Journal of Theoretical and Applied Information Technology
30th June 2024. Vol.102. No. 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DATA INTEGRITY CONCERNS, REQUIREMENTS, AND PROOFING IN CLOUD COMPUTING

[1]NABEEL AL-MILLI, [2]ZEYAD JOBAIR, [3]MOHAMMAD RASMI AL-MOUSA, [4]ALA'A AL-SHAIKH *, [5]MAHMOUD ASASSFEH, [6]RAED ALAZAIDAH, [7]ESSAM AL-DAOUD

Faculty of Information Technology, Zarqa University, 2000 Zarqa 13110, Jordan

Corresponding Author's E-mail: ashaikh@zu.edu.jo

ABSTRACT

Cloud computing enables users to maintain their data on remotely a cloud server and stay safe from harmful threats, such as impersonation attacks. Users opt for storing their data in external cloud storage, which provides them with usage and storage analysis services. Cloud computing is utilized to store vast volumes of images due to the rapidly growing use of images with an increasing level of detail brought on by improvements in images. The cloud has become a reality, nevertheless, there are many security issues or concerns that surround it, such as data integrity and unauthorized access, to name a few. Many provable data possession (PDP) and proofs of retrievability (POR) approaches have been put forth to evaluate the data integrity in the cloud. Moreover, distributing the data that is kept in the cloud amongst numerous users is still a problem because the company managing authentication and authorization for the cloud services is dishonest. In contrast to other publications that focus on the cloud as a whole, this paper lists concerns associated with data kept in cloud storage and resolutions to those matters.

Keywords: *Data Integrity, Cloud Computing, Data Security, Cloud Security, Integrity Protocols*

1. INTRODUCTION

Integrity for cloud-based data transactions is a crucial component of contemporary technology. Establishing trust and security in these transactions is crucial as more people and businesses use the cloud for data processing and storage. In essence, confidentiality, integrity, and availability, which are denoted by the CIA triad, constitute the basic requirements for information security (InfoSec) [1]. Data integrity, at its most basic level, is the assurance that data hasn't been tampered with or changed in any manner during transmission, processing, or storage [2]. Accordingly, InfoSec professionals seek to develop protocols and technology in cloud computing that guard against illegal data access, modification, and loss [3].

Scalability, accessibility, and cost-effectiveness are just a few of the primary benefits that cloud computing offers over conventional data storage techniques. Nevertheless, it also presents fresh difficulties for data security, such as problems with data privacy, data loss, and data theft [4]. Given the increasing reliance on cloud-based solutions for crucial processes, there have never been more pressing requirements for data integrity in transactions. These issues are being addressed by improvements in cloud security technologies including encryption, authentication, and access control, but continued work on the development and improvement of these technologies will be required to guarantee the integrity of data transactions in the future [5, 6].

There are many distinct facets to data security, yet our focus in this paper will be on data integrity [7]. Maintaining data integrity requires that the information must be kept in its original format. A lot of the research tackled accessibility, confidentiality, and integrity as they gained great attention from researchers [8]. As a result, message authentication codes, checksums, trapdoor hash functions, and Reed-Solomon codes were implemented. Moreover, digital signatures and other techniques have made it possible to effectively verify the integrity of data storage in conventional systems. No matter how safe cloud service providers (CSPs) claim their products are, the data holder still requires a way to confirm their data kept remotely on a trusted cloud server. In other words, a CSP needs to make it possible for independent third parties to verify the information. Consequently, the client's desired data can be easily obtained and downloaded from the server in this manner.

Data integrity is simply described as the absence of damage in the data that can be guaranteed with reliability and accuracy across time. Specifically, it may be said that the data should be recorded as the original and that the original recorded data should be sent at the time of retrieval. Therefore, it is necessary to maintain a controller to assert modifications that

are dynamically happening during the transmission, as changes might take place abruptly [9].

Unfortunately, massive amounts of data are very wasteful in terms of time usage and communication costs. In reality, collecting data sometimes may result in incomplete data, where some rows of the collected data have one or more missed values. Missing data occurs due to many reasons such as security attacks, connection errors, and sensor faults [10, 11]

Should customers need a third party to verify the information on behalf of him/her, data will be made available to the third party. Provable data possession (PDP) and proofs of retrievability (POR) are strategies that assist users in verifying the integrity of their data without requiring them to recover it [12, 13].

The CSP guarantees the integrity of the users' data by utilizing some measures such as virtual networks, firewalls, and other security regulations inside its perimeter. Practically, security is a key module of any cloud computing architecture. It is critical to make sure that only authorized users can access the data and that the data is protected. However, there are two important issues when outsourcing to a CSP is selected as the storage model [14], namely, (1) despite the owner has no control over the system, bad actors or even unauthorized users can access the crucial data by breaking into the system, and (2) in case the data is stored on the owner's property, the CSP might be the cause of the violation.

Cloud computing has become ubiquitous, people are using different cloud products and solutions to perform all their computing demands. However, there are too many concerns about the cloud security. People are suspicious about the existence of their work, files, and personal information on the cloud. One of the main concerns is that those people need guarantees that the data that are stored in the cloud are not susceptible to tampering by the CSP or a bad actor. In other words, people need proof of the integrity of their data. In this context, CSPs provide some measures to ensure the integrity of the data that are kept on the cloud.

In this paper, we'll look at some of the most recent research on the integrity of the data that is stored on the CSP premise and analyze it. Finally presents a big picture for this research issue, as well as a glimpse into the future.

The rest of this paper is organized as follows: in Sect. 2 we introduce some work that is related to cloud computing and security in cloud computing. In Sect. 3, we give a brief introduction to cloud computing, service models, and deployment models. Then, we discuss some of the security concerns regarding cloud computing in Sect. 4. The requirements for a secure cloud computing environment are discussed in Sect. 5. Some integrity-proofing techniques in cloud computing are discussed in Sect. 6. Finally, we draw some conclusions and recommend some future work in Sect. 7.

2. RELATED WORK

To make data secure, cloud computing must be developed carefully and with consideration for every area of security. Data integrity is a hot topic in cloud computing and presents an excellent research opportunity.

Attempting to provide a complete picture, Liu et al. [15] analyzed and provided authenticator-based verification of data integrity methodologies on cloud and Internet of Things (IoT) data. They tackled the research from several perspectives. Initially, they defined the research problem, then, they briefly described the study motives and methods. Afterward, they listed and compared some of the representative methodologies' most recent triumphs. Finally, they gave their viewpoint on likely future developments.

To guarantee the computation of dataflow integrity in cloud infrastructures, Du et al. [16] introduced a configurable runtime integrity attestation system that is referred to as RunTest. RunTest was presented to offer simple application-level attestation mechanisms to quickly determine the service of fraudulent providers and validate the legitimacy of computational effects. RunTest was created in the IBM System S data-flow system processing, and it was tested in the NCSU virtual computer lab. The tests demonstrated that the strategy was effective, and it had low performance when handling data flow in the cloud environment.

Many factors were established by the classification approach that was proposed by Shaikh and Sasikumar [17]. The parameters were specified in terms of a variety of dimensions and were provided based on the desired level of security. The storage is used to implement the relevant security measures based on the classification of the data set according to its dimensions. The effectiveness of the proposed categorization technique was evaluated using the sample dataset acquired.

The present worries and obstacles to cloud computing were anticipated by Puthal et al. [18]. The study consisted of three parts. Firstly, the cloud computing architectural style and the services that are offered are discussed. Secondly, based on the cloud computing service layer, the study identified a

variety of security issues. Finally, they highlighted several outstanding problems in light of the proliferation of cloud computing and its long-term effects. They concluded by highlighting the platforms that were accessible for cloud research and development.

In 2019, Mahmood et al. [19] created a novel cloud technique to enhance data security and address data integrity while guaranteeing dependability and a secure cloud storage service. A hidden image is encoded using the Advanced Encryption Standard (AES). The encrypted image is then placed into the host image using a steganography technique that combines the Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD) to produce the stage image. To ensure data integrity, a Secure Hash Algorithm 2 (SHA- 2) hash value was generated for the stego image before uploading it to the cloud. Once the image was downloaded from the cloud, the same procedure was utilized to generate the image's hash value. The secret image was then acquired by comparing the two hash values and checking to see if the cloud data had changed. The proposed scheme was demonstrated to be extremely client-friendly and secure through a thorough analysis of security and performance.

Garg et al. [20] suggested that the data integrity mechanism relies on the characteristics of bilinear pairings and employed Boneh Lynn Shacham (BLS) signatures for file block signatures. The results of the trials showed that the suggested protocol effectively lowered the calculation costs that were incurred by a client during the system startup phase, making it suitable for clients using devices like mobile phones, PDAs, etc. [21]

Ghallab et al. [22] examined eight different models to compare the cloud data integrity and security models. It emphasized virtually anthers for some of the current cloud security difficulties and challenges by making the fundamental methods of privacy-preserving audits public, including access control, attribute-based public key encryption, and access control. The pre-existing data integrity and security models, techniques, and strategies that have been applied in the literature on distributed cloud security were also allocated by the writers. Further study in the area of cloud security was recommended to address numerous security and data integrity issues.

Yang, et al. [23] investigated the problem of outsourced data integrity verification with verifiable data updates in a cloud computing context. They presented an effective integrity auditing technique for outsourced data using MSHT in the interim, which could also achieve block-based dynamic data updates. They used verifiable data replacement updates to overcome the issue of effective data integrity auditing. Yet, in some particular situations, renters wished to add a few fresh data blocks and remove a few out-of-date ones in addition to updating the data and auditing its integrity.

3. CLOUD COMPUTING

In this section, we summarize the characteristics of cloud computing in five main points. Firstly, cloud computing is an on-demand, self-service computing paradigm [24, 25], which means that a user can get the required resources without the need for human interaction or communication with the CSP.

Secondly, cloud computing offers broad network access, which enables devices with thin or thick clients, such as smartphones, laptops, and desktop computers, to access resources from any location using a standard mechanism [26].

Resource pooling is another important feature of cloud computing. Practically, resources are merged to permit sharing by several tenants. In a typical multi-tenant system, consumers are dynamically allocated resources, which can be used by them before being dynamically allocated to another consumer to satisfy the demand. Although resources are allocated and pooled to the consumers, they are unaware of the locations of those resources. Instead, they use the cloud infrastructure for their software applications [27].

Moreover, clients have no influence over the network, operating system, or storage, which together make up the physical environment of the cloud, even though they can run and use the software. The CSP is the only entity with the privilege to manage the underlying physical environment without the customer's participation. The user of this software can access it as a thin client through a web browser.

The last feature of cloud computing is the cloud service models [26]. Service models define how computing resources, such as networks, storage, and computation, are provisioned to the cloud users. There are three cloud service models as shown in Figure 1, namely, software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS).

3.1 Cloud Service Models

In SaaS, clients get access to software and cloud infrastructure so they can use it to execute their applications on the infrastructure. Clients can only run and consume the software in the cloud. Practically, clients do not influence the physical environment or the underlying infrastructure, such as the network, operating system, and storage. The

original physical environment is only under the control of the CSP. The client can access the software as a thin client through a web browser.

Like SaaS, the CSP maintains the infrastructure in PaaS. Unlike SaaS, clients can install and deploy applications that can be customized as per their preferences. The customization can be implemented using tools that are made available to the clients by the CSP. Using these tools, each user can configure his/her preferences by accessing the application settings.

The last service model, IaaS, allows the provisioning of computing resources such as networks, storage, and processing. In IaaS, clients can install the operating systems of their use. Furthermore, they can install and run their software. In other words, the client is provided with a machine and would have full control over that machine in a way that enables software installation/uninstallation, system configuration, etc.

Figure 1: The three service models of cloud computing.

3.2 Cloud Deployment Models

The National Institute of Standards and Technology (NIST) defined four deployment models for cloud computing [28]. The most prominent one is the public cloud model in which the cloud is located at the CSP's premises and is made available to the public via the Internet. The public model proved its scalability and stability, nevertheless, it presents some issues that might incur additional costs on the clients. Furthermore, clients are unaware of the CSP's storage technique, the location of data, and the company in which data is stored. It is noteworthy that decisions on the transition to the public cloud deployment model are tolerable for some security expenses.

The second deployment model is the private cloud model. The services of the private cloud are only provisioned to a single organization. The cloud measured can either be on-site or off-site. Private clouds address the same security concerns as public clouds when they are off-premises. Moreover, they add the overhead of running and managing the storage, controlling and monitoring the capacity, and services provisioning.

In the third deployment model, the community cloud, the cloud is controlled by the organizations or a third party, is located on or off premises, and services are only provided to a community of organizations with a shared interest. The disadvantage of this system is that there are many unresolved issues, such as contractor and security implications, service interruptions, and issues with data being spread across multiple companies and domains.

The hybrid cloud is the fourth deployment model which consists of two or more distinct cloud infrastructures, either private or public. Each cloud that the hybrid cloud comprises remains a separate and distinct entity, notwithstanding, they are connected by standardized or proprietary technology that facilitates data and application portability. The hybrid model has the control and protection of private clouds along with the accessibility and scalability of public clouds. Concerns about data security and confidentiality when transferring data from the public to the private setting or vice versa are some of the issues that confront the hybrid cloud since privacy standards in the public cloud differ significantly from those found in the private cloud.

3.3 Challenges in Cloud Computing:

Cloud computing has many advantages, and it shares a great contribution to the revolutionary computing model. It also plays an important role in digital transformation [29], yet it has some disadvantages. Some of the usage- and behavior-related security threats in the cloud include [30]:

- **Access:** When the data is unlawfully accessed, the potential for data modification emerges.
- **Availability:** The information must always be available. For clients not dealing with storage-related problems that result in client data loss.
- **Network load:** The system could malfunction because of the massive amount of data that is transferred back and forth between computers and servers on the cloud.
- **Integrity:** The most important factors that affect the cloud and heavily depend on the CSP are data accuracy, legality, and security.
- **Data location:** The customer is unaware of the precise location in which the data was kept or prioritized since it was dispersed over several locations, which is confusing. Assuring the client of the integrity of the data is one of the crucial issues in cloud computing that must be handled.

4. CLOUD SECURITY CONCERNS

Every layer must instantly trust the layer that is immediately behind it. Accordingly, the trust approach is a top-down trust model. Moreover, safe communication in the cloud must be enabled, which requires security guarantees at the operational, technical, procedural, and legal levels. However,

each service tier treats security as a separate issue. Therefore, trust could be viewed as a chain that extends from the application holder who trusts the provider through the end client [18]. In the following subsections, some security concerns of each service model are discussed.

4.1 Security Concerns Regarding SaaS

SaaS users must rely on the CSP to take the appropriate security measures to ensure their safety. On the other hand, CSPs must take the necessary precautions to prevent clients from reading each other's data. It is difficult for clients to ensure that the required safety steps are taken. Assurances that an application will be available and accessible upon request are difficult to guarantee from the client's side as well [31].

SaaS applications that are provisioned over the web can be good replacements for out-of-date software. Consequently, the main objective is to successfully transfer the required information while maintaining or improving the security benefits that are provided by the legacy application. The CSP either brings the application over the cloud or hosts it on its premises, i.e., the CSP's server farm [32].

A third-party vendor manages the framework on behalf of the client. By utilizing cloud computing and the pay-as-you-go payment method, the application administration provider can lower its interest in foundation benefits, freeing it up to focus on offering its clients superior services.

Over the last decade, the use of computers in organizations has increased exponentially [33], making computing and IT services commonplace. Nowadays, data and business processes, such as transactions, records, or pricing information, to name a few, are viewed as strategic assets that need to be safeguarded through access control and compliance regulations [34].

4.2 Security Concerns Regarding PaaS

By adopting PaaS, the CSP provides the client with considerable potential to develop applications directly on the cloud platform. The CSP guarantees that the data stays separate between the applications, but any security measures immediately below the application level, such as the assumption of a system disruption, will currently fall under the purview of the provider.

In essence, PaaS gives programmers the freedom to develop creative applications directly on the platform. Although it lacks functionality that is ready for use by customers, it is more easily expanded than SaaS. In terms of security, the stated capabilities of security skills are less complete, but there is more opportunity to create additional security, so there is a trade-off as well [35].

4.3 Security Concerns Regarding IaaS

The developer in IaaS has better security control if the virtualization director doesn't have any security issues. Nevertheless, there are numerous security issues even though virtual machines would be able to handle them in theory. The unalterable nature of the data that is stored on the supplier's equipment is another factor. Due to the information society's rapid virtualization. Maintaining total control over information will require a lot of expenditure. Few processes need to be paired to get the highest level of trust and security on a cloud asset [36].

5. CLOUD SECURITY REQUIREMENTS

The decision of whether to use a cloud system or a conventional one depends on how much the users trust the CSP and the services it offers. Assessing a provider's ability to address all the risks, such as those related to data security, virtual machines, and other legal and regulatory requirements, is the foundation of building trust [37].

Although the placement of cloud data has long been a topic of contention, no research has been done in this area. When a customer who is keeping sensitive data or planning to host applications in the cloud is uninformed of the original site, suitable place access control models for addressing such challenges have not yet been devised. Before adequate access control mechanisms for pass or multi domain Cloud are deployed, extensive research must be done [37].

Even though the works based on reputation systems, i.e., for CSP trust evaluation, and user trust evaluation are among the few in this field a comprehensive investigation and research in this area is expected to be completed shortly. Data or service compliance is another difficult issue with cloud computing. Because the security and privacy of the information handled by the CSP on behalf of the part of its business falls under their purview appropriate Service License Agreement (SLA) policy drafting and adherence to a particular territory by the CSP become imperative.

Another critical aspect of cloud security that must be thoroughly investigated is the cloud user's and CSP's lack of collaboration in uncovering and replying to security problems [38].

Trust in the CSP and its offerings is one of the most powerful motivators in a user's decision to move to a cloud solution. Trust is based on whether a CSP has considered all the risks, such as data security, virtual machine security, and other

government and compliance concerns. The three aspects that have been recognized for cloud system security assessment are the CIA triad.

The security needs of an existing cloud system under the CIA domain are the same standards that are commonly used to identify security issues of a traditional information system. To facilitate the understanding, mapping, and assessment of cloud-specific threats and the proposed solutions that are described in the following sections, a classification of security vulnerabilities has been proposed here [39].

5.1 Confidentiality

The responsibility to uphold the rights of people and organizations to privacy is represented by confidentiality. The information that is protected and whose uses are constrained by a duty to maintain confidentiality may be disclosed under certain circumstances by a person with authorization. Consequently, confidentiality reflects the obligation that researchers and other controllers and processors of personal data have to secure that personal information from unauthorized access and use, while privacy represents the right to restrict the sharing of one's personal information [40].

Confidentiality is looked at as the precise safeguard that prevents an asset from being revealed to unapproved users. In a cloud computing environment, those users could be clients who wish to acquire unlawful access to the data that the CSP has currently stored in the same dataset as their own. Also, the CSP may employ dishonest or curious staff members who could tamper with client-sensitive and important data. Furthermore, the virtual machine network, virtual machine image, and other components must adhere to strict confidentiality standards.

5.2 Integrity

The integrity of an asset makes sure that it has not been altered by individuals from a third party who do not have the authority to make changes. This characteristic guarantees the correctness and legality of an asset concerning its owner [41].

The integrity of an asset is frequently assumed to be impacted by addition, removal, and update operations. Because consumers access cloud-based services through web browsers, web-based attacks that change user data, database data, virtual machine data, or even Web Service Description Language (WSDL) files are frequent in cloud environments [40].

Practically, maintaining data integrity refers to keeping the data accurate, reliable, and consistent from creation to destruction. It is quite important to ensure data integrity for many reasons [42]. Firstly, integrity adds faith in the data. As long as businesses and organizations use data to guide decision-making, the decisions that are made using unreliable data may be inaccurate and may result in errors. Secondly, integrity ensures compliance, as many companies adhere to rules that demand data accuracy. There may be penalties and legal repercussions if these rules are broken.

Reputation is another reason for the importance of integrity. Data breaches usually harm organizations' reputations. By upholding data integrity, a firm can increase its reputation by fostering trust with clients and business partners. Moreover, efficiency adds to the importance of integrity. Correct data streamlines operations and increases efficiency by reducing the need for additional analysis and repairs. Last but not least, integrity is crucial to ensuring safety in a variety of sectors, such as healthcare. For the right treatment and diagnosis to be given, as well as to guarantee patient safety, accurate and trustworthy data is required.

Terabytes (TB) or even petabytes (PB) of data are needed for many data-centric activities that are handled by the cloud computing system. Because of this, it is crucial to address the data integrity problems associated with PaaS, SaaS, and data as a service (DaaS) appropriately. In practice, some data integrity issues are specific to the cloud [43]. First of these issues is the inability of the client to demonstrate that the CSP removed any legitimate data. Also, delivering the client incomplete data would be undetected.

Another issue is the attacks, and much specifically web-based attacks, such as SQL injection, which uses vulnerabilities in web servers to insert malicious software and change the elements of client data, or cross-site scripting (XSS), in which hackers introduce harmful scripts into dynamic websites so that the malicious code runs on the client's browser, giving them access to the customer's profile and jeopardizing the security of information and data.

Spoofing attacks might modify the web services description language (WSDL) content files to carry out particular actions for which might not have authorization, or they might attack the web service (WS) security policy by changing the WSDL to lower the security requirements.

5.3 Availability

One of the most crucial security components that a CSP must keep up with is availability. Different

commercial organizations that employ cloud-based services must guarantee the availability of their services. A small amount of downtime can result in large financial consequences that are irrecoverable. A typical SLA outlines the availability and response times that the CSP has committed to deliver. For instance, the SLA may state that resources would be accessible 99.9% of the time and that resources will be available on demand if more than 80 percent of any resource given is being utilized [44].

Data availability is the ability to obtain data when required. Overall, availability is critical to modern enterprises and organizations because it ensures operational continuity, disaster recovery, collaboration, innovation, compliance, and competitive advantage [45].

Virtualization lies at the heart of cloud computing. A physical server is divided into several virtual machines (VMs) and each VM functions as a separate computing environment [46]. Integrity management is essential in virtualized environments to guarantee that each VM is separated from the others and that the data and applications it runs are safe from hacking or other unauthorized access. This kind of integrity is referred to as virtualization integrity. Virtual machines are essential to modern IT infrastructures because they offer benefits including flexibility, scalability, improved disaster recovery, and cost-effective solutions for organizations to manage and deploy applications and services.

However, VMs can be relocated across data centers as needed to obtain more processing power or production capacity [47]. Yet, security precautions are required for VMs to function. A VM can move without a security policy, leaving it vulnerable to various attacks.

A breach in one VM could potentially affect the entire system, which implies that virtualization integrity is vital. To ensure the CIA of virtualized resources, virtualization integrity entails ensuring that (1) each VM is isolated and safe, (2) access control measures are put in place, and (3) security procedures are used [48].

Many steps are taken to ensure virtualization integrity, including, but not limited to, (1) network segmentation, (2) identity and access management, and (3) intrusion detection, prevention, and threat intelligence. Virtualized environments must also undergo regular audits, testing, and updates to maintain their security over time [48].

Virtual machines can be vulnerable in different ways, for example, an attacker might inject a specially-constructed malicious instance, or VM instance, into the cloud. When requests to that particular malicious VM are received, the CSP will redirect those requests to that malicious instance. As a result, the entire system is compromised [49].

It is crucial to assert the integrity of the services or VM instances and to know when they are compromised. It is worth mentioning that there exists a behavior in cloud computing that is referred to as VM rollback, which restores certain credibility problems in the VM [50]. Nevertheless, it is possible to reactivate security bugs (or wholes) that have already been patched or to regain access to passwords or account holders who had been taken offline by rolling back to earlier versions of the VM. Keeping VM snapshots is crucial for this purpose.

The CSP should be in charge of preserving the secrecy of the secured contents as well as the metadata servicing as the VM process live migration has already begun. The CSP must look at the VMs' life cycles and state changes. The states of VMs include defined, active, paused, and suspended [51]. There may also be unassigned VMs in storage without a state. Applying current security updates to VMs that have been stopped, turned off, or given no resources is crucial [52].

6. INTEGRITY-PROOFING TECHNIQUES

Data dignity in cloud computing is a hot research topic that is being extensively investigated by researchers. To ensure the trustworthiness and safe integrity of data in cloud computing, numerous studies are continuously being conducted [53]. Researchers have proposed a wide range of approaches to deal with problems with data integrity. The fundamental models of data integrity in the cloud are Provable Data Possession (PDP) and Proof of Retrievability (POR) [54]. The PDP model aims at ensuring the integrity of the client's data, that are kept on the CSP's side, and that the data has not been tampered with. On the other hand, POR aims to guarantee that the client can retrieve the requested data even if it is slightly corrupted [13].

6.1 Provable Data Possession (PDP)

With PDP, a client can verify whether a server still has the original data without having to retrieve it [55]. Accessing large files over the Internet is cumbersome as it involves many I/O operations to be executed by the server, as well as the long file transfer time that it will consume to send the file from the server to the client [56].

The PDP model uses a lightweight PDP protocol that is based on challenge/response [57]. The protocol comprises two phases, namely, (1) preprocess and store, and (2) verify server procession. In the first phase, shown in Figure 2, the client breaks the file F into n blocks, such that $F =$

$\{F_1, F_2, \cdots, F_n\}$. The file is then preprocessed by the client, which produces the preprocess file F', such that $F' = \{F'_1, F'_2, \cdots, F'_n\}$, as well as the metadata M. The preprocessed file F' is sent to be stored on the server side, the metadata M is kept on the client machine, and the client's copy of the file is deleted. Eventually, only one version of the file is kept on the server.

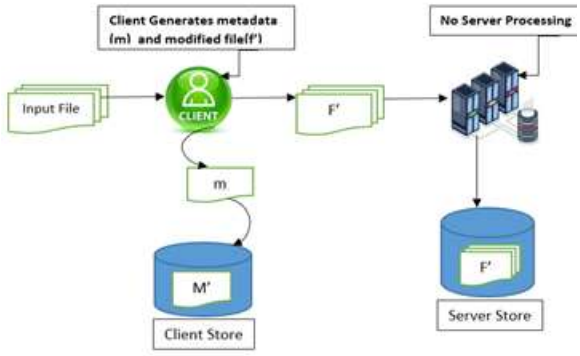

Figure 2. The preprocess and store phase of the PDP protocol

In the second phase, verify server possession, the metadata M will be used to verify that the server retains an intact version of the file. This phase starts when the client sends a random challenge R to the server to prove that it retains the file. When the challenge is received by the server, it computes a proof of possession P using some function, and sends the value of P to the client. The client verifies the response using the metadata M which is kept on its side. The verification server procession phase is shown in Figure 3.

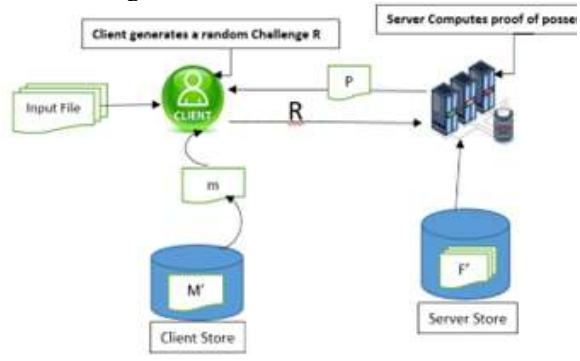

Figure 3: The verify server possession phase of the PDP protocol

Despite the many advantages that the PDP scheme has, it suffers a very important drawback that takes place when the client needs to update the contents of the file. The previous scheme is efficient and feasible when working on static data. In the case of the data with a changing nature, the client will need to repeat the first phase. The process becomes very time-consuming and loses flexibility when the rate of change of the data is relatively high [13]. Dynamic PDP [58] and Scalable PDP [59] were developed as improvements to the standard PDP to overcome its drawbacks.

6.2 Proofs of Retrievability

Similar to PDP, POR aims at providing proof from the server, which is referred to as the prover, to the client, which is referred to as the verifier, that the file F is intact without the need to send the file itself between the prover and the verifier, and thus providing less communication complexity. [60].

Like PDP, a challenge/response strategy is used in POR. The process starts with generating a secret key, which is used to encode the file and generate a file handle. The verifier issues a challenge using the secret key and the encrypted file handle. The challenge is given a value, say c, which is generated from the key and file handle as well as other parameters. The prover sends a response r to the verifier whose challenge is identified by the challenge value c. Finally, it is the responsibility of the verifier to accept whether or not r represents a valid response to c.

From the previous challenge/response mechanism, it is obvious that only the keys along with some values are communicated during the verify/prove operation, which illustrates the lower complexity of POR.

In general, integrity-proofing techniques are an open research area that needs much research effort to be done to fill the research gap in it. Moreover, the other two components of the CIA triad, namely, confidentiality and availability, are of no less importance than integrity. Not only do people have concerns about the integrity of their data, but they also need their data to be kept secret and not be exposed. The availability of data from wherever and whenever is also another important hot research topic that might urge researchers to more work in this field.

7. CONCLUSION AND FUTURE WORK

In this paper, we discussed cloud computing and some security concerns about cloud computing in light of the increasing demand for cloud computing, specifically cloud storage. We also discussed the alignment of those security concerns with the CIA trio, and what requirements need to be guaranteed by CSPs to build trust with clients. We also introduced two well-known techniques for proving the integrity of the data, or files, that are kept on the CSP's premise, namely PDP and POR. On the other hand, much work must be dedicated to identifying the risks

of keeping data on the cloud as well as cloud computing environments.

In future work, new PDP or POR techniques can be suggested to improve the trust between cloud users and CSPs. We can also study the incorporation of metaheuristic algorithms [61, 62, 63] in designing PDP and POR. Furthermore, we can investigate the use of machine learning [64, 65] and neural networks [66] in analyzing security threats in cloud computing.

REFERENCES

[1] A. Al-Jarrah, A. Al-Jarrah and A. Albsharat, "Dictionary Based Arabic Text Compression and Encryption Utilizing Two-Dimensional Random," International Arab Journal of Information Technology, vol. 19, no. 6, pp. 103-114, 2022.

[2] P.-C. Su and T.-C. Su, "Secure Blockchain-Based Electronic Voting Mechanism," International Arab Journal of Information Technology, vol. 20, no. 2, pp. 103-111, 2023.

[3] Q. Wang, C. Wang, J. Li, K. Ren and W. Lou, "Enabling public verifiability and data dynamics for storage security in cloud computing," in Computer Security – ESORICS 2009. ESORICS 2009. Lecture Notes in Computer Science, vol. 5789, M. Backes and P. Ning, Eds., Springer, Berlin, Heidelberg, 2009, p. 355–370.

[4] K. Hashizume, D. G. Rosado, E. Fernández-Medina and E. B. Fernandez, "An analysis of security issues for cloud computing," Journal of Internet Services and Applications, vol. 4, 2013.

[5] B. P. Kavin, S. Ganapathy, P. Suthanthiramani and A. Kannan, "13 - A modified digital signature algorithm to improve the biomedical image integrity in cloud environment," in Advances in Computational Techniques for Biomedical Image Analysis, D. Koundal and S. Gupta, Eds., Academic Press, 2020, pp. 253-271.

[6] M. Al-Khateeb, M. R. Al-Mousa, A. S. Al-Sherideh, D. Almajali, M. Asassfeha and H. Khafajeh, "Awareness model for minimizing the effects of social engineering attacks in web applications," International Journal of Data and Network Science, vol. 7, no. 2, pp. 791-800, 2023.

[7] N. Al-Milli and W. Almobaideen, "Hybrid Neural Network to Impute Missing Data for IoT Applications," in 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), Amman, Jordan, 2019.

[8] N. Al-Milli and B. H. Hammo, "A Convolutional Neural Network Model to Detect Illegitimate URLs," in 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2020.

[9] N. Al-Milli, A. Hudaib and N. Obeid, "Population Diversity Control of Genetic Algorithm Using a Novel Injection Method for Bankruptcy Prediction Problem," Mathematics, vol. 9, no. 8, 2021.

[10] Y. Li, Y. Yu, R. Chen, X. Du and M. Guizani, "IntegrityChain: Provable Data Possession for Decentralized Storage," IEEE Journal on Selected Areas in Communications, vol. 38, no. 6, pp. 1205-1217, 2020.

[11] M. R. Al-Mousa, "Generic Proactive IoT Cybercrime Evidence Analysis Model for Digital Forensics," in 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 2021.

[12] A. Sakthivel, "Enhancing Cloud Security Based On Group Signature," International Arab Journal of Information Technology, vol. 14, no. 6, pp. 923-929, 2017.

[13] I. Walker, C. Hewage and A. Jayal, "Provable Data Possession (PDP) and Proofs of Retrievability (POR) of Current Big User Data," SN Computer Science, vol. 3, 2022.

[14] H. A. Hussain and A. A. Yaseen, "OWARD TO BUILD STRONG IMAGE INTEGRITY SCHEME IN CLOUD COMPUTING ENVIRONMENT," Journal of Theoretical and Applied Information Technology, vol. 97, no. 2, pp. 398-408, 2019.

[15] C. Liu, C. Yang, X. Zhang and J. Chen, "External integrity verification for outsourced big data in cloud and IoT: A big picture," Future Generation Computer Systems, vol. 49, pp. 58-67, 2015.

[16] J. Du, W. Wei, X. Gu and T. Yu, "RunTest: Assuring Integrity of Dataflow Processing in Cloud Computing Infrastructures," in Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, Beijing, China, 2010.

[17] R. Shaikh and M. Sasikumar, "Data Classification for Achieving Security in Cloud Computing," Procedia Computer Science, vol. 45, pp. 493-498, 2015.

[18] D. Puthal, B. Sahoo, S. Mishra and S. Swain, "Cloud Computing Features, Issues, and Challenges: A Big Picture," in 2015 International Conference on Computational Intelligence and Networks, Odisha, India, 2015.

[19] G. S. Mahmood, D. J. Huang and B. A. Jaleel, "Achieving an Effective, Confidentiality and Integrity of Data in Cloud Computing," International Journal of Network Security, vol. 21, no. 2, pp. 326-332, 2019.

[20] N. Garg, S. Bawa and N. Kumar, "An efficient data integrity auditing protocol for cloud computing," Future Generation Computer Systems, vol. 109, pp. 306-316, 2020.

[21] A. Al-Shaikh, A. Shaheen, M. R. Al-Mousa, K. Alqawasmi, A. S. A. Sherideh and H. Khattab, "A Comparative Study on the Performance of 64-bit ARM Processors," International Journal of Interactive Mobile Technologies, vol. 17, no. 13, pp. 94-113, 2023.

[22] A. Ghallab, M. H. Saif and A. Mohsen, "Data Integrity and Security in Distributed Cloud Computing—A Review," in Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and

Applications. Advances in Intelligent Systems and Computing, vol. 1245, V. K. Gunjan and J. M. Zurada, Eds., Springer, Singapore.
[23] C. Yang, B. Song, Y. Ding, J. Ou, C. Fan and J. Xia, "Efficient Data Integrity Auditing Supporting Provable Data Update for Secure Cloud Storage," Wireless Communications & Mobile Computing, vol. 2022, 2022.
[24] A. Al-Shaikh, H. Khattab, A. Sharieh and A. Sleit, "Resource Utilization in Cloud Computing as an Optimization Problem," International Journal of Advanced Computer Science and Applications, vol. 7, no. 6, pp. 336-342, 2016.
[25] M. R. Al-Mousa, Q. Al-Zaqebah, A. S. Al-Sherideh, M. Al-Ghanim, G. Samara, S. Al-Matarneh and M. Asassfeh, "Examining Digital Forensic Evidence for Android Applications," in 2022 International Arab Conference on Information Technology (ACIT), Abu Dhabi, United Arab Emirates, 2022.
[26] V. Lahoura, H. Singh, A. Aggarwal, B. Sharma, M. A. Mohammed, R. Damaševičius, S. Kadry and K. Cengiz, "Cloud Computing-Based Framework for Breast Cancer Diagnosis Using Extreme Learning Machine," Diagnostics, vol. 11, no. 2, 2021.
[27] C. M. Mohammed and S. R. M. Zeebaree, "Sufficient Comparison Among Cloud Computing Services: IaaS, PaaS, and SaaS: A Review," International Journal of Science and Business, vol. 5, no. 2, pp. 17-30, 2021.
[28] C. Thota, G. Manogaran, D. Lopez and R. Sundarasekar, "Architecture for big data storage in different cloud deployment models," in Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing, USA, Information Resources Management Association, 2021, pp. 178-208.
[29] A. Al-Shaikh, R. Al-Sayyed and A. Sleit, "A Case Study for Evaluating Facebook Pages with Respect to Arab Mainstream News Media," Jordanian Journal of Computers and Information Technology, vol. 3, no. 3, pp. 142-156, 2017.
[30] M. Jangjou and M. K. Sohrabi, "A Comprehensive Survey on Security Challenges in Different Network Layers in Cloud Computing," Archives of Computational Methods in Engineering, vol. 29, p. 3587–3608, 2022.
[31] M. Á. Díaz de León Guillén, V. Morales-Rocha, F. Martínez and L. Felipe, "A systematic review of security threats and countermeasures in SaaS," Journal of Computer Security, vol., vol. 28, no. 6, pp. 635-653, 2020.
[32] Anjana and A. Singh, "Security concerns and countermeasures in cloud computing: a qualitative analysis," International Journal of Information Technology, vol. 11, p. 683–690, 2019.
[33] A. Al-Shaikh and A. Sleit, "Evaluating IndexedDB performance on web browsers," in 2017 8th International Conference on Information Technology (ICIT), Amman, Jordan , 2017.
[34] N. K. Sehgal and P. C. P. Bhatt, Cloud Computing, Springer, Cham, 2018.
[35] W. Isharufe, F. Jaafar and S. Butakov, "Study of Security Issues in Platform-as-a-Service (PaaS) Cloud Model," in 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Istanbul, Turkey, 2020.
[36] F. K. Parast, C. Sindhav, S. Nikam, H. I. Yekta, K. B. Kent and S. Hakak, "Cloud computing security: A survey of service-based models," Computers & Security, vol. 114, 2022.
[37] M. Choi, A. E. Azzaoui, S. K. Singh, M. M. Salim, S. R. Jeremiah and a. J. H. Park, "The Future of Metaverse: Security Issues, Requirements, and Solutions," Human-centric Computing and Information Sciences, vol. 12, 2022.
[38] P. Mishra, E. S. Pilli and R. Joshi, Cloud Security: Attacks, Techniques,Tools, and Challenges, CRC Press, 2021.
[39] H. A. COLOSI, C. COSTACHE and I. A. COLOSI, "Informational privacy, confidentiality and data security in research involving human subjects," Applied Medical Informatics, vol. 41, 2019.
[40] H. Alosert, J. Savery, J. Rheaume, M. Cheeks, R. Turner, C. Spencer, S. S. Farid and S. Goldrick, "Data integrity within the biopharmaceutical sector in the era of industry 4.0," Biotechnology Journal, vol. 17, 2022.
[41] V. Yesin, M. Karpinski, M. Yesina, V. Vilihura and K. Warwas, "Ensuring Data Integrity in Databases with the Universal Basis of Relations," Applied Sciences, vol. 11, 2021.
[42] B. Corcoran, M. Tan, X. Xu, A. Boes, J. Wu, T. G. Nguyen, S. T. Chu, B. E. Little, R. Morandotti, A. Mitchell and D. J. Moss, "Ultra-dense optical data transmission over standard fibre with a single chip source," Nature Communications, vol. 11, 2020.
[43] J. B. Hong, A. Nhlabatsi, D. S. Kim, A. Hussein, N. Fetais and K. M. Khan, "Systematic identification of threats in the cloud: A survey," Computer Networks, vol. 150, pp. 46-69, 2019.
[44] R. Girão-Silva, L. Martins, T. Gomes, A. Alashaikh and D. Tipper, "Improving Network Availability—A Design Perspective," in Third International Congress on Information and Communication Technology. Advances in Intelligent Systems and Computing, vol. 797, X. Yang, S. Sherratt, N. Dey and A. Joshi, Eds., Springer, Singapore, 2019, p. 799–815.
[45] I. M, M. Kaur, M. Raj, S. R and H.-N. Lee, "Cross Channel Scripting and Code Injection Attacks on Web and Cloud-Based Applications: A Comprehensive Review," Sensors, vol. 22, 2022.
[46] Y. Alghofaili, A. Albattah, N. Alrajeh, M. A. Rassam and B. A. S. Al-rimy, "Secure Cloud Infrastructure: A Survey on Issues, Current Solutions, and Open Challenges," Applied Sciences, vol. 11, no. 19, 2021.
[47] P. Ranaweera, A. D. Jurcut and M. Liyanage, "Survey on Multi-Access Edge Computing Security and Privacy," IEEE Communications Surveys & Tutorials, vol. 23, no. 2, pp. 1078-1124, 2021.
[48] I. van de Poel, "Core Values and Value Conflicts in Cybersecurity: Beyond Privacy Versus Security," in

The Ethics of Cybersecurity, M. Christen, B. Gordijn and M. Loi, Eds., Springer, Cham, 2020, p. 45–71.

[49] A. Piplai, P. Ranade, A. Kotal, S. Mittal, S. N. Narayanan and A. Joshi, "Using Knowledge Graphs and Reinforcement Learning for Malware Analysis," in 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020.

[50] H. Tang, S. Feng, X. Zhao and Y. Jin, "VirtAV: An agentless antivirus system based on in-memory signature scanning for virtual machine," in 2016 18th International Conference on Advanced Communication Technology (ICACT), 2016 18th International Conference on Advanced Communication Technology (ICACT), 2016.

[51] E. Farr, R. Harper, L. Spainhower and J. Xenidis, "A Case for High Availability in a Virtualized Environment (HAVEN)," in 2008 Third International Conference on Availability, Reliability and Security, Barcelona, Spain, 2008.

[52] K. C. Apostolakis, G. Margetis, C. Stephanidis, J.-M. Duquerrois, L. Drouglazet, A. Lallet, S. Delmas, L. Cordeiro, A. Gomes, M. Amor, A. D. Zayas, C. Verikoukis, K. Ramantas and I. Mark, "Cloud-Native 5G Infrastructure and Network Applications (NetApps) for Public Protection and Disaster Relief: The 5G-EPICENTRE Project," in 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Porto, Portugal, 2021.

[53] G. Cheng, Y. Lin, J. Yan, J. Zhao and L. Bai, "Model-Measurement Data Integrity Attacks," IEEE Transactions on Smart Grid.

[54] N. Garg, S. Bawa and N. Kumar, "An efficient data integrity auditing protocol for cloud computing," Future Generation Computer Systems, vol. 109, pp. 306-316, 2020.

[55] Y. Zhu, H. Hu, G.-J. Ahn and M. Yu, "Cooperative Provable Data Possession for Integrity Verification in Multicloud Storage," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 12, pp. 2231-224, 2012.

[56] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson and D. Song, "Provable data possession at untrusted stores," in CCS '07: Proceedings of the 14th ACM conference on Computer and communications security, Alexandria, Virginia, USA, 2007.

[57] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson and D. Song, "Remote data checking using provable data possession," ACM Transactions on Information and System Security, vol. 14, no. 1, p. 1–34, 2011.

[58] C. C. Erway, A. Küpçü, C. Papamanthou and R. Tamassia, "Dynamic Provable Data Possession," ACM Transactions on Information and System Security, vol. 14, no. 4, 2015.

[59] G. Ateniese, R. Di Pietro, L. V. Mancini and G. Tsudik, "Scalable and Efficient Provable Data Possession," in Securecomm08: Fourth International Conference On Security on Privacy for communication Networks, Istanbul, Turkey, 2008.

[60] K. D. Bowers, A. Juels and A. Oprea, "Proofs of retrievability: theory and implementation," in CCSW '09: Proceedings of the 2009 ACM workshop on Cloud computing security, Chicago, Illinois, USA, 2009.

[61] A. Al-Shaikh, B. A. Mahafzah and M. Alshraideh, "Metaheuristic Approach using Grey Wolf Optimizer for Finding Strongly Connected Components in Digraphs," Journal of Theoretical and Applied Information Technology, vol. 97, no. 16, pp. 4439-4452, 2019.

[62] A. Al-Shaikh, H. Khattab and S. Al-Sharaeh, "Performance Comparison of LEACH and LEACH-C Protocols in Wireless Sensor Networks," Journal of ICT Research and Applications, vol. 12, no. 3, pp. 219-236, 2018.

[63] A. Al-Shaikh, B. A. Mahafzah and M. Alshraideh, "Hybrid harmony search algorithm for social network contact tracing of COVID-19," Soft Computing, 2021.

[64] O. AlShorman, M. Masadeh, M. B. B. Heyat, F. Akhtar, H. Almahasneh, G. M. Ashraf and A. Alexiou, "Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection," Journal of Integrative Neuroscience, vol. 21, no. 1, 2022.

[65] R. Alazaidah, G. Samara, S. Almatarneh, M. Hassan, M. Aljaidi and H. Mansur, "Multi-Label Classification Based on Associations," Applied Sciences, vol. 13, no. 5, 2023.

[66] K. E. Alqawasmi and A. M. Alsmadi, "Estimation of ARMA Model Order Using Artificial Neural Networks," Circuits, Systems, and Signal Processing, 2023.