Cloud Storage and Security Overview

Oussama Arki
LIRE Laboratory, Constantine 2 University
Ali Mendjli, 25000 Constantine, Algeria
oussama.arki@univ-constantine2.dz

Abdelhafid Zitouni
LIRE Laboratory, Constantine 2 University
Ali Mendjli, 25000 Constantine, Algeria
abdelhafid.zitouni@univ-constantine2.dz

Abstract

Cloud storage is one of the cloud services, it allows users to store and manage their data remotely in the cloud. Nowadays, cloud storage has become an attractive storage scheme for users to store their data. When users store their files remotely in a cloud storage system, they still worried about the security of their files and the guarantee of the confidentiality and the integrity of them. In this work, we propose a study of cloud storage and the problem of security in this service. For that, first, we give an overview of this concept, second, we talk about the problem of security in cloud storage, and then we summarize the different techniques and methods that have been proposed in order to ensure the security in cloud storage systems.

Keywords Cloud Storage, Information Security, Confidentiality, Integrity, Availability, Encryption .

1 Introduction

In the last years, information technology has been widely developed, aiming to increase the power of computing and decrease their cost, which led to the emerging of new technologies.
In that context, cloud computing rapidly appeared s as IT solution of choice for many companies and individuals. According to the national institute of standards and technology, "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" [PT11].
Cloud storage is one of the cloud services, it allows the users to store and manage their data remotely in the cloud servers, Cloud storage is a model of networked online storage where data is stored on multiple virtual servers [PP13]. The cloud storage users are perplexed about how security can be guaranteed in the cloud storage system. Moreover, they are worried about hosting their sensitive data in the cloud storage systems. The characteristics of cloud computing can create a serious data risk as the same resources are used among the different users [CS16].
The adaptation of cloud storage needs a good comprehension of this service and the risks behind the outsourcing of our sensitive data in the cloud. In this paper, we give an overview of cloud storage system, and we discuss the problem of security in cloud storage, and then we classify the existing techniques that aim to secure the cloud storage into different categories.
The rest of this paper is organised as follows. Section 2 is an overview of cloud storage systems. Section 3 discusses the problem of security in cloud storage. In section 4, we classified the existing solutions which aim to secure the cloud storage. Finally, a conclusion is given in Section 5.

2 Cloud Storage overview

The cloud storage is one of the cloud services, it can be considered as a part of the infrastructure as a service, in this section we give a brief study of cloud storage.

2.1 Cloud storage

Cloud storage is a model of networked online storage where data is stored on multiple virtual servers, generally hosted by third parties, rather than being

Copyright © by the paper's authors. Copying permitted for private and academic purposes.

In: Proceedings of the 3rd Edition of the International Conference on Advanced Aspects of Software Engineering (ICAASE18), Constantine, Algeria, 1,2-December-2018, published at http://ceur-ws.org

CEUR-WS.org/Vol-2326/paper3.pdf

hosted on dedicated servers [PP13] , It allows users to store their data at remote disks and access them anytime from any place [VMB14]. Through based uses a Web-Cloud storage services may be accessed through a web service application programming interface(API), or interface[PP13].

2.2 Cloud Storage Architecture

Cloud storage architectures are primarily about delivery of storage on demand in a highly scalable and multi-tenant way. Generically, cloud storage architectures consist of[VMB14] :

front end that exports an API to access the storage. In traditional storage systems, this API is the SCSI protocol; but in the cloud, these protocols are evolving. There, you can find Web service front ends, file-based front ends, and even more traditional front ends.

the storage logic behind the front end is a layer of middleware that is called the storage logic. This layer implements a variety of features, such as replication and data reduction, over the traditional data-placement algorithms .

back end implements the physical storage for data. This may be an internal protocol that implements specific features or a traditional back end to the physical disks. Figure 01 presents the cloud storage architecture.

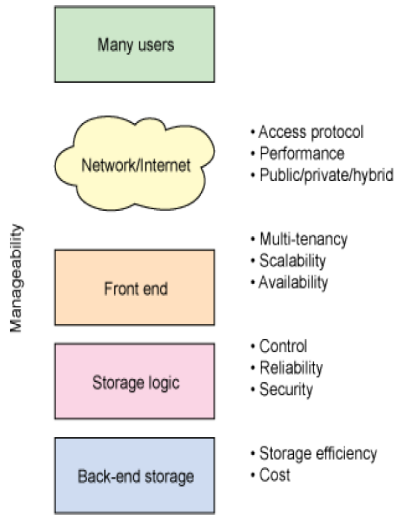

Figure 1: Cloud Storage Architecture

2.3 Definition of Cloud Storage Services

Many cloud storage providers are active on the market, offering various kinds of services to their customers. This study distinguishes between two types of cloud storage services[MTM+12].

Basic cloud storage services are generally not designed to be accessed directly by users but rather incorporated into custom software using application program-ming interfaces" (API).

Advanced cloud storage services mostly employ basic cloud storage services for the actual storage of data, and provide interfaces such as client or web applications which greatly simplify the use of the service for the customer.

2.4 The Existing Interface for cloud Data Storage

Since cloud Storage, cloud service providers began to make their own implementations available to users. As a result, a multitude of interfaces have been supplied that have been re-purposed for cloud storage, such as block-based access via iSCSI; POSIX interfaces (NFS, CIFS, and WebDAV); object-based CRUD (Create, Read, Update, Delete) interfaces over HTTP; and a plethora of proprietary interfaces for database or table access. Figure 02 presents the cloud storage existing interfaces[Zap12].

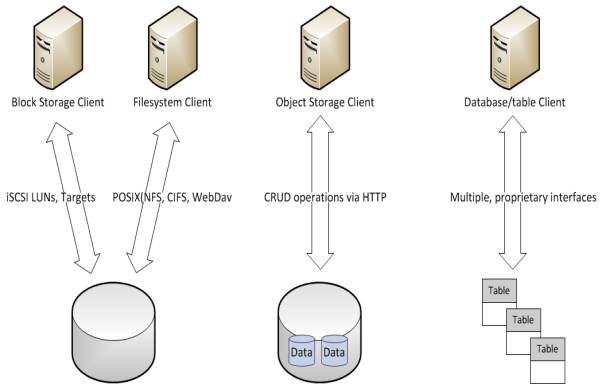

Figure 2: Existing Interfaces for Cloud Data Storage

CDMI

As cloud storage provides benefits , such as scalability and cost savings, the adoption of cloud storage is growing. However, each cloud storage provider offers its own cloud storage interface. As a result, multiple standards exist, which locks clients into proprietary solutions. The Storage Networking Industry Association (SNIA)'s response has been to develop the Cloud Data Management Interface (CDMI), an extensible standard that accommodates vendors' requirements and ensures consistency and interoperability for users [Zap12]. Figure 03 shows the cloud storage reference model.

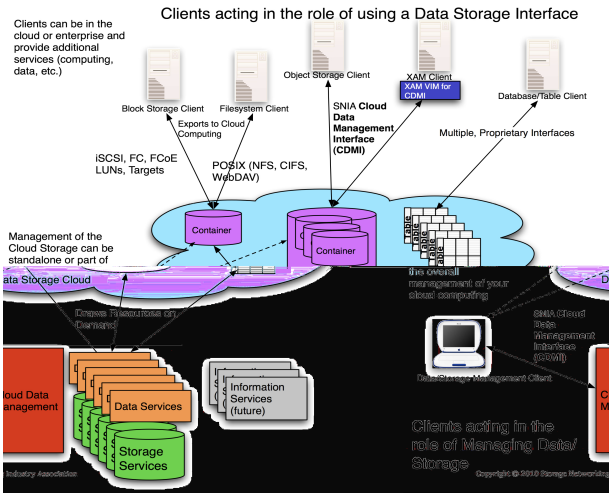

Figure 3: Cloud Storage Reference Model [2]

3 Cloud storage and security

In this section, we discuss the problem of cloud storage security.

3.1 Information Security

This term means protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide integrity, confidentiality and availability. Here, integrity means guarding against improper information modification or destruction and includes ensuring information non-repudiation and authenticity. Confidentiality means preserving authorized restrictions on disclosure and access, including means for defending proprietary and personal privacy information. Availability means ensuring timely and reliable access to and use of information[LSMMM13].

-Confidentiality: It means keeping user data secret in the Cloud systems. It ensures that the data of the user, which reside in the Cloud, cannot be accessed by an unauthorized person. There are two basic approaches to achieve such confidentiality, physical isolation and cryptography. Confidentiality can be achieved through proper encryption technique: symmetric and asymmetric algorithms [PS13].

-Data Integrity: Data integrity is one of the most critical elements in any information system. Generally, data integrity means protecting data from unauthorized deletion, modification, or fabrication. Managing entitys admittance and rights to specific enterprise resources ensures that valuable data and services are not abused, misappropriated, or stolen [YJYG14].

-Availability: Data should be available when it is requested via the owner. It ensures that user can be able to use the service anytime from any place. Two strategies called hardening and redundancy are mainly used to enhance the availability [PS13].

3.2 Security Challenges in Cloud Storage

With adoption of a cloud model, users lose control over physical security. Security overall covers mainly three aspects: confidentiality, integrity and availability (CIA). These aspects are the top most considerations in designing a security measure to ensure maximum protection. [FVRG14].

-Securing access to protected data is restricted to certain level of user authorised to access it. This requires mechanisms to be in place to control the access of protected data. The foundation on which access control mechanisms are built starts with authentication, authorization and encryption.

-Protecting data from loss and leakage envolves integrity of many parties involved in providing the resources. It is suggested to practice auditing techniques such as Proof-of-Retrivebility(POR) and Proof-of-Data Possession(PDP) to enable verification.

-High-available as access and data are getting secured, it is important to keep the hardware high-available. The hardware is the infrastructure hosting the services to store data and information. Without ensuring failover, the services are unable to meet the uptime and comply with service level managements.

4 Cloud Storage Security Solutions and Techniques

In the literature, many techniques and methods are proposed to provide cloud storage security. Those techniques and methods can be classified into encryption methods, identity and access management (IAM) techniques, data protection techniques and availability techniques.

4.1 Ensuring Confidentiality

Data confidentiality in cloud storage security refers to the property that information stored in the cloud storage is not made available or disclosed to unauthorized individuals, entities, or processes. Access control and data encryption have been widely deployed to protect data confidentiality [CtLZ+14].

4.1.1 Encryption of data

Cryptography is the art of keeping message secure by changing the data into non-readable forms [NAK16].

Traditional Cryptographic Techniques Traditional cryptography consists of three algorithms, Symmetric-key algorithms, Asymmetric-key algorithms and Hashing :

- Asymmetric cryptography is a class of cryptographic algorithms which requires two separate

keys, one of which is secret (or private) and one of which is public. Although different, the two parts of this key pair are mathematically linked. The public key is used to encrypt plaintext or to verify a digital signature; whereas the private key is used to decrypt cipher text or to create a digital signature [ZAH+15].

- Symmetric-key algorithms are a class of algorithms for cryptography that use the same cryptographic keys for both encryption of plaintext and decryption of cipher text. The keys may be identical or there may be a simple transformation to go between the two keys. The keys, in practice, represent a shared secret between two or more parties that can be used to maintain a private information link [ZAH+15].

- A hash function takes a data of variable length and produces a data of fixed length. It produces small and static length data which is unique for each data. The hash code is also specified as message digest or Hash value. Any kind of change to any bits in the data consequences in a huge alteration to the hash code [PPPS17].

Advanced Cryptographic Techniques In the beginning of the Cloud Computing, common encryption Technique like Public Key Encryption was applied. This traditional technique does not provide expected result as it support one to one encryption type communication [RDD15].

- Searchable Encryption: A searchable encryption scheme is applied at high level in order to encrypt the content that is available in search index so that it can hidden from others except the party that provide the authorised tokens.

- Homomorphic Encryption: the Homomorphic encryption scheme allows executing computations on the encrypted data. It is only of the advanced cryptographic technique. It has a slow processing time during computation.

- Identity based Encryption: In Identity Based Encryption, an identity of the user plays a vital role. The sender who sends the message only needs to know the receivers identity attribute in order to send the encrypted messages. However, key revocation is not achieved in Identity Based Encryption.

- Attribute-based Encryption: Attribute-based Encryption come up with access control. In Attribute Based Encryption, data owner uses a set of attributes to encrypt the data and only the authorized users who has the predicted or certain attributes can decrypt the data.

4.1.2 Identity and Access Management (IAM)

IAM remains one of the greatest challenges in cloud computing, IAM refers to the processes, technologies, and policies that manage access of identities to digital resources and determine what authorization identities have over these resources [EO16].
Identity Management (IdM) is the process of creating, managing, and using identities, and the infrastructure that provides support for these processes. In IdM, each person or application is identified by a credential, which represents a set of attributes, issued by a reliable source [JCC17]

- Identity in the Cloud Model: identity provider and service provider merge also in this model. This means for the cloud case that the cloud service provider, which hosts the application, is also responsible for the identity management [BTK14]. Figure 4 illustrates this model.

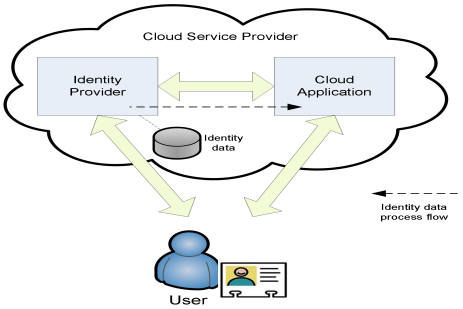

Figure 4: The Identity in the Cloud Model

- Identity to the Cloud Model: Also in this model, the identity provider takes over the tasks regarding identity management for the service provider. However, the main difference in this model is that the service provider and its applications are cloud-based [BTK14].Figure 5 illustrates this model.

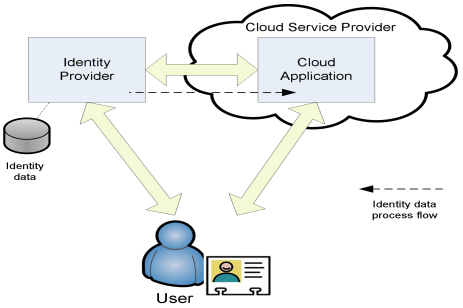

Figure 5: The Identity to the Cloud Model

- Identity from the Cloud Model: The identity from the cloud model fully features the cloud computing paradigm. In this case, both the cloud application and the identity provider are operated in the cloud. However, in contrast to the Identity in the Cloud-Model both entities are operated by distinct cloud service providers [BTK14]. Figure 6 illustrates this model.

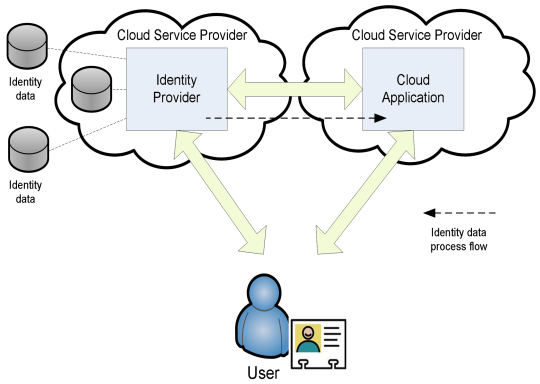

Figure 6: The Identity from the Cloud Model

- Cloud Identity Broker Model: The cloud identity broker model can be seen as an extension to the Identity from the Cloud-Model. In this Cloud Identity Broker-Model, the identity provider in the cloud acts now as an identity broker in the cloud. In other words, the cloud identity broker is some kind of hub between one or more service providers and one or more identity providers[BTK14].Figure 7 illustrates this model.

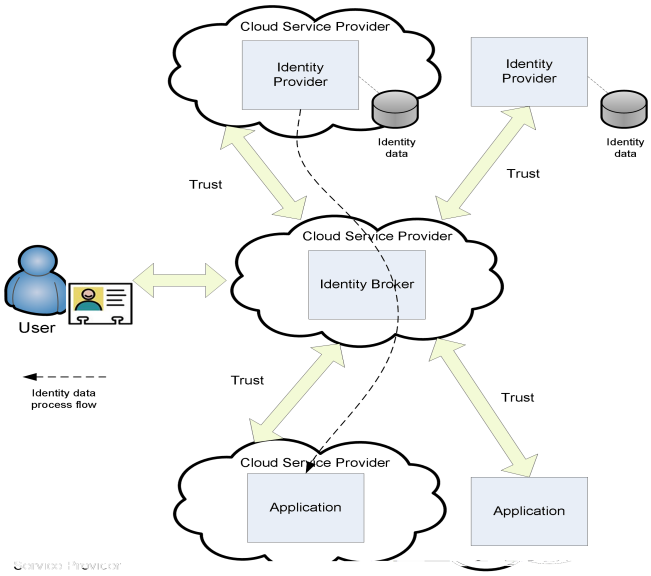

Figure 7: The Cloud Identity Broker Model

- Federated Cloud Identity Broker Model: The federated cloud identity broker model combines the traditional federated identity model with the newly Cloud Identity Broker-Model. This combined model aims on eliminating the drawbacks of the central Cloud Identity Broker-Model. The general architecture is illustrated in Figure 8, showing the federation of two different cloud identity brokers. Compared to the simple Cloud Identity Broker- Model, in this federated model users and service providers do not need to rely on one and the same identity broker. Actually, both the user and the service provider can rely on the individual broker of their choice[BTK14].

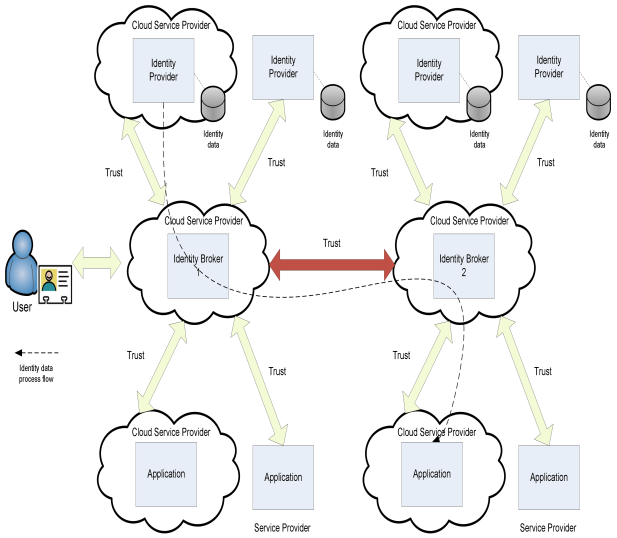

Figure 8: The Federated Cloud Identity Broker Model

Access Control Access control has been one of the key mechanisms to protect data confidentiality in traditional data networks. It is designed to block unauthorized users and malicious hackers from accessing data [CtLZ+14]:

- Attribute-Based Access Control (ABAC): it is based in attribute-based encryption (ABE).In ABAC model, access is granted based on attributes of the user. When applied to cloud storage, access control is enforced on data encrypted using ABE schemes. In an ABE system, a users keys and ciphertexts are labeled with sets of descriptive attributes. A particular key can decrypt a particular ciphertext only if there is a match between the attributes of the ciphertext and the users key [CtLZ+14].
- Role Besed Access Controle(RBAC): has also been commonly adopted in traditional storage system in order to simplify management of permissions. Its access policy is determined based

on different roles assigned to users by the system, while the data owner can specify a set of permissions of their data to different roles. By separation the tasks of role assignment and permission assignment, RBAC is much more efficient and scalable compared to other access control based on individual users, because the number of roles are usually significantly less than the number of users [CtLZ+14] .

4.2 Data Protection

A number of different techniques and mechanisms have been proposed and designed for cloud data integrity verification process. The mainstream of research in this field belongs to POR and Provable Data Possession (PDP). The two methods originally emerged with a similar concept but different approaches [CtLZ+14] .These techniques allow detecting data integrity damages without requiring a copy of the user local data. The idea was to encode the protocol with the data before storing it[FVRG14].

Provable Data Possesion (PDP): In this client precomputes tags for each block of a file. Then stores the file and its tags with a server. Later the client can verify that the server possesses the file by generating a random challenge against a randomly selected set of file blocks. Using the queried blocks and their corresponding tags, the server generates a proof of possession. The client is thus convinced of data possession, without actually having to retrieve file blocks [CGB14]

Proof of Retrievability (PoR): In this scheme ,first file is divided into blocks and then encoded with error correcting codes. Then check blocks called sentinels are embedded for each block. Encryption is performed to make check blocks indistinguishable from other file blocks. The verifier challenges the prover by specifying the positions of a collection of sentinels. The prover returns the respective sentinels. If prover has modified or deleted a substantial portion of file, then with high probability , it will have also suppressed a number of sentinels. Using error correcting code file can be recovered. otherwise it is tampered [CGB14].

4.3 Cloud Storage Availability

Availability refers to the system uptime and the system capability to operate continuously. Different techniques can be implemented to increase the system availability [WMF17].In the cloud data is stored using RAID (Redundant Array of Independent Disk). It provide way to store same data in different places in multiple disk (Redundantly) [Gor16].

RAID Features in Data Storage

RAID is a technology that combines independent physical disk drives into a single hard drive for the purpose of read/write speed improvement or reliability of stored data enhancement, or both . There are 2 implementations of RAID [MPNL16]:

Hardware RAID requires a RAID controller that controls Input/Output. Hardware RAID is used for host servers. It is high-performance yet expensive.

Software RAID: The operating system controls Input/Output. Software RAID is implemented on computers to boost the performance with low-cost solution.

RAID Levels

RAID uses many different architecture called levels, each level have a different scenario of disk and storage technique, depending on the balance between fault tolerance and performance. In cloud architecture levels of RAID describe how data distributed across the drives, there are 7 levels of RAID with different features, established on two basis levels RAID 0 and RAID 1 [MPNL16].

RAID 0: RAID 0 consists of at least 2 similar disks, which creates an array of n disks ($n \geq 2$). Data is split up evenly and get written across all devices in the array. Each disk stores 1/n data. The size of the array is the size of the smallest drive multiples the number of drives . Advantages: Read/Write transfer rate enhancement: Each disk has to Read/Write 1/n of the data. Theoretically, performance is n times higher. Disadvantages: Lower reliability. If one drive fails, all data in RAID 0 array is lost. Data loss rate of RAID 0 array is n times higher than the single-disks. Figure 9 shows the RAID 0 model.

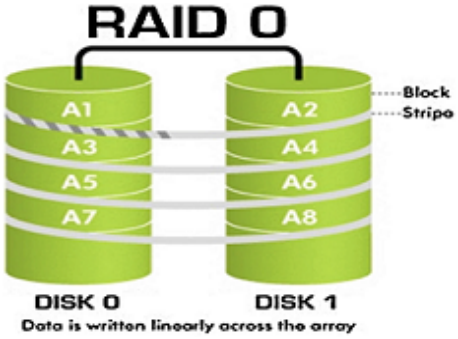

Figure 9: RAID 0 Storage Level

RAID 1: This is the simplest RAID level that provides data reliability. Like RAID 0, RAID 1 requires at least 2 drives to operate. Data are stored twice in 2 drives (Mirroring). If one disk fails, the other will continue to operate. Therefore, broken drive can be replaced without any worry of data loss. RAID 1 is not high-performance; however, it is essential for administrations and individuals that manage important data. An RAID 1 array capacity is the size of a single drive. Figure 10 shows the RAID 1 model.

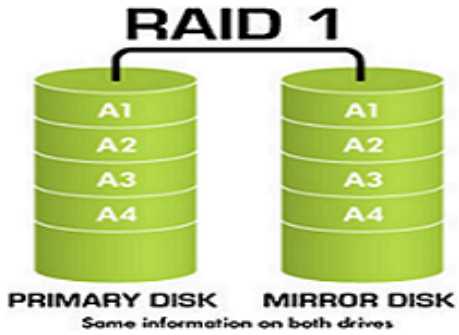

Figure 10: RAID 1 Storage Level

RAID 10: RAID 10 combines the approaches of RAID 1 and RAID 0. It requires a minimum of 4 drives to set up an RAID 10 array. Data are written on 4 drives at the same time: using Striping (RAID 0) on 2 drives, and using Mirroring (Raid 1) on the two others.RAID 10 is fast and secure. Performance are improved while reliability is ensured even if 1 drive fails. However, RAID 10 has its disadvantages of high cost, effective space is 1/2 of total size of 4 drives . Figure 11 shows the RAID 10 model.

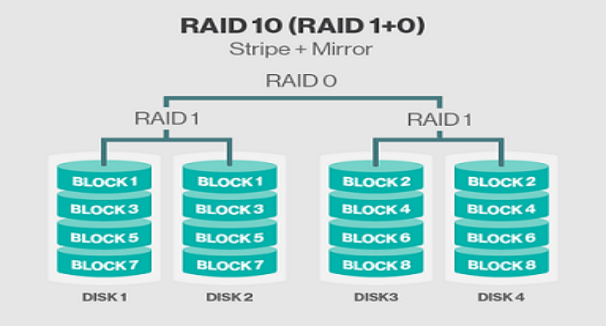

Figure 11: RAID 10 Storage Level

5 Conclusion

In cloud storage systems, there is always a big concern about data security. The guarantee of security is the main challenge for the cloud storage provider. First; they have to secure the access to the user's data, second; they have to guarantee the integrity of this data, and then; they have to provide a continuous and permanent access to this data. In this paper, we gave an overview of cloud storage and security. In the first step; we provided a whole study of this service; like the definition, it's architecture and the different forms of this service. In the second step; we discussed the problem of security in this service. In the last step, we summarized the cloud storage security technique according to the CIA properties of security. Table 1 summirizes the cloud storage security solutions and techniques.

Table 1: Cloud Storage Security Techniques.

Cloud Data Storage Security Techniques		
Encryption	Traditional	Symmetric
		Asymmetric
		Hashing
	Advanced	searchable
		Holomorphic
		Role based
		Attribute
Identity and Access Management	Identity Management	Identity in the Cloud-Model
		Identity to the Cloud-Model
		Identity from the Cloud-Model
		Cloud Identity Broker-Model
		Federated Cloud Identity Broker-Model
	Access Control	Attribute-based access control (ABAC)
		Role-based access control (RBAC)
Integrity	Provable Data Possesion (PDP)	
	Proof of Retrievability (PoR)	
Availability	RAID	Hardware RAID
		Software RAID

References

[BTK14] Zwattendorfer Bernd, Zefferer Thomas, and Stranacher Klaus. An overview of cloud identity management-models. In *10th International Conference on Web Information Systems and Technologies (WEBIST)*, pages 3–6, 2014.

[CGB14] V. Desai Charmee and Jethava Gordhan B. Survey on data integrity checking techniques in cloud data storage. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(12):293, December 2014.

[CS16] Prakash Chandan and Dasgupta Surajit. Cloud computing security analysis: Challenges and possible solutions. In *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, page 3, 2016.

[CtLZ+14] Huang Chun-ting, Huang Lei, Qin Zhongyuan, Yuan Hang, Zhou Lan, Varadharajan Vijay, and Jay Kuo C.-C. Survey on securing data storage in the cloud. *APSIPA Transactions*

on Signal and Information Processing, 3(2014):4,7–9, May 2014.

[EO16] Sturrus Edwin and Kulikova Olga. *Identity and Access Management*. Encyclopedia of Cloud Computing, 2016.

[FVRG14] Yahya F., Chang V., Walters R.J., and Wills G.B. Security challenges in cloud storage. In *IEEE 6th International Conference on Cloud Computing Technology and Science*, pages 1052–1054, 2014.

[Gor16] Ranvir Gorai. Deep dive into cloud computing. *International Journal of Research in Engineering, Technology and Science*, VI(Special Issue):4, July 2016.

[JCC17] Werner Jorge, Merkle Westphall Carla, and Becker Westphall Carlos. Cloud identity management: a survey on privacy strategies. *Computer Networks*, 122:3–4, July 2017.

[LSMMM13] Akter Lipi, Rahman S M Monzurur, and Hasan Md. Information security in cloud computing. *International Journal of Information Technology Convergence and Services (IJITCS)*, 3(4):18, August 2013.

[MPNL16] Le Quang Minh, Huy Anh Phan, Anh Chuyen Nguyen, and Khanh Duong Le. Research on enhancing security in cloud data storage. In *ICTA: International Conference on Advances in Information and Communication Technology*, pages 511–512, 2016.

[MTM+12] Borgmann Moritz, Hahn Tobias, Herfert Michael, Kunz Thomas, Richter Marcel, Viebeg Ursula, and Vowe Sven. On the security of cloud storage services. Technical report, Fraunhofer Institute for Secure Information Technology SIT, March 2012.

[NAK16] Hassan Hussein Nidal, Khalid Ahmed, and Khanfar Khalid. A survey of cryptography cloud storage techniques. *Int. Journal of Computer Science & Mobile Computing*, 5(2):186, February 2016.

[PP13] O. Balbudhe Pravin and O. Balbudhe Pradip. Cloud storage reference model for cloud computing. *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR)*, 2(3):83, March 2013.

[PPPS17] Parisha, Khanna Pooja, Sharma Puneet, and Rizvi Sheenu. Hash function based data partitioning in cloud computing for secured cloud storage. *Int. Journal of Engineering Research and Application*, 7(7):3, July 2017.

[PS13] Yadav Poonam and Sujata. Security issues in cloud computing solution of ddos and introducing two-tier captcha. *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, 3(3):29, June 2013.

[PT11] Mell Peter and Grance Timothy. The nist definition of cloud computing. Technical report, National Institute of Standards and Technology, September 2011.

[RDD15] Kirubakaramoorthi R., Arivazhagan D., and Helen D. Survey on encryption techniques used to secure cloud storage system. *Indian Journal of Science and Technology*, 36(8):2–4, December 2015.

[VMB14] Spoorthy V., Mamatha M., and Santhosh Kumar B. A survey on data storage and security in cloud computing. *International Journal of Computer Science and Mobile Computing*, 3(6):307–311, June 2014.

[WMF17] Bajaber Wejdan, AlQulaity Manahil, and S. Alotaibi Fahd. Different techniques to ensure high availability in cloud computing. *International Journal of Advanced Research in Computer and Communication Engineering*, 6(11):6, November 2017.

[YJYG14] Sun Yunchuan, Zhang Junsheng, Xiong Yongping, and Zhu Guangyu. Data security and privacy in cloud computing. *International Journal of Distributed Sensor Networks*, 2014:3, 2014.

[ZAH+15] Kartit Zaid, Azougaghe Ali, Idrissi H.Kamal, Marraki M.El, Hedabou M., Belkasmi M., and Kartit A. Applying encryption algorithm for data security in cloud storage. In *The International Symposium on Ubiquitous Networking*, pages 6–7, 2015.

[Zap12] Vytautas Zapolskas. Securing cloud storage service. Master's thesis, KTH Royal Institute of Technology, 2012.