IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 6, June 2022

Data Storage and Security in Cloud Computing: A Survey

Sara Dhingra
B. Tech Student (CSE)
Dronacharya College of Engineering, Gurgaon, Haryana, India

Abstract: *Cloud computing is envisioned as the next generation of IT enterprise design. Cloud computing distributes application software and data bases to massive data centres, where data and service management may not be completely reliable. This creates a slew of new security issues that have yet to be properly addressed. In this article, we primarily focus on features for ensuring data storage security in the cloud, as well as architecture for data storage that is applied by other cloud service providers vendors, and essential points for establishing data storage security.*

Keywords: Cloud Computing, Cloud Storage Techniques, Security Techniques, Architecture, S3, API, TPAI

I. INTRODUCTION

Several factors are ushering in the era of Cloud Computing [9], which is computer technology development and use based on the Internet. Data centres are being transformed into massive pools of computing services as processors get cheaper and more powerful, combined with the software as a service (SaaS)[8] computing architecture. Users can now subscribe to high-quality services using data and software that resides only on remote data centres, thanks to rising network capacity and stable yet flexible network connections. Moving data to the cloud has a lot of advantages.

Users benefit from the convenience of not having to worry about the intricacies of direct hardware administration. Amazon Simple Storage Service and Amazon Elastic Compute Cloud are the first cloud computing companies.

While new internet-based online services offer vast amounts of storage space and customised processing power, this shift in computing platforms also eliminates the need for local workstations to maintain data.

As a result, consumers' data availability and integrity are at the mercy of their cloud service providers. The recent outage of Amazon's S3[4] service is one example of this. Cloud storage has a number of advantages. There's no need to spend money on storage devices. There's no need for a technical specialist to keep track of storage, backup, replication, and, most crucially, disaster recovery. Allowing people access to your data will result in collaborative work rather than solo work.

II. SERVICES IN CLOUD COMPUTING

1. **SaaS:** Software as a Service (SaaS)[8] is the most popular and user-friendly kind of cloud computing. SaaS is a delivery model that leverages the Internet to offer programmes that are managed by a third-party vendor and have a client-facing interface. The majority of SaaS apps may be accessed immediately from a Web browser, without the need for any downloads or installations. SaaS reduces the requirement for individual PCs to install and operate apps. Because everything can be controlled by vendors, organisations may streamline their maintenance and support using SaaS. This includes apps, runtime, data, middleware, O/S, virtualization, servers, storage, and networking. Gmail, Google Apps, Microsoft Office 365, Google+, Facebook, and Yahoo are all examples of SaaS.
2. **PaaS:** Platform as a Service (PaaS)[8] uses a platform to supply computational resources. PaaS provides developers with a framework upon which they may create or customise apps. PaaS eliminates the need to purchase the underlying layers of hardware and software, making application development, testing, and deployment rapid, simple, and cost-effective. One distinction between SaaS and PaaS is which components of the service must be controlled by customers rather than providers: Vendors continue to manage runtime, middleware, operating systems, virtualization, servers, storage, and networking with PaaS, while consumers

DOI: 10.48175/IJARSCT-4982

control applications and data. AWS Elastic Beanstalk, Windows Azure, Heroku, Force.com, and Google App Engine are all examples of PaaS.

3. **Infrastructure as a Service (IaaS):** Infrastructure as a Service (IaaS)[8] provides computer infrastructure (such as a platform virtualization environment), storage, and networking. Users can acquire software, servers, and network equipment as a fully outsourced service that is often priced according to the quantity of resources used consumed. In essence, a third party permits you to establish a virtual server on their IT infrastructure in exchange for a leasing fee. Users of PaaS and IaaS must manage more than SaaS users: applications, data, runtime, middleware, and operating system. Virtualization, servers, hard drives, storage, and networking are still managed by vendors. Users benefit from IaaS since it provides infrastructure on which they may install any desired platforms. If new versions are released, users are responsible for upgrading them.
4. **Storage as a Service (SaaS):** Storage as a Service (StaaS)[5] enables cloud applications to extend beyond their constrained server capacity. Users can store their data on remote discs and access it from anywhere with StaaS. Cloud storage systems are expected to meet a number of stringent requirements for preserving users' data and information, including high availability, reliability, performance, replication, and data consistency; however, no single system can meet all of these requirements simultaneously due to their incompatibility.
5. **Amazon S3:** It is another option. Amazon S3[4] is an internet storage service. It's intended to make web-scale computing more accessible to programmers. Amazon S3 is a simple web services interface that allows you to store and retrieve unlimited amount of data from anywhere on the internet at any time. It provides any developer with the same highly scalable, dependable, secure, fast, and low-cost infrastructure that Amazon utilises to host its global network of websites. The service tries to optimise scalability benefits and pass them on to developers. According to the Spring 2010 Storage magazine/Search Storage Purchasing Intentions poll, 14% of respondents are currently using cloud storage, with disaster recovery being the most popular use (6 percent). However, 4% of companies use it to store primary data from their data centres, and an equal amount use it for near-line data storage. However, there are a few things you should know before signing up with a cloud storage service provider. Is it safe to store data on the cloud? How much will it set you back? What services are most beneficial to small businesses? We've gathered our best suggestions and professional guidance in one spot in our cloud storage services guide for beginners so you can receive answers to your most pressing problems. Learn about cloud backup, archiving, disaster recovery, and storing primary data on the cloud.

III. CLOUD STORAGE MODELS

There are cloud storage methods that allow customers to keep control of their data. Cloud storage [2] has evolved into three categories, one of which allows for the cost-effective and secure combination of two categories. Storage infrastructure is presented as a leasable commodity by public cloud storage providers (both in terms of long-term or short-term storage and the networking bandwidth used within the infrastructure). Private clouds are based on the same principles as public clouds, but in a format that may be safely incorporated within a user's firewall. Finally, hybrid cloud storage combines the two approaches, allowing policies to specify which data must be kept private and which data can be secured in public clouds (see Figure 1).

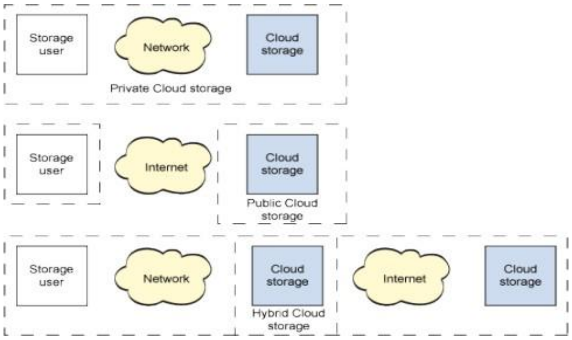


Figure 1: Cloud storage models

DOI: 10.48175/IJARSCT-4982

Figure 1 depicts the cloud models graphically. Amazon is an example of a public cloud storage provider (which offers storage as a service). IBM[1], para scale, and Clever safe (all of which provide software and/or hardware for internal clouds) are examples of private cloud storage providers. Finally, Egnyte is one of the hybrid cloud providers.

IV. DATA STORAGE SECURITY TECHNIQUES IN CLOUD COMPUTING

This paper discusses a number of existing strategies [2]. Cloud storage is defined as a network of distributed data centres that uses virtualization technologies and provides a data storage interface.

4.1 Implicit Storage Data Security in the Online World

In cloud computing, providing implicit storage security to data is more helpful. For providing such security, a data partitioning strategy employing the roots of a polynomial in a finite field is used. Data is partitioned in this approach in such a way that each chunk is implicitly secure and does not require encryption. These parts are stored on distinct network servers that are only known by the user. Access to each server, as well as knowledge of where the data sections are stored, are required for data reconstruction. Several variations of this technique are described, including one in which encryption keys are stored implicitly rather than data, and another in which a subset of the data is stored.

4.2 Authentication on the Basis of a Person's Identity

For IBHMCC, there are identity-based encryption (IBE) and decryption techniques, as well as identity-based signature (IBS).

Consumers have access to a wide range of resources and services. As a result, several security threats are possible. As a result, cloud security necessitates user and service authentication. It becomes extremely difficult when SSH Authentication Protocol (SAP) is used in the cloud. As an alternative to SAP, a new identity-based authentication protocol with a corresponding signature and encryption technique was presented. The sequence of stages is limited by the identify-based authentication technique. The client C sends a client Hello message to the servers in step 1. Cn.session identifier ID and c specification are included in the message.

4.3 Complete Data Dynamic Support for Public Auditing

Data integrity verification on unreliable servers is a major worry in cloud storage. A trusted organisation with experience and capabilities that data owners lack can be assigned as an external audit party to examine the risk of outsourced data when needed. It also gives data owners a transparent and cost-effective way to earn trust in the cloud. The existing proof read of PDF (or) POR technique is upgraded to achieve dynamic data support by spoofing the underlying Markel Hash tree (MHT).

4.4 Appropriate Third-Party Auditing (TPA)

Because cloud users save data on a cloud server, security and data storage accuracy are top priorities. Data owners with large amounts of outsourced data may find auditing the data's correctness in a cloud environment challenging and costly. To allow third-party auditing, in which users may safely delegate integrity-checking responsibilities to third-party auditors (TPA)[2], this scheme can practically guarantee instantaneous data error localisation (i.e. the identification of misbehaving servers). To give security to various cloud kinds, an innovative and homogenous structure is introduced. Before outsourcing data to the cloud, the BLS (Bonch-Lynn-Sachems) algorithm is used to sign data blocks to ensure data storage security.

4.5 A Method for Dynamically Storing Data in the Cloud

Because the clients did not have a local copy of the data saved in the cloud, data storage in the cloud may not be totally trustworthy. To address these concerns, a new protocol system based on the data reading protocol algorithm was presented to check data integrity. The proposed effective automatic data reading algorithm assists clients in checking data security. These systems use homomorphism tokens, blocking erasure and unblocking factors, and distributed erasure coded data to create a flexible distributed storage integrity auditing mechanism (FDSIAM).

4.6 Storage Protocol That Is Both Effective and Secure

Users are increasingly outsourcing data to service providers with sufficient storage space and lower storage costs. The proposal is for a safe and efficient storage protocol that ensures data storage confidentiality and integrity. This protocol was created using the elliptic curve cryptography building method, and a sober sequence is used to verify data integrity[2]. Cloud users do a data and software process protocol step to add the privacy enforcement structure to software and data before moving them to the cloud. The challenge response mechanism is credentialed such that the contents of the data are not exposed to outsiders. Data dynamic activities are also employed to maintain the same level of security and provide respite.

4.7 Data Storage Security

The data is secured in the server using the user's preferred security technique, ensuring that data is given top priority and resources are shared between servers to ensure data security in the cloud. Because of intruder attacks, sending data over the internet is risky. In a cloud setting, data encryption is critical. Implemented a secure cross platform and introduced a consistent and new structure for offering security to cloud kinds. Our scheme achieves the integration of storage correctness insurance and data error location (i.e., the identification of misbehaving server) using the proposed effective and flexible two-way handshakes based on token management by utilising the homomorpic token with distributed verification of erasure coded data.

4.8 A Reliable and Secure Storage Service

Customers can save their data on the cloud and use any of the available high-quality applications without having to worry about data storage. Despite the benefits to cloud providers, such a service relinquishes customer control over their data, introducing new valuability risks to cloud data accuracy. The homomorphism token and distributed coded-data were used to create a customizable distributed storage integrity auditing mechanism. Furthermore, the suggested approach allows for secure and efficient dynamic operations on external data, such as block change, deletion, and add.

4.9 Cloud storage systems that are optimal

Individuals, businesses, and institutions are increasingly turning to cloud data storage for backup and synchronisation since it needs no effort. At a high level, the suggested system specifies a possible architecture for a cryptographic storage service. A data processor (DP) analyses data before it is transmitted to the cloud, a data verification (DV) examines whether data in the cloud has been tampered with, and a token generator (TG)[2] generates tokens that allow cloud storage providers to recover portions of user data.

4.10 Small file access and storage process with storage

Hadoop distributed file system server reasons for small file issue of native Hadoop distributed file system are investigated in order to provide comprehensive support for services. Nane is under a lot of pressure. The node of the HADOOP distributed file system is enforced by a high number of small files, and there is no prefetching method for data placement correction. To address these small-scale issues, a method was proposed. In a big cluster, hundreds of computers both host directly associated storage and execute user application tasks, which enhances tiny file efficiency on Hadoop distributed file system. The resource can grow in response to demand by distributing storage and computing over multiple servers.

4.11 Management of file storage security

To ensure the security of data kept in the cloud, a solution based on a distributed architecture was presented. A master server and a group of slave servers are part of the proposed system. In the proposed approach, there are no direct commutation links between clients and slave servers. The master server is in charge of processing client requests, while the slave server is in charge of chunking operations in order to offer data backup for future file recovery. Client files are kept on the main server as tokens, and files are chunked on the slave server for file recovery.

V. CLOUD STORAGE ARCHITECTURE

Cloud storage architectures [3] are primarily concerned with providing on-demand storage in a highly scalable and multitenant environment. Cloud storage architectures, in general (see Figure 1), comprise of a front end that exposes an API for accessing the storage. This API is the SCSI protocol in traditional storage systems; however, these protocols are changing in the cloud. Web service front ends, file-based front ends, and even more typical front ends can all be found there (such as Internet SCSI, or iSCSI). The storage logic is a layer of middleware that sits behind the front end. Over typical data-placement techniques, this layer incorporates a range of characteristics, such as replication and data reduction (with consideration for geographic placement).

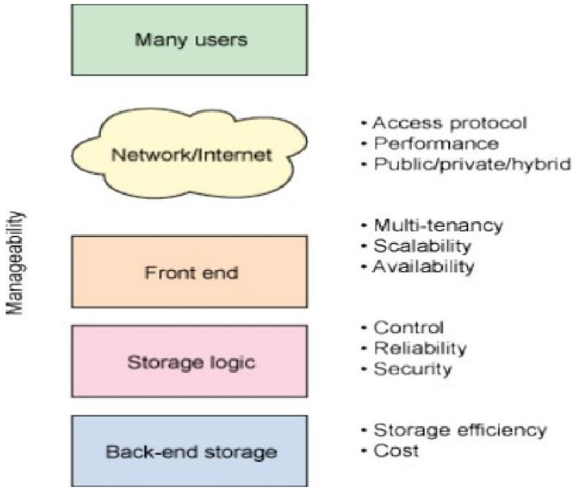
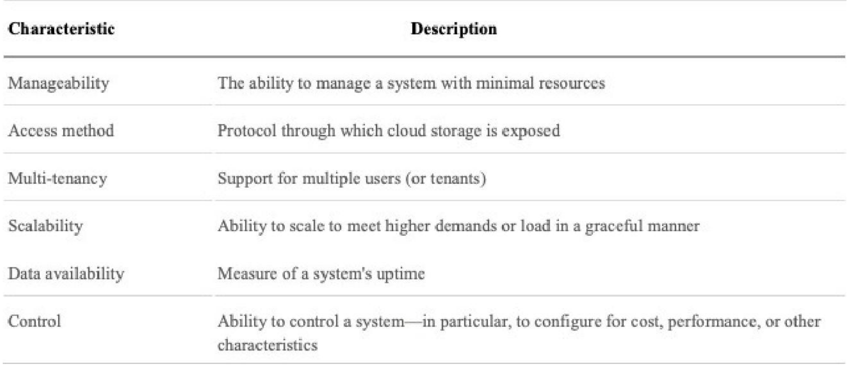


Fig 2: Cloud storage architecture

Some of the characteristics of contemporary cloud storage architectures can be seen in Figure 2[3]. It's worth noting that no qualities are limited to the layer in question; rather, they serve as a reference for the specific issues it addresses. Table 1 lays forth these qualities.

Table 1: Cloud Storage Characteristics



Characteristic	Description
Manageability	The ability to manage a system with minimal resources
Access method	Protocol through which cloud storage is exposed
Multi-tenancy	Support for multiple users (or tenants)
Scalability	Ability to scale to meet higher demands or load in a graceful manner
Data availability	Measure of a system's uptime
Control	Ability to control a system—in particular, to configure for cost, performance, or other characteristics

VI. CLOUD STORAGE API (APPLICATION PROGRAMMING INTERFACE)

A Cloud Storage Application Programming Interface (API)[7] is a way for gaining access to and using cloud storage. REST (Representational State Transfer) is the most common of them, however there are others that are based on SOAP (Simple Object Access Protocol). All of these APIs are related to making service requests via the Internet. REST is a well acknowledged approach to designing "high-quality" scalable APIs. The fact that REST is a "stateless" architecture is one of its most fundamental characteristics. This means that the request contains everything needed to fulfil the request to the storage cloud, eliminating the need for a session between the requestor and the storage cloud. It's crucial because the Internet is notoriously obfuscated (it has an unpredictable response time and it is generally not fast when compared to a local area network). REST is a style of programming that is quite similar to how the Internet operates. Because of the delay, traditional file storage access mechanisms such as NFS (network file system) or CIFS (Common

Internet File System)[7] do not work over the Internet. Files are stored in cloud storage, which some refer to as objects and others refer to as unstructured data. Consider the files on your computer, such as photos, spreadsheets, and documents. These are unstructured due of their extreme variety. Block or structured data is the other type of data. Consider data from a database, which is used to feed transactional systems that demand a specific level of guaranteed or low-latency performance. This is not a circumstance where cloud storage would be appropriate. Unstructured data accounts for over 70% of all machine-stored data worldwide, according to the Industrial Design Centre (IDC), and is also the fastest increasing data type. As a result, Cloud Storage is online storage for files. This does not exclude you from accessing Cloud Storage over a private network or LAN, which may also allow you to access a storage cloud using other methods such as NFS or CIFS. It does imply that a REST API is the most common and preferred method of access. REST APIs are language agnostic, which means they can be used by developers working in any programming language. A URL can be used to access resources within the system. As a result, an API is not a "programming language," but rather the method by which a programming language can access a storage cloud.

REST APIs are also about modifying the status of resources by interacting with their representations. In a functional sense, they aren't about invoking web service methods. The URLs that define the resources and the structure of the representations are the main distinctions between different Cloud Storage APIs. Amazon S3 APIs, Eucalyptus APIs, Rackspace Cloud Files APIs, Mezeo APIs, Nivanix APIs, Simple Cloud API, and more, as well as the Cloud Storage Technical Work Group of the Storage Networking Industry Association (SNIA), and more.

VII. CONCLUSION

Because of its availability, scalability, performance, portability, and functional needs, cloud data storage is more favourable than traditional storage. We primarily focused on the data storage considerations that cloud service providers utilise to store data, as well as the security considerations that must be supplied for data kept in the cloud. We investigated the Amazon S3 [4] and third-party auditing (TPA)[2] systems for data storage and security in the cloud.

DOI: 10.48175/IJARSCT-4982