decryption is required before doing any type of computation. Encryption overcomes major problems, but the power of the cloud can only be used if the user is able to do computation on encrypted data [10,12]. Furthermore, it is desired to protect data on cloud servers through the use of encryption-oriented technologies. Because of the larger key size and longer computational time of asymmetric cryptography, public key cryptography is used only for key exchange and symmetric key cryptography is used for further encryption/decryption. The computational time of cryptography techniques is further classified as encryption/decryption time, key generation time, and key exchange time. The time required for encryption/decryption is calculated by converting plaintext (message) to ciphertext and vice versa [12,13]. Many studies presented various cryptographic solutions for cloud data safety, including Fully Homomorphic Encryption (FHE) and Attributed-Based Encryption (ABE). This form of secure system can successfully protect data from target attackers, such as external hostile activities and internal inappropriate operations; nevertheless, the efficiency of data processing may suffer due to the additional computations necessary in huge data [14,15]. Some processes are simply impossible to complete due to technical constraints, such as sounds in FHE]. These are currently two approaches to solving security challenges on the cloud side: Using regulatory compliance measures to limit employee behavior and prevent data leakage with encryptions such as FHE and ABE. This paper presents an overview of an encryption approach based on formal methods for securing distributed big data storage in the cloud

II. Big Data

As society becomes increasingly instrumented, companies are producing and storing massive volumes of data. Managing, getting insights and safeguarding such massive data is a challenge and vital to competitive advantage. Analytics solutions that mine structured and unstructured data are significant since they can assist firms discover insights not just from their privately acquired data but also from enormous amounts of data publicly available. The ability to cross-relate private information on consumer preferences and predicts with information from tweets, blogs, product evaluation and data from social network open a wide range of possibilities for organizations to understand the needs of their customers, predicts the needs and demands and optimize the use of resources. Despite its gargantuan nature, big data applications are almost widespread, ranging from marketing to scientific research to customer interest, among other things. Today, we can see Big Data in action practically everywhere. From Facebook, which processes over 40 billion images from its user base, to CERN's Large Hydron Collider (LHC), which generates 15.1 billion per year, to Walmart, which processes over 1 billion consumer transactions every hour [16].

Simply described, big data is a collection of enormous datasets that cannot be analyzed using conventional computing approaches. Big data has evolved into a comprehensive field that includes a variety of tools, methodologies, and frameworks. [16, 17] define big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze." These definitions demonstrate that the writers consider large data in terms of how it is analyzed rather than how many exact terabytes of space it occupies. Other meanings are more focused on the data. [18], opined that for data to be classified as big data it must possess the three Vs: Volume, Variety, and Velocity. However, big data is a massive collection of data collected over a short period of time that is complex and difficult to process and manage using traditional database management tools. Its format can be classified as structured, which is derived from various research efforts and other traditional databases, semi-structured, which is derived from extensible markup language (XML), and finally unstructured, which is derived from social media discussions, emails, environmental data generation, and other sources [19].

2.1 Characteristics of Big Data

Big Data can be defined in three words: volume, velocity, and variety [20]. Each attribute is described below.

Volume: Volume concerns the large quantities of data that is generated continuously or simply put 'amount of data'. Because of its enormous volume, storing such data used to attract high storage cost. One might ask where the huge volume of data is being generated from, it simply comes from e-commerce, social networking websites and our devices like the smart phones. This data can easily be distinguished as structured (formal schemas and data models), unstructured (no pre-defined data models) and semi-structured data (lacks strict data model structure) and Mixed (various types together) [20].

Velocity: Velocity refers to the rate at which data is created and processed. Previously, data was processed in batches, therefore the coming rate must be slower than the batch processing rate. However, at the moment, the rate at which such massive volumes of data are generated is extremely fast. According to internet live statistics, Google now processes over 1.2 trillion queries each year globally. Others, such as Facebook, create over 2.7 billion activities and 300 million photographs per day [20, 21].

Variety: Data formats, which include databases, spreadsheets, documents, images, and videos, to name a few, are categorised as structured, unstructured, or semi-structured. Previously, data structure was critical in data processing, but it can no longer be enforced before the data is used or analysed. It has been observed that appropriate data for scientific and business reasons has been made public over the years. Examples include historical data on traffic conditions in major cities, product ratings and comments, and demographics [20].

Veracity: The amount of data that may be trusted given the dependability of its source is referred to as its veracity. How reliable is all of this information? Consider all of the Twitter posts that contain hash tags, abbreviations, typos, and so on, as well as the dependability and correctness of all of that stuff. Gleaning massive amounts of data is pointless if the quality or trustworthiness is lacking. Another excellent example is the usage of GPS data. As you travel through a city, the GPS will frequently "drift" off track. Signals from satellites are lost when they bounce off tall buildings or other structures. When this occurs, location data must be combined with other data sources, such as road data or accelerometer data, to offer correct results [22].

Value: The monetary value that a firm can gain from using Big Data computing is referred to as value. Having an infinite amount of data is one thing, but it is pointless unless it can be converted into value. While there is an obvious link between data and insights, this does not automatically imply that Big Data is valuable. The most crucial aspect of starting a big data endeavour is understanding the costs and advantages of gathering and analysing data to ensure that the data gathered can eventually be monetized. Looking at the characteristics and classification of big data, one can conclude that the paradigm shift from static structured data management to the "Big Data" trend is mostly due to technical development dynamics and, to a lesser extent, Cloud Computing [22].

III. Cloud Computing

Various authors have approached Cloud Computing from various angles [23]. Cloud computing is a computing approach in which enormously scalable IT-related capabilities are delivered as a service to various external consumers via the internet. While [24] defines cloud computing as a parallel and distributed system comprised of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements negotiated between the service provider and the consumers. The National Institute of Standards and Technology (NIST) provides the most widely accepted definition of cloud computing: "Cloud computing is a network access model that enables 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 interaction from service providers. This cloud model is made up of five key elements, three service models, and four deployment models " [25].

3.1 Characteristics Cloud Computing

According to the NIST standard standards, cloud computing has five basic features, while there are more characteristics based on other sources. The following are the characteristics [25]:

On- demand self Service: This is a situation in which a user can request one or more services as needed without having to engage with humans, which means it can be accomplished using an online control panel and paid for using the "pay-and-go" technique.

Resource Pooling: Resource pooling provides a collection of resources that replicate the behavior of a single blended resource; in other words, the user does not know and does not need to know the location of the resources being pooled. Storage, computation, memory, network bandwidth, and virtual machines are examples of resources [25, 26].

Broad Network Access: This is more of a case of resource and service availability, because all of the resources and services provided that are located in different vendor areas in the cloud can be available and accessible from a wide range of locations via standard mechanisms (e.g., mobile phones, laptops, PDAs), for example, the terms "easy-to-access standardized mechanisms" and "global reach capability" are used to refer to this characteristic [25, 27,28].

Rapid Elasticity: Elasticity is fundamentally synonymous with scalability; elasticity denotes the ability to scale up or scale down resources as needed. Users can request as many services and resources as they require at any time. This quality is so admirable that Amazon, a well-known cloud service provider, dubbed one of its most popular and widely used services the Elastic Compute Cloud (EC2) [25].

Measured Service: All components of the cloud should be automatically regulated, monitored, optimized, and reported at multiple abstract levels for the resources of both vendors and users in measured service [25].

Multi-Tenacity: The cloud security alliance proposed this attribute. It was suggested that models for policy-driven enforcement, isolation, governance, segmentation, service levels, and chargeback/billing for various customer groups be developed [25].

Auditability and Certifiability: Logs and audit trails should be included as part of services; this will go a long way toward assessing how well policies and regulations are followed.

3.2 Benefits of Cloud Computing

Cloud computing provides numerous advantages, including the following [29]:

Lower prices: All resources are shared, including pricey networking equipment, servers, and IT workers, leading in lower expenses, particularly for small to mid-sized applications.

Shifting Capital Expenses to Operating Expenses: Cloud computing allows businesses to shift money from capital expenses to operating expenses, allowing them to focus their money and resources on innovation.

Agility (On Demand Service): On-demand provisioning allows for speedier setup on an as-needed basis. When a project is funded, the customer can begin servicing, and when the project is completed, the customer can simply end the cloud contract.

Scalability: With a more cost-effective pay-as-you-go model, many cloud services can easily and efficiently scale to accommodate the growing nature of the organization. This is also referred to as elasticity.

Simplified Maintenance: Patches and upgrades are quickly deployed throughout the common infrastructure, as well as backups.

Diverse platform support: Many cloud computing services include built-in support for a wide range of client platforms, including as browsers, mobile devices, and others. This extensive platform compatibility allows programs to reach a broader range of people.

Faster development: Cloud computing platforms supply many of the essential services that would normally be implemented in-house under traditional development methodologies. These services, together with templates and other tools, can drastically shorten the development time.

Large scale prototyping/Testing: Cloud computing greatly simplifies large-scale prototyping and load testing. A client can easily launch 1,000 cloud servers to load test your application and then release them as soon as they are finished, and then try doing the same with owned or corporate servers.

3.3 Cloud Security Threats

According to a 2013 study conducted by the Cloud Computing Alliance, cloud computing faces several security threats, including traffic hijacking, insecure interfaces and APIs, denial of service (DoS) attacks, malicious insiders, cloud service abuse, insufficient due diligence, shared technology vulnerabilities, data breaches, an unknown risk profile, and a broken perimeter security model [25].

3.4. Cloud Computing Deployment Models (Types)

According to NIST's official definition of cloud computing, there are four types of cloud computing deployment models based on the interaction between the service provider and the consumer who uses the service. There are four types of deployment models: public, private, communal, and hybrid [25].

Public Cloud: This is the most often used deployment architecture; under this technique, the cloud is managed by independent providers and is open to the broader public. This implies that the cloud owner delivers public services on the internet in the great majority of cases based on present rules, policies, and pricing models. The cloud infrastructure is available for general public use. It could be owned, controlled, and operated by a business, academic institution, or government agency, or some mix of these. It exists on the cloud provider's premises.

Private Cloud: Private clouds are an option for businesses who already possess data centres and built IT infrastructure and have specific security or performance requirements. In many ways, they are a better solution for the company data centre than legacy servers, providing several benefits from virtualization and automation. However, they also provide challenges and disadvantages, most notably the necessity for the organization to relocate or re-factor programs in order to benefit from Cloud automation. In other words, the cloud infrastructure is set up for a single corporation with several users to use exclusively (e.g., business units). It could be owned, managed, and operated by the organization, a third party, or a mix of the two, and it could be on or off premises.

Community Cloud: A community cloud deployment model is utilized when a group of organizations with similar policies and concerns establish a cloud for their community members to use. These people of the community might also be referred to as customers. In this model, either a third party or the community itself provides the necessary infrastructure for cloud computing services. The majority of the expenses are shared among community members, cutting costs. This deployment technique offers the benefits of low cost and excellent security [30].

Hybrid Cloud: The hybrid cloud deployment methodology combines two or more types of clouds (private, community, or public). In this approach, infrastructures may rarely retain their unique qualities, but may require standardized functions to communicate with one another in terms of application and data interoperability and portability. In order to meet business demands, a company may, for example, bridge its internally controlled private cloud with other public clouds using standardized or proprietary technologies [25]. It is vital to remember that each service and deployment type is better suited to some business models than others. Private clouds will benefit large enterprises, while public clouds will assist smaller businesses. Businesses will continue to migrate back and forth between these four primary paradigms as cloud computing evolves.

3.5 Cloud Computing Service Models

Cloud computing, as defined by the National Institute of Standards and Technology (NIST), is a model for providing 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. First, we define Service, Platform, and Infrastructure, and then we go over the Service models in depth. A service is a mechanism that can deliver one or more capabilities that can be used in accordance with provider-defined limits and regulations and via an interface. A platform is a core computer system that contains hardware, operating systems, and, in some situations, the ability to deploy and run programs. While infrastructure refers to the underlying physical components required for a system to accomplish its functions [31]. These components in information systems can include processing, storage, network equipment, and, in some situations, database management systems and operating systems. There are three deployment models in this cloud computing service model [32].

***Software as a Service (SaaS)**:* The consumer is given the option to use the provider's applications that are hosted on a cloud infrastructure. Through a program interface or a thin client interface like a web browser (for example, web-based email), the programs can be accessed from a variety of client devices. With the possible exception of a small number of user-specific application configuration choices, the customer does not manage or control the underlying cloud infrastructure, which includes the network, servers, operating systems, storage, or even specific application capabilities.

***Platform as a Service (PaaS)**:* The ability to deploy consumer-created or acquired applications made using programming languages, libraries, services, and tools supported by the provider on the cloud infrastructure is a feature offered to the customer. The consumer has control over the deployed apps and possibly the configuration options for the application-hosting environment but does not manage or control the underlying cloud infrastructure, including the network, servers, operating systems, or storage.

***Infrastructure as a Service (IaaS)**:* The ability to deploy and operate arbitrary software, including operating systems and applications, is supplied to the consumer in the form of provision of processing, storage, networks, and other basic computer resources. Although the consumer has no management or control over the underlying cloud infrastructure, they do have some limited influence over some networking components, operating systems, storage, and deployed applications (e.g., host firewalls) [32].

IV. Cloud Storage

Cloud storage is a networked online storage model in which data is stored on multiple virtual servers, which are typically hosted by third parties, rather than dedicated servers. Hosting businesses run big data centres, and people that need their data hosted buy or lease storage capacity from them to meet their needs. In the background, data centre operators virtualize resources based on the needs of the client and expose them as storage pools, which customers can utilize to store files or data objects [33].

4.1 Characteristics of Cloud Storage

Cloud storage features that are related to cloud computing characteristics are listed and described below:
On Demand Self Service: As the name implies, consumers can do all computing functions on their own without requiring human connection with the service provider.
Broad Network Access: Cloud services can be accessed via the network using normal techniques. ***Resource Pooling:*** In a multi-tenant architecture, cloud computing resources are pooled and shared by all service users.
Rapid Elasticity: Depending on the demands of the user, their capabilities scale up and down rapidly and elastically. Metering capabilities are used to optimize resource utilization in measured services [34].

4.2 Cloud Storage Architecture

The cloud computing paradigm change has resulted in the introduction of new programs designed to run on smart phones, tablets, personal computers, and other devices. The majority of these programs have back ends that are stored in the cloud and may be accessed via web-enabled interfaces. Various writers have proposed various cloud storage architectures. Figure 1 depicts a generic architecture of cloud storage [35].

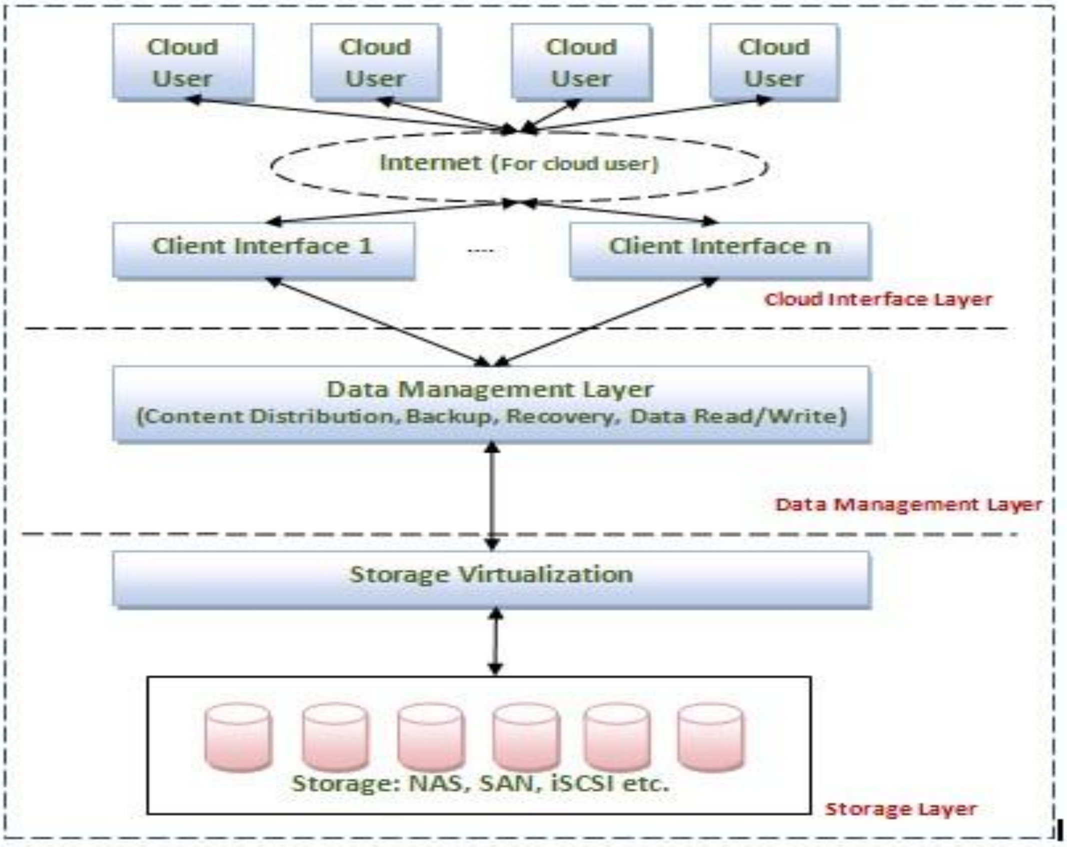

Figure 1: Generalized Architecture of Cloud Storage [35].

The cloud interface layer, the data management layer, and the storage layer comprise the architecture. The cloud interface layer is the first software layer offered by the cloud storage provider to link cloud users to storage via the internet; it is where user authentication and permission takes place. The second layer is a software layer known as the data management layer, which is responsible for data manipulation, data partitioning, synchronization, replication, data control across network, backups, recoveries, and meta data maintenance. The storage layer, which includes virtualization and basic storage, is the final layer of interest. Storage virtualization creates the appearance of unified storage. It maps scattered heterogeneous storage devices to a single continuous storage space and generates a shared dynamic platform, whereas the basic storage consists of database servers and various storage devices [35].

4.3 Cloud Storage Models

There are cloud storage options that allow customers to retain control over their data. Public cloud storage, private cloud storage, and hybrid cloud storage are the three types of cloud storage. Storage infrastructure is presented as a leasable commodity to which clients are expected to subscribe by public cloud storage providers. It could be for lengthy or short periods of time. In addition, the networking bandwidth required within the infrastructure is included. Private clouds, on the other hand, utilise public cloud storage concepts but in a form that can be securely incorporated within a user's firewall. Finally, hybrid cloud storage allows the two models (public and private) to blend, allowing policies to determine which data must be kept privately and which can be kept secure on the public cloud. Amazon is one example of a public cloud storage provider (which offers storage as a service). IBM is one of the private cloud storage providers [36]. Finally, Egnyte is one of the hybrid cloud providers. as seen in Figure 2:

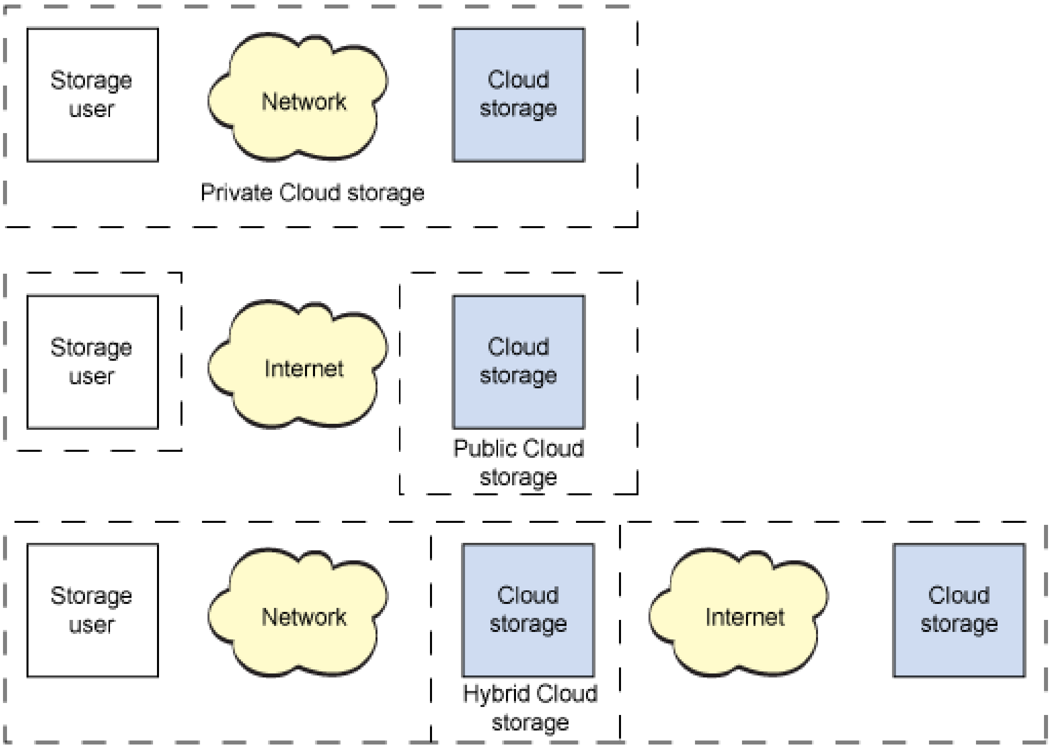

Figure 2: The cloud Storage models [36].

4.4 Benefits of Cloud Storage

Cost: Purchasing of physical storage might be costly. Without the requirement for hardware, cloud storage is far less expensive per GB than external drives.

Accessibility: Storing files in the cloud allows one to access them from any location with an internet connection.

Recovery: One can access his/her files on the cloud in the event of a hard drive failure or other hardware problem. It serves as a backup solution for his/her physical storage drives.

Syncing and Updating: When one works with cloud storage, any changes he/she makes to a file are synchronized and updated across all of his/her devices that access the cloud.

Security: Cloud storage providers provide extra layers of protection to their services. Because there are so many people who have information stored in the cloud, many companies go to great measures to ensure that your files are not viewed by someone who should not [37].

4.5 Cloud Data Storage Issues

Cloud computing relocates application software and databases to massive data centres, where data and service management may not be completely trustworthy. Cloud storage originated as an infrastructure in cloud computing, and it has been hailed for its low cost, which indicates less capital investment and the ability to scale up or down as needed. However, there will be difficulties. As a result, the authors discussed the following topics: trust management, security providers, privacy protection, ownership, data location and relocation, data recovery, performance and availability, data backup, and data portability and conversion [38].

Trust Management: Trust management is defined as relying on a person or thing's integrity, strength, competence, and assurance. Clients are reluctant to commit their data to Internet Service Providers (ISP).

Security Provider (ISP): Cloud service providers use data storage and transmission encryption, as well as user identification and authorisation. Many clients are concerned about the security of remote data in the hands of thieves and hackers. Cloud providers are well aware of the problem and are deploying significant resources to address it.

Privacy protection: Cloud computing makes use of virtual computing technology, in which users' data is distributed over multiple virtual data centres, as opposed to traditional systems where data is located in the same physical location. Because data centres are sometimes located across continents, the issue of data privacy may present a difficulty to many legal regimes [38, 39].

Ownership: Another problem with cloud data storage Some Cloud clients are concerned about losing their rights once their data has been transferred to the cloud. Although this problem can be solved with well-trained user-sided agreements.

Data Location and Relocation: Cloud computing allows for significant levels of data mobility. Some clients are concerned since consumers do not always know where their data is located [38-40].

Multiplatform Support: For IT departments adopting managed services, the more important issue is how the cloud-based service interacts across numerous platforms and operating systems, such as OS X, Windows, Linux, and thin clients. As more user interfaces become web-based, the need for multiplatform support will diminish [38].

Data Integrity: When it comes to data security, cloud service providers should include systems to assure data integrity and be able to explain what happened to a certain dataset and when it happened.

Data Recovery and Backup: An incident such as a server failure may result in data destruction or loss for users. To mitigate the challenge, data should be backed up and recovered in the future. Users of the cloud can keep a backup of crucial data on a local computer. Many service providers now offer data dumps onto media or allow users to back up data via frequent downloads.

Performance and Availability: Businesses are concerned about the acceptable standards of performance and availability of cloud-hosted applications.

Data Portability and Conversion: Because switching service providers is a problem of cloud data storage, porting and converting data is heavily dependent on the nature of the cloud provider's data retrieval format, especially when the format is not easily revealed.

Data Control: Every Data Administrator wishes to have control over their data. Because the data is located outside of the enterprise's infrastructure, there is a sense of "loss of control" over the data, which may eventually result in data loss. This is because cloud services and service delivery standards are still in their infancy, rendering these worries hypothetical and psychological rather than real.

Control and Interoperability: In computer jargon, interoperability is the capacity of computer systems or software to interchange and use information. Rajan et al (2012) underlined the fact that using cloud storage is difficult and "is not plug-and-play" . This is due to the fact that each vendor has distinct access methods and non-standard Application Programming Interfaces (API), making integrating apps like archiving or sharing files with cloud storage complicated and costly. Despite the fact that some vendors offer software clients that implement popular network file sharing protocols such as Network File System (NFS), they are unable to bridge between different cloud services. Furthermore, the lack of standard protocols for accessing cloud storage implies that there is no interoperability between cloud storage providers, complicating data movement [29].

Suitability and Applications: Cloud storage is ideally suited for more static data in active data, such as applications that involve online backup and archiving. In practice, archived files function well with cloud storage because they do not change frequently and may not require high-speed transactional access. Second, employing data reduction methods, mass data can be simply compressed. Applications that require minimal I/O performance and can tolerate low downtime are good candidates for cloud storage. Audio, video, user files, and email repositories are examples of such applications that best fit into cloud storage [29].

Security and Performance: Cloud data access is plainly constrained by network speed and latency, and despite advancements in internet performance, it remains inferior to local network storage. Although some suppliers use local caching and compression techniques to increase performance, these do not reduce internet latency. Data security is the most serious concern with cloud storage in terms of cloud security. If there is a security breach, whether in transfer or within a shared infrastructure, experts recommend that all data stored in a cloud be encrypted, however this is easier said than done depending on the application [29].

Other concerns are as follows: [37]:

Internet Access: Having an internet connection is required for cloud storage. Accessing your storage may be difficult if you are on a sluggish network. You will be unable to view your files if you are not connected to the internet.

Costs: Uploading and downloading files from the cloud incur additional charges. If you often access a large number of files, this can soon build up.
Hard Drives: Cloud storage is meant to reduce our reliance on hard drives. Some enterprise cloud storage providers, on the other hand, demand physical hard drives as well.
Support: Cloud storage support is not great, especially if one is utilizing a free version of a cloud service. Many service providers will direct him/her to a knowledge base or frequently asked questions (FAQs).
Privacy: When one utilizes a cloud service, one's data is no longer physically stored on its computer. So, who is accountable for ensuring data security? That is a grey region that is still being explored.

4.6 Securing Data in Cloud

Moving data to a Cloud environment presents an opportunity to achieve tremendous cost savings compared to the cost to purchase an equivalent amount of data for a locally hosted data centre. As with virtual machines, a customer's data is stored over a shared infrastructure that may be distributed throughout multiple Cloud datacentres. Adequate security measures must be in place to ensure unauthorized users cannot access data either intentionally or accidently [41].

V. Review of Methods of Securing Data on Cloud Environment

Security concerns have permeated almost every layer of cloud computing, from networks to system management. Because of the linkages between technical applications, such as the use of Virtual Machines, many security vulnerabilities in networks and data storage are also applicable to cloud computing (VM). Many academics investigated security issues and solutions from a variety of viewpoints, including Data Management, Monitoring and Protecting Cloud Operators, and Information Management [42- 44].
Data Management: First, data management security is an aspect of securing data in cloud computing that frequently focuses on encryption preparations or data classifications for security purposes [45]. Some ways, such as employing extensible Access Control Markup Language (XACML) management policy, have been developed to enable safe query processing for Resource Description Framework (RDF) [46]. Furthermore, selective data encryption is regarded as a method of lowering computer costs while securing data in clouds. For example, classifying data into different ranks using searchable encryption allows users to choose whether or not the data needs to be encrypted. Most existing data management methods, however, assume that cloud operators do not abuse the data or have limited access to it. In other cases, recovering information is possible even if the data is encrypted on the cloud side [47- 49].
Monitoring and Protecting Cloud Operators: Another facet in safeguarding cloud data is monitoring and protecting data storage, which takes into account data processing or processes that occur in the clouds. It suggests that the cloud operators' actions are being scrutinized. One method is to use Attributed-Based Encryption (ABE) to secure private information when data is exchanged across several clouds. However, limiting cloud operators' access scale might lead to other issues, such as data integration and data integrity. If cloud service providers are completely prohibited, the risks of data loss or operation failure would increase [17,48]. Others explored incorporating the concept of local synchronization into asynchronous spiking neural P systems. This method was developed to improve the computational capability of distributed parallel computing systems [50, 51].
Information Protection: Other studies concentrate on information security, such as access control systems and trust management. For example, utilizing trust level classification approaches, an approach was presented to secure instant community data access. When users determined the trust communication configurations for Instant Social Networking (ISN), which can be supported by reputation assessments, this technique was effective [52 - 54]. Another recent study employed ontology-based authentication classifications to develop an inter crossed access control strategy for safeguarding multimedia big data in cloud computing. These studies, however, were primarily concerned with securing data transmissions and authentications. When data is stored on the cloud, the techniques have little control. Furthermore, it is intended to safeguard data on cloud servers through the use of encryption-oriented technologies [55]. Previous studies, such as FHE and ABE, have also examined this topic. Despite the fact that these types of safe measures can successfully protect data from target attackers, such as external harmful activities and internal inappropriate operations, the performance of data processing might be negatively damaged by the additional computations [51, 56]. Some operations are even impossible to complete due to technical constraints, such as sounds in FHE [57].
Network: Because cloud computing is on-demand, it is vulnerable to security concerns such as data loss, data leakage, Denial-of-Service (DoS), account or service traffic hijacking, and malevolent insiders. Because of a careless cloud service provider, a malicious hacker may modify crucial data. To mitigate such danger, essential data must be encrypted and encryption keys must be kept secure [58, 59]. If an intruder acquires access to a client's credentials stored in the cloud, it can eavesdrop on transactions and activities, maliciously change the data, return false information, and divert the customer to hostile sites. DoS attacks are another important risk for

cloud platforms because most firms rely on one or more services to be available 24 hours a day, seven days a week. Customers may incur costs if one or more services are denied, especially if they are billed based on disk space utilization and compute cycles [60]. Another important vulnerability that clouds platform face is account or service hijacking. Hijacking a service allows a hostile person to get access to critical and sensitive portions of a deployed service, potentially jeopardizing its integrity, availability, and confidentiality. A malicious insider, such as a current or former employee, a business partner, or a contractor, may obtain harmful access to the data, network, or system [59]. When the cloud service provider is completely accountable for data protection, the situation deteriorates. Because of their distributed nature, cloud platforms attract more threats. It is preferable that the data (in this case, video contents) be safeguarded and only accessible in encrypted form. Directly hiding the data in encrypted HEVC video streams to maintain its quality can prevent video content leaking. This data concealment capability can solve the security and privacy problems connected with cloud computing. A cloud server can include additional video information, such as video notation and data authentication, inside an encrypted version of the HEVC standard [42, 61].

Once the data has been concealed, the server can validate its integrity without knowing the original contents. As a result, the encrypted data's confidentiality and privacy are safeguarded. Various works on encrypting data in videos can be found in the literature. [62] presents a novel encryption method for intra and inter frames in MPEG films. The authors stated that encrypting the entire video is necessary for highly sensitive and private movies. As a result, not only the Intra frames, but also the Inter frames, had to be encrypted. An MPEG video encryption technique is proposed. The scheme proposed was based on the Advanced Encryption Standard-128bit (AES128) algorithm. Only the Intra frames of a certain video were encrypted because the Inter frames are worthless without knowing the corresponding Intra frames [42].

A unique approach for H.264 data encryption based on selective encryption is given. The proposed approach provided transparent encryption as well as security against a variety of threats. When it came to retaining the formation and length of video streams, the proposed encryption and decryption were relatively faster. Various clouds, including Google App Engine and Amazon Web Services, have faced security breaches in recent years. Illegal users take advantage of these security flaws to steal confidential information or disrupt the normal operation of the Internet. As a result, devices interacting with cloud platforms must use robust and lightweight authentication and authorization schemes. Cloud computing is a variant of client-server architecture model, where, thousands of clients use the same infrastructure at a lot larger scales. Identity and access control management is a basic requirement for cloud computing [63].

As a result, better authentication is necessary in comparison to the traditional client-server interaction architecture. For mobile attacks, the authors [43] presented a cloud-assisted privacy-preserving key management approach. Their proposed approach safeguarded the identity and location of patients. Furthermore, the proposed approach did not take critical privacy and updating considerations into account. As a result, their method proved incompatible with a real-time cloud computing platform. Recently, bilinear pairing in an elliptic curve has received interest in the development of an ID-based cryptosystem. This cryptosystem addressed the high-cost authentication and public key management issues that traditional public key cryptosystems had. Each user's identification was utilized as a public key. As a result, a user did not incur additional computational costs when confirming the public keys of other users. Furthermore, no additional storage space on the user's device was required to keep the other users' public keys and related certificates [43]. Several studies have recently used ID-based cryptosystems in various cloud contexts. Wang et al (2013b) proposed a novel ID-based authentication mechanism for cloud environments. Although the proposed approach was appropriate for a dispersed mobile cloud service environment, it lacked user privacy and traceability support. The majority of elliptic curve or bilinear pairing-based authentication techniques are developed for client-server environments. They are not applicable to distributed service environments where several service providers compete for the providing of diverse services. The user must manage various secret keys obtained from different service providers. To fix this problem, all service providers must use the same secret key. If an enemy has the secret key, it may appear as a legitimate service provider in order to fool the users. Furthermore, an intruder who obtains the secret key may also obtain the session keys. The attacker may eavesdrop on sensitive information exchanged between the user and another service provider after obtaining one or more session keys [43, 64].

VI. Conclusion

Most current active ways to addressing data abuse on the cloud-side have two options. The first option is to use regulatory compliance measures to limit employees' behaviour. This paradigm is not effectively managed by technical techniques, and cloud operators may fail to comply with security regulations. The other way is to use encryption to protect data against information leaking. However, due to computational overload and the lower operational efficiency level, this sort of data security cannot meet most current industry expectations; thus, more study on safeguarding huge data storage is required that takes care of the challenge with minimum overload and latency. This paper had only presented an overview for securing distributed big data storage in the cloud environment and the need for more research.

Reference

[1]. Glass Erin, (2022) https://www.digitalocean.com/community/tutorials/a-general-introduction-to-cloud-computing. Retrieved December, 2022.

[2]. Pansotra, E.A & Singh, E.S.P. (2015). Cloud Security Algorithms. International Journal of Security and its Applications, 9, 353-360.

[3]. Potey, M. M., Dhote, C & Sharma, D.H (2016). Homomorphic Encryption for security of Cloud Data. Procedia Computer Science, 79,175 – 181.

[4]. Li, Y., Gai., K., Qui L, M. & Zhao, H. (2017). Intelligent cryptography approach for secure distributed big data storage in cloud computing, Information Sciences, Volume 387,2017,Pages 103-115,ISSN 0020-0255,, https://doi.org/10.1016/j.ins.2016.09.005, (https://www.sciencedirect.com/science/article/pii/S0020025516307319)

[5]. Ghazali, Osman, 2017 Security Measurement as a Trust in Cloud Computing Service Selection and Monitoring. Journal of Advances in Information Technology Vol. 8, No. 2

[6]. Atayero, A. A, & Feyisetan, O. (2011). Security Issues in cloud computing: The Potentials of homomorphic encryption, Journal of Emerging Trends in Computing and information Sciences, 2, 546-552.

[7]. Wang, B., Li, M., Chow, S.S & Li, H (2013). "Computing encrypted cloud data efficiently under multiple keys. Communications and Network Security (CNS)" IEEE Conference, pp 504-513.

[8]. Samanthula, B.K, Howser, G., Elmehdwi., Y & Madria, S. An efficient and secure data sharing Framework using homomorphic encryption in the cloud. Proceedings of the 1st International Workshop on Cloud intelligence, 2012. ACM, 8.

[9]. Gital, A. Y, Ismail, A.S. Chen. M. & Chiroma. H (2014). "A Framework for the Design of Cloud Based Collaborative Virtual Environment Architecture". Proceedings of the International Multiconference of Engineers and computer Scientist, 2014.

[10]. Wei, L. Zhu., H. Cao, Z. Dong, X, Jia, W., Chen, Y. & Vasilakos. W.V (2014). "Security and privacy for storage and computation in cloud computing", Journal of Information science, 258, 371-386.

[11]. Van Dijk, M. & Juels, A. (2010). "On the impossibility of Cryptography Alone for Privacy-Preserving Cloud Computing". HotSec., 10, pp 1-8.

[12]. Brakerskl, X. (2012). "Fully homomorphic encryption without modulus switching from classical GapSVP. Advances in Cryptology-CRYPTO". Springer.

[13]. Yagisawa, M. (2015). "Fully Homomorphy Encryption without bootstrapping. IACR Cryptology eprint" Archive, 2015, pp 474.

[14]. Modi, C., Patel, D.C Borisaniya., Bo., Patels, A. & Rajarajan, M. (2013). "A survey on security issues and Solutions at different layers of Cloud computing". The Journal of supercomputing, 63 561-592.

[15]. Herrera-Viedma, E., Cabrerizo, F.J., Kacprzyk, J. & Pedrycz, W. (2014). "A review of soft consensus models in a fuzzy environment". Information Fusion, 17, pp 4-13.

[16]. Siemens, G., and Long P., 2011. Penetrating the Fog: Analytics in Learning and Education. Educause Review, Vol. 46, No. 5.

[17]. Chen H, Chiang RHL, Storey VC (2012) "Business intelligence and analytics: from big data to big impact". MIS Quarterly 36(4):1165–1188

[18]. Russom P (2011) Big data analytics. TDWI Best Practices Report, Fourth Quarter, 1-35

[19]. Vasudeva R. & Chanfrashekara S. N (2016) A comparative study on big data storage and information retrieval: A Review International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835

[20]. Doug Lancey (2001) "Metagroup paper; 3d data management: Controlling Data Volume, Variety, and Velocity" Gartner.

[21]. Open source search through goggle; http://www.internetlivestats.com/google-search-statistics/ Retrieved October 2022.

[22]. Jenn Cano (2014) "The V's of Big Data: Velocity, Volume, Value, Variety, and Veracity" (https://www.xsnet.com/blog/bid/205405/The-V-s-of-Big-Data. Retrieved October 2022.

[23]. Plummer, D. C., Smith, D., Bittman, T. J., Cearley, D. W., Cappuccio, D. J., Scott, D., ... & Robertson, B. (2009). Gartner highlights five attributes of cloud com-puting. Vol G00167182, 1-5.

[24]. Buyya, R., C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic(2009) "Cloud Computing and Emerging IT Platforms:Vision, Hype, and Reality for Delivering Computing as the 5th utility," Future Generation Computer Systems(25)6, pp.599-616.

[25]. Mell, Peter & Grance, Tim. (2011). The NIST definition of cloud computing (Draft). NIST Special Publication. 800. 145.

[26]. Wischik, D., Handley, M. & Braun, M. B (2008). "The resource pooling principle SIGCOMM" Computer Communication Review, 38, 47-52.

[27]. Hamdaqa, M., & Tahvildari, L. (2012). Cloud Computing Uncovered: A research Landscape in H. Ali & M. Atif (Eds), Advances in Computers Elsevier. pp. 41-85.

[28]. Yakimenko, O. A., Slegers, N. J., Bourakov, E. A., Hewgley, C. W., Bordetsky, A. B., Jensen, R. P., Robbinson, A. B., Malone, J. R. & Heidt, P. E, (2009). Mobile system for precise aero delivery with global reach network capability. In Control and Automotion, 2009, ICCA 2009, IEEE International Conference pp. 1394-1398.

[29]. Rajan R. A. P & Shanmugapriyaa (2012) "Evolution of Cloud Storage as Cloud Computing:" IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661 Vol 1(1) (2012), pg. 38-45

[30]. Dillion, T., Chen, W., & Chang, E. (2010). "Cloud Computing: issues and challenges, in Advanced Information Networking and Applications (AINA)" 24th IEEE International Conference pp. 27-33.

[31]. Shin, H. and E. Ellinger, A. (2013), "The effect of implicit service guarantees on business performance", Journal of Services Marketing, Vol. 27 No. 6, pp. 431-442. https://doi.org/10.1108/JSM-02-2012-0037.

[32]. Jula A., Elankovan S. & Zalina O. (2014). Cloud computing service composition: A systematic literature review. Expert Systems Applications. Pp 3809 - 3824

[33]. Balbudhe, P.O., & Balbudhe P. O. (2013), Cloud Storage Reference Model for Cloud Computing in International Journal of IT, Engineering and Applied Sciences Research (IJIEASR) 2(3) page 81.

[34]. Lin. I., Tong. J., Bohn. R., Messina. J, Badger L., & Leaf. D (2011). NIST Cloud Computing References Architecture. NIST Special Publication.

[35]. Arora, I., & Gupta, D.A. (2012). Opportunities, Concerns and Challenges in the Adoption of Cloud Storage.

[36]. Sivasakthi, T., Prabakaran, N., & Chennai, T. (2014). Applying Digital Signature with Encryption Algorithm of User Authentication for Data Security in Cloud Computing.

[37]. Mathew Mister (2022) "Advantages and Disadvantages of Cloud Storage" https://www.promax.com/blog/10-advantages-and-disadvantages-of-cloud-storage. Retrieved January 2023.

[38]. Malik N (2007) "Future challenges in context-aware computing" IADIS International Conference.https://scholar.google.com.pk/citations?view_op=view_citation&hl=en&user=oYdExzwAAAAJ&citation_for_view. Retrieved January 2023.

[39]. Wang. C, Ren. K, Lou. W & Li (2012) "Toward publicly auditable secure cloud data storage services. Network", IEEE Proc. pp.19-24.

[40]. Subashini S, Kavitha V (2010)"A survey on security issues in service delivery models of cloud computing". Journal Network Computer Application.

[41]. Todd Steiner (2012) "An Introduction to Securing a Cloud Environment" The San Institute www.sans.org/reading-room/whitepapers/cloud/introduction-securing-cloudenvironment-34052. Retrieved January 2023.

[42]. Pedraycz, W. (2014). Allocation of information granularity in optimization and decision-making models: towards building the foundations of granular computing. European Journal of Operational Research, 232, 137-145.

[43]. Wang, C Chow, S.S, Wang, Q., Ren, K. & Lou, W. (2013b). Privacy-preserving public auditing for secure cloud storage. IEEE Transactions on Computers, 62, 362-375.

[44]. Wang, C., Wang, Q., Ren K. & Lou. W (2010). Privacy-preserving public auditing for data storage security in cloud computing. INFOCOM, Proceedings IEEE, pp, 1-9

[45]. Dayandanda, R and Someswar, G. M (2015). "Time-Based Proxy Re-encryption Scheme for Secure Data Sharing in a Cloud Environment". International Journal of Emerging trends in Science and Technology, 2.

[46]. Chadwick, D.W. & Fatema, K. (2012). "A primary preserving authorization system for the cloud." Journal of Computer and System Sciences, 78, pp 1359-1373.

[47]. Cao, N, Wang, C, Li, M., Ren K & Lou, W. (2014). Privacy- preserving multi-keyword ranked search over encrypted cloud data. IEEE Transactions on Parallel and distributed systems, 25, 222-233.

[48]. Gai, K. Qiu, M. Thuraisingham, B. & Tao, L (2015b). "Proactive attribute-based secure data schema for Mobile cloud in financial industry. High performance Computing and Communications (HPCC)," IEEE 7th International Symposium on Cyberspace Safety and Security (CESS), pp 1332-1337.

[49]. Wang, B., Yu, S., Lou, W. & Hou. Y.T (2014) "Privacy-Preserving multi-keyword fuzzy search over encrypted data in the cloud. IEEE INFOCOM– IEEE Conference Computer Communications, pp 2112-2120.

[50]. Song T., Pan L., Paun G. (2013). "Asynchronous spiking neural P systems with local synchronization" Journal of Information System and Sciences. 219, pp 197- 207.

[51]. Plantard, T., Susilo, W. & Zhang, Z. (2013). Fully homomorphic encryption using hidden ideal lattice. IEEE Transactions on information forensics and security, 8 2127-2137.

[52]. Yan, Z., Chen, Y. & Shen, Y. (2013). A Practice Reputation System for Pervasive Social Chatting. Journal Of Computer and System Sciences, 79, 556-572.

[53]. Yan, Z Wang. M & Zhang, P.A Scheme to Secure Instant Community Data Access Based on Trust and Contexts, Computer and Information Technology (Cit), 2014 IEEE International Conference On, 2014a. pp 646-651.

[54]. Yan, Z., Zhang, P. & Vasilakos, A.V (2014b). A Survey on Trust Management for Internet of Things Journal of Network and Computer Applications, 42, pp 120-134.

[55]. Li, Y., Gai, K., Ming, Z., Zhao, H. & Qiu, M. (2016a). "Inter crossed Access Controls for Secure Financial Services on Multimedia Big. Data in Cloud System" ACM Transactions on Multimedia Computer, communications, and Applications (TOMM),12, pp 67.

[56]. Aliev, R.Pedrycz, W., Fazlollahi, B., Huseynov, O.H, Alizadeh, A.V. & Guirimov, B (2012). "Fuzzy Logic-Based Generalized Decision Theory with Imperfect Information", Journal of Information Sciences, 189, pp18-42.

[57]. Yagisawa, M. (2015). "Fully Homomorphy Encryption without bootstrapping". IACR Cryptology eprint Archive, 2015, pp 474.

[58]. Gital, A. Y, Ismail, A.S. Chen. M. & Chiroma H.A (2014). "Framework for The Design of Cloud Based Collaborative Virtual Environment Architecture." Proceedings Of the International Multi Conference of Engineers and Computer Scientist.

[59]. Alahmadi, A., Abdelhakim, M. Ren, J. & Li, T. (2004). "Defense against primary user emulation attacks in cognitive radio networks using advanced encryption standard." IEEE Transactions on Information forensics and security, 9, pp 772-781.

[60]. Gai, K., Qiu, M., Chen L—C & Liu, M (2015a). "Electronic health record error prevention approach using ontology in big data. High Performance Computing and Communications (HPCC)," IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conference on Embedded Software and System (ICESS), IEEE 17th International Conferences, pp 752-757.

[61]. Gai., Qiu. M., Zhao, H. & Dai, W (2015c). "Anti-Counterfeit Scheme Using Monte Carlo Simulation for E-Commence in Cloud System. Cyber Security and Cloud Computing (CS Cloud), IEEE 2nd International Conference, pp 74-79.

[62]. Parakh, A. & Kak, S. (2009). "Online data storage using implicit security". Journal of Information Science, 179, pp 3323-3331.

[63]. Gai, K., Qiu, M., Zhao, H., Tao L. & Zong, Z. (2016b). "Dynamic Energy-aware cloudlet-based mobile cloud computing model for green computing:" Journal of Network and Computer Applications, 56, pp 46-54.

[64]. Qi, W., Zhou, P., & Ye, W. (2021). "Analysis and Research of Data Encryption Technology in Network Communication Security." Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, 392 LNICST, 503–511. https://doi.org/10.1007/978-3-030-87903-7_62.

Anah Hassan Bijik, et. al. "A Review on Securing Distributed Big Data Storage in Cloud Environment." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 25(1), 2023, pp. 17-28.