Goswami *et al. Journal of Cloud Computing* (2024) 13:45
https://doi.org/10.1186/s13677-024-00605-z

Journal of Cloud Computing:
Advances, Systems and Applications

REVIEW **Open Access**

Investigation on storage level data integrity strategies in cloud computing: classification, security obstructions, challenges and vulnerability

Paromita Goswami[1,2], Neetu Faujdar[2*], Somen Debnath[3], Ajoy Kumar Khan[1] and Ghanshyam Singh[4]

Abstract

Cloud computing provides outsourcing of computing services at a lower cost, making it a popular choice for many businesses. In recent years, cloud data storage has gained significant success, thanks to its advantages in maintenance, performance, support, cost, and reliability compared to traditional storage methods. However, despite the benefits of disaster recovery, scalability, and resource backup, some organizations still prefer traditional data storage over cloud storage due to concerns about data correctness and security. Data integrity is a critical issue in cloud computing, as data owners need to rely on third-party cloud storage providers to handle their data. To address this, researchers have been developing new algorithms for data integrity strategies in cloud storage to enhance security and ensure the accuracy of outsourced data. This article aims to highlight the security issues and possible attacks on cloud storage, as well as discussing the phases, characteristics, and classification of data integrity strategies. A comparative analysis of these strategies in the context of cloud storage is also presented. Furthermore, the overhead parameters of auditing system models in cloud computing are examined, considering the desired design goals. By understanding and addressing these factors, organizations can make informed decisions about their cloud storage solutions, taking into account both security and performance considerations.

Keywords Cloud computing, Data integrity, Security attacks, Cloud storage, Data auditing, Security challenges

Introduction

Cloud computing's appeal lies in its dynamic and flexible Service Level Agreement (SLA) based negotiable services, allowing users to access virtually limitless computing resources [1]. According to the National Institute of Standards and Technology (NIST), cloud computing offers a swiftly provisioned pay-per-use model, enabling on-demand, accessible, and configurable network access to shared pool resources, requiring minimal interactions from service providers and reduced management efforts [2]. Cloud computing models include private, public, hybrid, and community clouds, with services categorized into Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). IaaS providers like Google Compute Engine, Windows Azure Virtual

*Correspondence:
Neetu Faujdar
neetu.faujdar@gmail.com
[1] Department of Computer Engineering, Mizoram University, Aizawl, MZ 796004, India
[2] Department of Computer Engineering and Application, GLA University, Mathura, UP 281406, India
[3] Department of Computer science and Enginering, Tripura University, Agartala, Tripura 796022, India
[4] Centre for Smart Information and Communication Systems Department of Electrical and Electronic Engineering Science, University of Johannesburg, Auckland Park Campus, PO Box. 524, Johannesburg 2006, South Africa

Machines, and Amazon Elastic Cloud Compute offer network resources and computing storage, enhancing performance and reducing maintenance costs to meet specific customer demands [3, 4]. This evolution in cloud computing has transformed various sectors. Businesses and healthcare organizations benefit from services like cost reduction through resource outsourcing [3, 4], performance monitoring [5, 6], resource management [7], and computing prediction [8]. Additionally, cloud computing facilitates tasks such as resource allocation [9], workload distribution [10–12], capacity planning [13], and job-based resource distribution [14, 15]. This transformative impact underscores the significance of cloud computing in modern digital landscapes, empowering organizations with unprecedented efficiency and scalability in resource utilization [3–15].

Despite the availability of various data services, data owners are apprehensive about entrusting their valuable data to cloud service providers (CSPs) for third-party cloud storage due to concerns about the integrity of the CSPs [13, 16, 17], and the shared nature of cloud storage environments. Cloud computing primarily encompasses data storage and computation, with Infrastructure as a Service (IaaS) closely linked to cloud storage. When accessing IaaS, cloud users often lack visibility into the precise location of their outsourced data within the cloud storage and the machines responsible for processing tasks. Consequently, data privacy within cloud storage is a significant security challenge, exacerbated by the presence of malicious users, resulting in data integrity and confidentiality issues. This poses a critical security challenge for cloud storage, and trust in remote cloud data storage is crucial for the success of cloud computing. Data integrity, encompassing completeness, correctness, and consistency, is vital in the context of Database Management Systems (DBMS) and the ACID (Atomicity, Consistency, Isolation, Durability) properties of transactions. The issue arises when CSPs cannot securely guarantee clients the accuracy and completeness of data in response to their queries [18].

Researchers are actively advancing the field of data integrity in cloud computing by refining data integrity verification techniques and bolstering data privacy-preserving methods. These verification techniques primarily encompass Proof of Work (PoW), Proof of Data Possession (PDP), and Proof of Retrievability (PoR). Notably, the introduction of Message Authentication Code (MAC) using a unique random key within the data integrity framework marked a deterministic approach to data integrity verification, mitigating the inefficiencies associated with remote data integrity schemes that employed RSA-based encryption. This approach addressed issues related to significant computation time and long hash value transfer times for large files [19]. To enhance the security of data integrity schemes, Provable Data Possession (PDP) concepts were introduced to establish the legitimacy of data possession by a cloud server. Various subsequent research efforts have continually refined these algorithms, introducing innovations like the Transparent PDP scheme [20], DHT-PDP [21], Certificateless PDP Protocol for Multiple Copies [22–24], and Dynamic Multiple-Replica PDP [25]. Concurrently, the Proof of Retrievability (PoR) concept was introduced in 2007 to address error localization and data recovery issues [26]. Additionally, Proof of Original Ownership (PoW) emerged in 2011 through the Merkle hash tree protocol to prevent malicious adversaries, leading to a plethora of subsequent research endeavors with diverse improved algorithms aimed at the same goals [27–29].

Fully homomorphic encryption (FHE) was proposed to maintain the privacy preservation of outsourced data and in that case, original data were converted into ciphertext through an encryption technique that supports multiplication and additional operation over the ciphertext [30]. Meanwhile, drawbacks in [22] such as practically infeasible due to complex operations, were then solved by [31] Somewhat Homomorphic Encryption (SHE) scheme. Many more research works have been established in these few years such as biometrics face recognition approach [32], privacy-preserving auditing scheme for Cloud Storage using HLA [33], An Etiquette Approach for Preserving Data [34], etc.

Recently, Google cloud has introduced Zebra technologies based on a security command center (SCC) and security operation center (SOC) to point out some harmful threats such as crypto mining activity, data exfiltration, potential malware infections, brute force SSH attacks, etc. to maintain data integrity of business organization's information [35].

In recent years, numerous cloud data integrity schemes have emerged, along with several survey papers, albeit with limited parameters to comprehensively address specific aspects of data integrity. Some of these surveys include data auditing from single copies to multiple replicas [36], Proof of Retrievability [37], various data integrity techniques and verification types for cloud storage, and different data integrity protocols [38]. However, these surveys often fall short in providing a comprehensive understanding of data integrity strategies and their classification. A concise taxonomy of data integrity schemes was presented in a survey paper [39], which discussed a comparative analysis of existing data integrity schemes, their evolution from 2007 to 2015, and covered fewer physical storage issues, fewer security challenges, and design considerations. This survey paper aims to address this gap by offering an in-depth discussion on the security challenges within physical cloud storage, potential threats, attacks, and their mitigations. It will also categorize data integrity schemes, outline their phases and characteristics, provide a comparative analysis, and project future trends. This comprehensive approach underscores the significance of data integrity schemes in securing cloud storage.

Discussion

Although there are several articles arise on similar issues, our research work differs from all mentioned research works in the following ways: Unlike [36, 37, 39], our research work focused on different types of storage-based attacks and also comprised up-to-date methods to resist storage-based attacks which always violate data integrity schemes on physical cloud storage. Like [37], it includes storage-based security issues, threats, and it's existing mitigation solutions. Unlike [36, 37, 39] our research work focused on the different types of proposals of data integrity verification which is broadly classified into file-level verification, entire blocks verification, metadata verification, and randomly block-level verification.

Unlike [37], our survey work is not constricted to only proof of retrievability (POR). It covers all verification types like the power of ownership (PoW), proof of retrievability (POR), and provable data possession (PDP). It also includes different types of auditing verifications techniques to elaborate job roles on the TPA's side and DO's side. It also includes a discussion of the benefit of public auditing to reduce the overhead of computational and communication overhead of DO. Unlike [36–38, 40–43], our survey work reviews a wide range of quality features of data integrity schemes that have individually prime importance in cloud storage security. Unlike [36, 37, 41], we focused on different types of security challenges according to existing symptoms, effects, and probable solutions of data integrity schemes. Like [42–44], we include a discussion about malicious insider attacks, forgery attacks, and dishonest TPA and CSP. Unlike [41, 43, 44], in comparative analysis, we introduce here different performance analysis parameters of existing works based on the work's motivations and limitations in addition to a discussion of public and private data auditing criteria. Like [32], we include all existing data integration methods briefly in the Comparative analysis of data integrity strategies section.

Research gap

According to the above discussion, this research focuses on the following points to summarize the research gaps:

- In contrast to [36, 37, 39], our research included current strategies to fend against storage-based attacks, which consistently compromise data integrity techniques on physical cloud storage.
- Our research, in contrast to [36, 37, 39], concentrated on the various approaches to data integrity verification, which is categorised into four categories: file-level verification, full block verification, metadata verification, and randomized block-level verification.
- Our survey study is not limited to proof of retrievability (POR), in contrast to [37]. It includes all forms of verification, including proven data possession (PDP), proof of retrievability (POR), and power of ownership (PoW). Different Key Management Techniques used in cloud storage to improve security at cloud storage were also added here .
- In contrast to [36–38, 40–43], our survey work examines a variety of data integrity scheme quality features, each of which is crucial to the security of cloud storage.
- In contrast to [36, 37, 41], we concentrated on various security issues based on the impacts, symptoms, and likely fixes of data integrity techniques.
- In contrast to [41, 43, 44], we present here various performance analysis parameters of previous efforts based on the goals and constraints of the work together with a discussion of auditing criteria for both public and private data.

Contribution

On the basis of our knowledge, this is the first attempt to overlook all the related issues of cloud data storage with possible directions under a single article. The Key contributions of this research paper are summarized below:

- Identification of possible attacks on storage level services which may arise on physical cloud storage mitigating explored solutions
- Summarizing of possible characteristics of data integrity strategies to examine data integrity auditing soundness, phases, classification, etc. to understand and analyse security loopholes
- Literature review on comparative analysis based on all characteristics, motivation, limitation, accuracy, method, and probable attacks
- Discussion on design goal issues along with security level issues on data integrity strategy to analyse dynamic performance efficiency, different key management techniques to achieve security features, to analyse server attacks, etc.
- Identification of security issues in data integrity strategy and its mitigation solution
- Discussion about the future direction of new data integrity schemes of cloud computing.

This review article is described in 8 sections. Issues of physical cloud storage section, discusses issues of physical cloud storage, and attacks in storage level service. Key management techniques with regards to storage level in cloud section describes some existing key management techniques to enhance security of cloud storage. Potential attacks in storage level service section describes possible potential attacks in cloud storage. Phases of data integrity technique section phases of the data integrity scheme and summarizes all possible characteristics of the data integrity strategy. Classification of data integrity strategy section describes a classification of data integrity strategy. Characteristics of data integrity technique section describes characteristics of data integrity technique. Challenges of data integrity technique in cloud environment section describes Challenges of data integrity technique in cloud Environment. Desire design challenges of data integrity strategy section describes Desire design challenges of data integrity strategy. Comparative analysis of data integrity strategies section represents a comparative analysis of existing research works of data integrity strategy. At the end,design goal issues and future trends of cloud storage based on existing integrity schemes using a timeline infographic from 2016 to 2022 in Future trends in data integrity approaches section.

Issues of physical cloud storage

Generally, the physical cloud storage in terms of IaaS services gives cloud users the opportunity of using computing resources at a minimum cost without taking any responsibility for infrastructure maintenance. But in the actual scenario, CSP and other authorized users have no trusted actors in cloud computing. Hence, cloud storage is an attack-prone area due to the malicious intentions of CSP and insider-outsider attackers. We have listed here cloud storage issues along with possible attacks. Table 1 shows below all possible mitigating solutions.

- In capability of CSP: Managing big cloud storage may create a data loss problem for CSP due to lack of insufficient computational capacity, sometimes cannot meet user's requirement, missing a user-friendly data serialization standard with easily readable and editable syntax, due to changes of a life cycle in a cloud environment [66].
- Loses control of cloud data over a distributed cloud environment may give vulnerable chances to unauthorized users to manipulate valuable data of valid one [67].
- Lack of Scalability of physical cloud storage: Scalability means all hardware resources are merged to provide more resources to the distributed cloud system. It might be beneficial for illegitimate access and modify cloud storage and physical data centers [68].
- Unfair resource allocation strategy: Generally, monitoring data is stored in a shared pool in a public cloud environment which might not be preferable to cloud users who are not interested to leave any footprint on their work distribution/data transmission by a public cloud-hosted software component which will be the reason for a future mediocre of original data fetching [69].
- Lack of performance monitoring of cloud storage: Generally, monitoring data is stored in a shared pool in a public cloud which might not be preferable to cloud users who are not interested to leave any footprint on their work distribution/data transmission by a public cloud-hosted software component [70].
- Data threat: Cloud users store sensitive data in cloud environments about their personal information or business information. Due to the lack of data threat prevention techniques of cloud service providers, data may be lost or damaged [64, 71].
- Malicious cloud storage provider: Lack of transparency and access control policies are basic parameters of a cloud service provider being a malicious storage provider. Due to the missing of these two parameters, it's quite easy to disclose confidential data of cloud users towards others for business profit [72].
- Data Pooling: Resource pooling is an important aspect of cloud computing. Due to this aspect, data recovery policies and data confidentiality schemes are broken [73].
- Data lock-in: Every cloud storage provider does not have a standard format to store data. Therefore, cloud users face a binding problem to switch data from one provider to another due to dynamic changes in resource requirements [39].
- Security against internal and external malicious attack: Data might be lost or data can be modified by insider or outsider attacks [49, 74–76].

Key management techniques with regards to storage level in cloud

In order to prevent data leakage and increase the difficulty of attack, this paper presents a method combining data distribution and data encryption to improve data storage security. We have listed here some key techniques used in cloud storage to enhance security and transparency between cloud storage, cloud users.

- Hierarchical Key Technique: Some research articles [77] provide secret sharing and key hierarchy derivation technique in combination with user password to enhance key security, protecting the key and preventing the attacker from using the key to recover the data.

Table 1 Potential Types of Vulnerable Attacks and Threats at Storage Level Data Integrity with Mitigating Solutions

Potential Attacks	Storage Issues	Threats	Mitigation Solution with references	Applied Methods
DoS	No prediction format to formulate required time/storage to store/process data into cloud storage, data threat	Vulnerable service takes place instead of original service	Proposed authentication & authorization protocol [45, 46] Proposed signature-based scheme [47] Proposed intrusion detection/prevention scheme [48–50]	Kerberos protocol Attribute-Based Proxy Signature Improved Dynamic Immune Algorithm(IDIA)
Phishing	Lack of storage monitoring, Unaccredited access to physical cloud storage	Data confidentiality disclose	Propose phishing detection technique [51]	a hybrid classifier approach and hyper-parameter classifier tuning
Brute Force Attack / online dictionary Attack	Unaccredited access to physical cloud storage	Data confidentiality disclose, Violation of Data Authenticity	Propose data obfuscation scheme [52]	Least Significant Bit(LSB) substitution method
MITC Attack	Improper security against internal and external malicious attacks	Abnormality in service availability	Propose string authentication technique [53]	Chaotic maps and fuzzy extractors
Port Scanning	Improper security internal and external malicious attack	Abnormality in service availability	firewall policies [54]	Distributed firewalls/Controllers
Identity Theft	Unaccredited access to physical cloud storage, Untrusty cloud storage, data threat	SLA violation, security policies violation	Password based authentication scheme [55–57], privacy beach prevention [58]	key-based semantic secure Bloom filter (KSSBF), compact password-authenticated key exchange protocol (CompactPAKE), OTP,Evolutionary System Model based Privacy Preserving-(EMPPC)
Risk Spoofing	Incapability of CSP's monitoring, Untrusty cloud storage, data lock-in	Lack of internal security, logging violation	Monitoring secure data policies [59]	Symmetric Searchable Encryption (SSE) or Attribute-Based Encryption (ABE)
Data Loss/Leakage	Incapability of CSP, continuous storage monitoring, lack of scalability	Malicious Insider, Malicious Cloud Storage Provider	Propose Data integrity technique [60–64]	Data encryption method, Public data auditing technique
Shared Technology issue	Unfair resource allocation strategy, no standard data storing format, Shared Technology Issue	VMs are become vulnerable due to loose control of a hypervisor	Virtual Machine Monitoring Scheme [65]	Xen, KVM

- Private Key Update Technique:This identity-based encryption technique [78] helps to update the private keys of the non-revoked group users instead of the authenticators of the revoked user when the authenticators are not updated, and it does away with the complex certificate administration found in standard PKI systems.
- Key Separation Technique: This cryptographic method aids in maintaining the privacy of shared sensitive data while offering consumers effective and efficient storage services [79].
- Attribute-based Encryption Key Technique: Instead of disclosing decryption keys, this method achieves the conventional notion of semantic security for data secrecy, whereas existing methods only do so by establishing a lesser security notion [80, 81]. It is used to share data with users in a confidential manner.
- Multiple Key Technique:This k-NN query-based method improves security by assisting the Data owner(DO) and each query user in maintaining separate keys and not sharing them [82]. In the meantime, the DO uses his own key to encrypt and decrypt data that has been outsourced.

Potential attacks in storage level service

Storage level service in cloud computing offers services of resource computation, virtual network, shared storage over the internet in lease. It provides more flexible and scalable benefits than on-premise physical hardware. Due to these two aspects of the cloud, storage-level services can be the victim of malicious attacks attempting to steal computing resources for the publication of original data or data exfiltration in data braces. If attackers can successfully enter into the infrastructure services of an organization, they can then grip those parts to obtain access to other important parts of the enterprise architecture causing security issues of data integrity. We have listed here possible attacks on storage-level services.

- DoS/DDoS: Ultimate purpose of this attack is to do unavailable original services towards users and overload the system by flooding spam results in a single cloud server. Due to the high workload, the performance of cloud servers slumps, and users lose the accessibility to their cloud services.
- Phishing: Attackers steal important information in the form of a user's credentials like name, password, etc. after redirecting the user to a fraud webpage as an original page.
- Brute Force attack/ Online dictionary attack: It's one type of cryptographic hack. Using an exhaustive key search engine, malicious attackers can violate the privacy policy of the data integrity scheme in cloud storage.
- MITC: Man in the cloud attack helps attackers to gain the capability to execute any code on a victim machine through installing their synchronization token on a victim's machine instead of the original synchronization token of a victim machine and using this token, attackers get control over target machine while target machine synchronizes this token to the attacker's machine.
- Port scanning: Attackers perform port scanning methods to identify open ports or exposed server locations, analyze the security level of storage and break into the target system.
- Identity theft: Using password recovery method, attackers can get account information of legitimate users which causes loss of credential information of the user's account.
- Risk spoofing: Resource workload balancing is a good managerial part of cloud storage but due to this aspect of cloud computing, attackers can steal credential data of cloud users, able to spread malware code in host machines and create internal security issues.
- Data loss/leakage: During data transmission time by external adversaries, incapability of cloud service providers, by unauthorized users of the same cloud environment, by internal malicious attackers, data can be lost or manipulated.
- Shared technology issue: Compromising hypervisors, cloud service providers can run concurrently multiple OS as guests on a host computer. For the feebleness of hypervisor, attackers create vulnerabilities like data loss, insider malicious attacks, outsider attacks, loss of control on machines, and service disruption by taking control over all virtual machines.

Phases of data integrity technique

Data integrity always keeps the promise of data consistency and accuracy of data at cloud storage. Its probabilistic nature and resistance capability of storing data from unauthorized access help cloud users to gain trust for outsourcing their data to remote clouds. It consists of mainly three actors in this scheme: Data owner (DO), Cloud Storage/Service Provider (CSP), and Third-Party Auditor(optional) [39] as depicted in Fig. 1. The data owner produces data before uploading it to any local cloud storage to acquire financial profit. CSP is a third-party organization offering Infrastructure as a service (IaaS) to cloud users. TPA exempts the burden of management of data of DO by checking the correctness and intactness of outsourced data. TPA also reduces communication overhead costs and the computational cost of the data owner [83, 84]. Sometimes, DO ownself takes

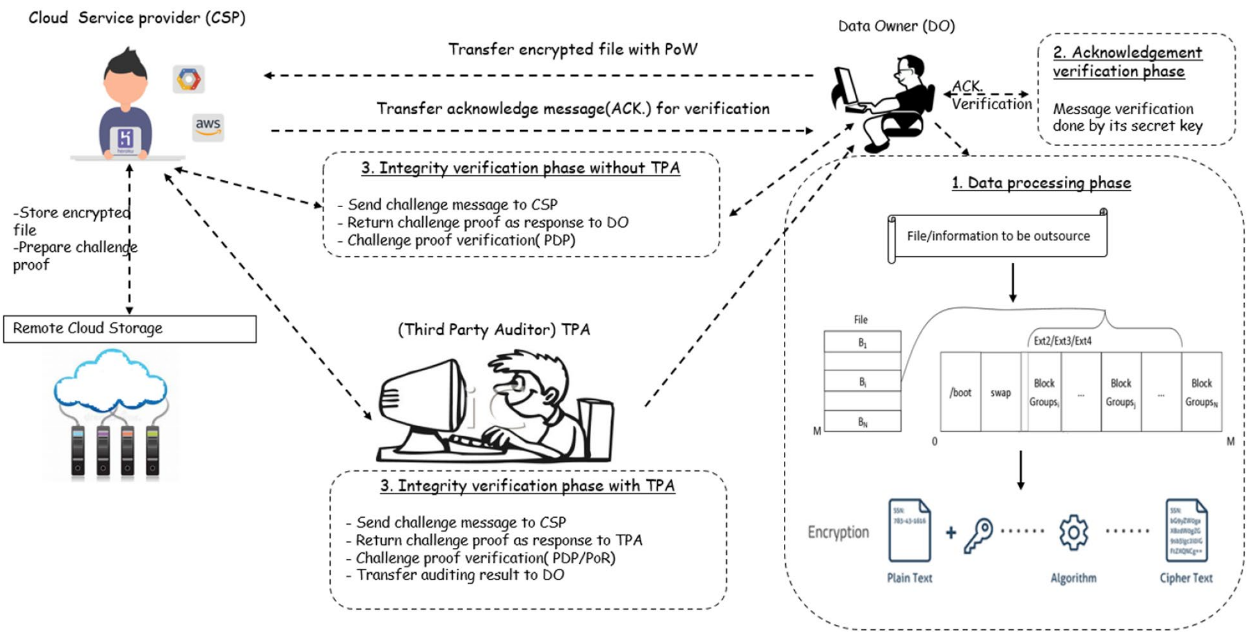
Cloud Service provider (CSP)
aws
Transfer encrypted file with PoW
Transfer acknowledge message(ACK.) for verification
Data Owner (DO)
ACK. Verification
2. Acknowledgement verification phase
Message verification done by its secret key
-Store encrypted file
-Prepare challenge proof
3. Integrity verification phase without TPA
- Send challenge message to CSP
- Return challenge proof as response to DO
- Challenge proof verification(PDP)
1. Data processing phase
File/information to be outsource
Remote Cloud Storage
(Third Party Auditor) TPA
File
B_1
B_i
B_N
M
Ext2/Ext3/Ext4
/boot
swap
Block Groups$_i$
...
Block Groups$_j$
...
Block Groups$_N$
0
M
3. Integrity verification phase with TPA
- Send challenge message to CSP
- Return challenge proof as response to TPA
- Challenge proof verification(PDP/PoR)
- Transfer auditing result to DO
Encryption
+
Plain Text
Algorithm
Cipher Text

Fig. 1 Entire Cycle of Data Integrity Technique

responsibility for data integrity verification without TPA interference. There are three phases in data integrity strategy described below in Table 2:

- Data processing phase: In data processing phase, data files are processed in many way like file is divided into blocks [60], applying encryption technique on blocks [90], generation of message digest [87], applying random masking number generation [88], key generation and applying signature on encrypted block [93] etc. and finally encrypted data or obfuscated data is outsourced to cloud storage.
- Acknowledgement Phase: This phase is totally optional but valuable because sometimes there may arise a situation where CSP might conceal the message of data loss or discard data accidentally to maintain their image [88]. But most of the research works skip this step to minimize computational overhead costs during acknowledgment verification time.
- Integrity verification phase: In this phase, DO/ TPA sends a challenge message to CSP and subsequently, CSP sends a response message as metadata or proof of information to TPA/DO for data integrity verification. The audit result is sent to DO if verification is done by TPA.

Classification of data integrity strategy

Classification of data integrity depends on a variety of conceptual parameters and sub-parameters. Table 3 shows all parameters, and sub-parameters with references to give a clear idea about data integrity strategy. The deployment setup of data integrity strategy is dependent on the environment of the proposed system. Clients can store their data in public cloud set up [98], multi-cloud setup [99, 100] or hybrid cloud set up [101]. If data are placed in a public cloud setup, clients lose access control visibility on data due to the outsider data management policy of CSP. As a result, data integrity problems arise because both CSP and public cloud storage are not honest in practical scenarios. Multi-cloud means more than one cloud service, more than one vendor in the same heterogeneous cloud architecture. A hybrid cloud is also a combination of private and public clouds. Hence, in the shared storage structure of multi and hybrid cloud environments, security issues of data integrity is a genuine concern. The guarantee of data integrity scheme can be proposed in two types: deterministic and probabilistic approaches. The performance of probabilistic verification is better than deterministic verification because of its higher accuracy in error correction of blocks without accessing the whole file and low computational overhead [102]. But, the deterministic approach gives adequate accuracy of data integrity whereas the probabilistic approach gives less than data integrity accuracy of deterministic approach [39].

a) Type of proposal

- File level verification: This is a deterministic verification approach. Here, data integrity verification is generally done by either TPA or the client. The client submitted an encoded file to the storage server

Table 2 Classified Phases of Data Integrity Schemes

Ref.	Technical Methods	Data Processing Phase			Acknowledgement Phase	Auditing Phase			
		Initial Phase	Key & Signature Generation Phase	Encryption		Using TPA	Using Data Owner/ Client	Challenge phase	Proof Verification Phase
[85]	Confidentiality Preserving Auditing	Yes	Yes	Yes	No	Yes	No	Yes	Yes
[60]	Ensuring of confidentiality and integrity data	Yes	No	Yes	No	Yes	No	No	No
[86]	Privacy preserving integrity checking model	Yes	Yes	Yes	No	Yes	No	Yes	Yes
[61]	Verifying Data Integrity	Yes	Yes	Yes	No	No	Yes	Yes	Yes
[87]	Data auditing mitigating with data privacy and data integrity	Yes	No	Yes	No	Yes	No	Yes	Yes
[88]	Public Verification of Data Integrity	Yes	Yes	Yes	No	Yes	No	Yes	Yes
[89]	Ternary Hash Tree Based Integrity Verification	Yes	Yes	Yes	No	Yes	No	Yes	Yes
[84]	Third-party auditing for cloud service providers	Yes	Yes	No	No	Yes	No	Yes	Yes
[90]	Identity-Based Integrity Auditing and Data Sharing	Yes	Yes	Yes	No	Yes	No	Yes	Yes
[91]	A Secure Data Dynamics and Public Auditing	Yes	No	Yes	No	Yes	No	Yes	Yes
[83]	Oruta: privacy-preserving public auditing	Yes	Yes	No	No	Yes	No	Yes	Yes
[92]	Dynamic Auditing Protocol	Yes	Yes	No	No	Yes	No	Yes	Yes
[93]	Dynamic Data Integrity Auditing Method	Yes	Yes	No	No	Yes	No	Yes	Yes
[94]	Algebraic Signatures-Based Data Integrity Auditing	Yes	Yes	No	Yes	Yes	No	Yes	Yes
[95]	Efficient public verification on the integrity	Yes	Yes	Yes	No	Yes	No	Yes	Yes
[96]	Attribute-Based Cloud Data Integrity Auditing	Yes	Yes	No	No	Yes	No	Yes	Yes

Table 2 (continued)

Ref.	Technical Methods	Data Processing Phase			Acknowledgement Phase	Auditing Phase			
		Initial Phase	Key & Signature Generation Phase	Encryption		Using TPA	Using Data Owner/ Client	Challenge phase	Proof Verification Phase
[78]	Efficient User Revocation in Identity-Based Cloud Storage Auditing	Yes	Yes	No	No	Yes	No	Yes	Yes
[97]	Secure and Efficient Data Integrity Verification Scheme	Yes	Yes	No	No	Yes	No	Yes	Yes

and for data integrity verification a verifier verified the encoded file through the challenge key and secret key which is chosen by the client [103].

- Block Level Verification : This type of verification is a deterministic verification approach. Firstly, a file is divided into blocks, encrypted, generated message digest, and sent encrypted blocks to CSP. Later, CSP sends a response message to TPA for verification and TPA verifies all blocks by comparing the newly generated message digest with the old message digest generated by the client [87].
- Randomly block level verification: This is a probabilistic verification approach. In this verification, a file is divided into blocks, next generate anyone signatures or combination of any two signatures of hash [86], BLS [88], HLA [124], random masking [88], or ZSS [97] for all blocks and submits both of them to cloud storage. Later, TPA generates a challenge message for randomly selected blocks which will be verified for data integration checking and sent to CSP. Next, CSP sends a proof message to TPA for verification. The proof message is verified by TPA for randomly selected blocks by generating new signatures and comparing old and new signatures of particular blocks [61, 86].
- Metadata verification: In this deterministic approach, firstly cloud users generate a secret key, and using this secret key, cloud users prepare metadata of the entire file through HMAC-MD5 authentication. Later, the encrypted file is sent to CSP, and metadata is sent to TPA. Later this metadata is used for integrity verification via TPA [85].

b) Category of data

- Static data: In static nature, no need to modify data that are stored in cloud storage. In [105], a basic RDPC scheme is proposed for the verification of static data integrity. In remote cloud data storage, all static files are of state-of-the-art nature which gets the main attention but in practical scenarios, TPA gets permission to possess the original data file creates security problems. In [106], the RSASS scheme is introduced for static data verification by applying a secure hash signature (SHA1) on file blocks.
- Dynamic Data: Data owners don't have any restriction policy for applying updation, insertion and deletion operations on outsourced data for unlimited time which is currently stored in remote cloud storage. In [111], a PDP scheme is introduced by assuming a ranked skipping list to hold up completely dynamic operation on data to overcome the problem of limited no. of insertion and query operation on data which is described in [118]. In [117], dynamic data graph is used to restrict conflict of the dynamic nature of big-sized graph data application.

c) Verification type

- Proof of ownership verification: The proof of ownership (PoW) scheme is introduced in the data integrity scheme to prove the actual data ownership of original data owner to server and to restrict unauthorized access to outsourced data of data owner from valid malicious users in the same cloud environment. PoW scheme is enclosed with data duplication scheme to reduce security issues about an illegal endeavor of a malicious user to access unauthorized data [27]. Three types of PoW scheme is defined: s-POW, s-Pow1, s-Pow2 in [29] which have satisfactory computation and I/O efficiency at user side but I/O burden on the remote cloud

Table 3 Taxonomy of applicable Data Integrity Phases

Type of Proposals				Category of data		Verification Types				
Deterministic			Probabilistic	Static Data	Dynamic Data	PDP	PoW	PoR	Auditing	
File level Verification	Entire Block Level Verification	Meta data Verification	Random Block level Verification						Public Auditing	Private Auditing
[89, 103]	[87, 104]	[85, 91, 96] [97, 107]	[61, 86, 88] [89, 90, 93] [94, 78, 117] [107, 114]	[105, 106]	[89, 92, 93] [94, 104, 111] [114, 118, 119] [117]	[89, 104, 107] [112–114] [120]	[88, 108]	[89, 103, 109] [108, 115] [121]	[88–90] [92–94] [78, 95, 97] [114, 122] [123]	[87, 104, 110] [116]

are significantly increased and this problem was overcome in [28] through establishing a balance between server and user side efficiency.

- Provable of data possession: Provable of data possession (PDP) scheme promises statically the exactness of data integrity verification of cloud data without downloading on untrusted cloud servers and restricts data leakage attacks at cloud storage. In [104], research work described aspects of the PDP technique from a variety of system design perspectives like computation efficiency, robust verification, lightweight and constant communication cost, etc. in related work. In [112], certificateless PDP is proposed for public cloud storage to address key escrow problems and key management of general public key cryptography and solve the security problems(verifiers were able to extract original data of users during integrity verification time) of [113, 120].
- Proof of retrievability verification: Proof of retrievability(PoR) ensures data intactness in remote cloud storage. Both PoR and PDP perform similar functions with the difference that PoR scheme has the ability to recover faulty outsourced data whereas PDP only supports data integrity and availability of data to clients [108]. In [109], IPOR scheme is introduced which ensures 100% retrieval probability of corrupted blocks of original data file. DIPOR scheme also supports data retrieval technique of partial health records along with data update operation [115].
- Auditing verification: Verification of cloud data which is outsourced by the data owner is known as the audit verification process. Data integrity scheme supports two types of verification: Private auditing verification(verification is done between CSP and data owner i.e. cloud user) and Public auditing verification (cloud user hiers a TPA to reduce computational and communication overhead at ownside and verification is done between CSP and TPA) [122]. Privacy-preserving public auditing [83, 122], certificateless public auditing [125],optimized public auditing scheme [123] ,bitcoin-based public auditing [88], S-audit public auditing scheme [108], shared data auditing [83], Dynamic data public auditing [126] Non-privacy preserving public auditing scheme [127], digital signature(BLS, hash table, RSA etc.) based public auditing scheme [88, 119, 128] etc. are some types of public auditing schemes. A private auditing scheme was first proposed by [110] called SW method and further reviewed by some research works[[87, 116].

Characteristics of data integrity technique

In this review article, focuses on several quality features of data integrity, which have individually prime importance in cloud storage security. These are:

- Public Auditability: The auditability scheme examines the accuracy of stored outsourced data from data owner at cloud storage by TPA according to the request of data owners [94, 95].
- Audit correctness: The proof message of CSP can pass the validation test of TPA only if CSP and TPA are being honest and CSP, data owner properly follow the pre-defined process of data storing [89, 78].
- Auditing soundness: The one and only way to pass TPA's verification test is that CSP has to store the data owner's entire outsourced data at cloud storage [90].
- Error localization at block level: It helps to find out error blocks of a file in cloud storage during verification time [89].
- Data Correctness: It helps to rectify error data block with available replica block's information in cloud storage [89].
- Stateless Auditor: During verification, a stateless auditor is not necessary to maintain, store or update previous results of verification for future usages [88, 95].
- Storage Correctness: CSP prepares a report which shows that all data is entirely stored in cloud storage even if the data are partially tempered or lost. Therefore, the system needs to guarantee data owners that their outsourced data are the same as what was previously stored [129].
- Robustness: In probabilistic data integrity strategy, errors in smaller size data should be identified and rectified [39].
- Unforgeability: Authenticated users can only generate a valid signature/metadata on shared data [129].
- Data Dynamic support: It allows data owners to insert, edit and delete data in the cloud storage by maintaining the constant level of integrity verification support like previous [89].
- Dependability: Data should be available during managing all the file blocks time [89].
- Replica Audibility: It helps to examine the replicas of the data file stored in the cloud storage by TPA on demand with data owners [89].
- Light Weight: It means that due to the occurrence of a large number of blocks and the presence of multiple users in the system, signature process time should be short to reduce the computational overhead of clients[88, 97].
- Auditing Correctness: It ensures that the response message from the CSP side can pass only the veri-

fication trial of TPA when CSP properly stores outsourced data perfectly into cloud storage [97].
- Privacy Protection: During verification, the auditing scheme should not expose a user's identity information in front of an adversary [90, 97].
- Efficient User Revocation: The repeal users are not able to upload any data to cloud storage and can not be authorized users any more [78].
- Batch Auditing: In the public auditing scheme, batch auditing method is proposed for doing multiple auditing tasks from different cloud users which TPA can instantly perform [95].
- Data Confidentiality: TPA can not acquire actual data during data integrity verification time [90].
- Boundless Verification: Data owners never give TPA any obligate condition about a fixed no. of verification interaction of data integrity [88].
- Efficiency: The size of test metadata and the test time on multi-owner's outsourced data in cloud computing are both individualistic with the number of data owners [95].
- Private Key Correctness: Private key can pass verification test of cloud user only if the Private key Generator (PKG) sends a right private key to the cloud user [90].
- Blockless Verification: TPA no need to download entire blocks from cloud storage for verification [95].

Challenges of data integrity technique in cloud environment

Security challenges of data integrity technique in cloud computing always come with some fundamental questions:

- how outsourced data will be safe in a remote server and how data will be protected from any loss, damage, or alteration in cloud storage?
- how security will assure cloud data if a malicious user is present inside the cloud?
- On which location of shared storage, outsourced data will be stored?
- Will legitimate access to the cloud data be by an authorized user only with complete audit verification availability?

All the above questions are associated with the term privacy preservation of data integrity scheme and that's why data integrity in cloud computing is a rapidly growing challenge still now. Refer Table 4, for existing solutions to security challenges and corresponding solutions of data integrity techniques.

a) **Risk to integrity of data**: This security is divided into three parts:

   - during globally acquiring time, cloud services are hampered by many malicious attacks if integrity of database, network etc. are properly maintained.
   - Data availability and integrity problems occur if unauthorized changes happened with data by CSP.
   - Segregation problem of data among cloud users in cloud storage is another problem of data integrity. Therefore, SLA-based patch management policy, standard validation technique against unauthorized use and adequate security parameters need to be included in data integrity technique [131].

Table 4 Security Challenges of Cloud Storage with its Solutions

Types of Security Issues	Symptoms	Affects	Solution with references
Risk to integrity of data	Unauthorized access, segregation problem of data, lack of maintenance of database	hamperness of cloud storage service, lack of data integrity	Data encryption method, Public data auditing technique [83, 88–90]
Dishonest TPA	tempering of original file, by the generation of wrong audit message, spoiling of CSP's intention	Lack of data confidentiality, lack of data integrity	hash function with collision resistant property [87], Secure Hash Algorithm (SHA-2) [87], RSA algorithm [91], Threat Model [83], panda public auditing(PPA) [130]
Dishonest CSP	Data leakage, data modification, loss of data	loss of reputation of CSP, data unavailability, lack of data integrity	Zero knowledge proof [117], data possession verification scheme [97], string authentication technique [53], firewall policies [54]
Forgery Attack	Forge audit message, forge proof message	Violate data integrity policy, lack of reputation of CSP	soundness criteria [88], one way hash function [83], metadata of storage block order [89], hardness of diffie-Hellman computation in bilinear group [117], pass the challenge proof with non negligible probability [90]
Malicious Insider Attack	Data leakage, data modification, data loss	Violate data integrity policy	string authentication technique [14], digital signature [88, 97, 124, 125]

b) **Dishonest TPA**: A dishonest TPA has two prime intentions:

- TPA can spoil the image of CSP by generating wrong integrity verification messages.
- TPA can exploit confidential information with the help of malicious attackers through repeated verification interaction messages with cloud storage.

  Hence, an audit message verification method must be included in a data integrity verification scheme to continuously analyze the intentional behavior of TPA

c) **Dishonest CSP**: An adversary CSP has three motives: i) CSP tries to retrieve either the original content of the entire data file or all block information of the data file and this leakage data information are used by CSP for business profit. ii) CSP can modify the actual content of a file and use it for personal reasons. But in both cases, the data owner can not detect the actual culprit. iii) CSP always tries to maintain its business reputation even if outsourced data of owner are partially tempered or lost Particularly, for that reason, an acknowledged verification method, an error data detection method and an error data recovery method should be included in data integrity scheme to maintain intactness of data and confidentiality of data [89, 132].

d) **Forgery Attack at Cloud Storage**: Outsider attacker may forge a proof message which is generated by CSP for the blocks indicated by challenge message to respond TPA. Malicious auditors may forge an audit-proof that passes the data integrity verification [88, 90].

e) **Data modification by an insider malicious user into cloud storage** : An insider malicious user can subvert or modify a data block at his/her will and can fool the auditor and data owner to trust that the data blocks are well maintained at the cloud storage even if that malicious user alters the interaction messages in the network channel. Hence data confidentiality scheme or obfuscation data technique should be included in data integrity technique [92].

Desire design challenges of data integrity strategy

Below are the main design issues for data integrity schemes:

a) **Communication overhead**: It means total outsourcing data, which is transferred from client to storage server, transfer of challenge message to CSP, transfer of the proof message to TPA, transfer of audit message to client all are communication overhead. Table 5 ,compares the communication overhead incurred during public auditing by DO, LCSP, and RTPA. Since DO always sends either their original file, an encrypted file, or an encrypted file with a signature to a cloud server, most articles here consider communication overhead for creating challenge messages and challenge-response messages, which is not included in DO's communication overhead.

b) **Computational overhead**: Data preprocessing, signature generation and audit message verification from data owner side or trusted agent side, challenge message generation, data integrity verification and audit message generation from the TPA side, prof message generation from CSP side all are computational overhead. In [97], the computational overhead of client, CSP and TPA are less than [124] because ZSS signature requires less overhead of power exponential and hash calculation than BLS signature. Table 6 compares the computational overhead incurred during public auditing by DO, LCSP,

Table 5 Comparison of Communication Overhead between DO, CSP and TPA During Auditing Phase

Ref.	Data Owner	Cloud Service Provider	Third Party Auditor
[88]	Not Considerable	$log2_c + 160$	$(s + 1)p$
[89]	Not Considerable	$2j\lvert k\rvert + \lvert r\rvert$	$2\lvert hash\rvert + 2j\lvert k\rvert + 360$
[90]	Not Considerable	$\lvert p\rvert + \lvert q\rvert$	$c.(\lvert n\rvert + \lvert p\rvert)$
[133]	Not Considerable	$log2_c + (c + 1)log2_p$	$(s + 1)log2_p$
[78]	Not Considerable	$n\lvert p\rvert + n\lvert q\rvert$	$(c + 1).\lvert q\rvert + \lvert p\rvert + c\lvert id\rvert$
[87]	Not Considerable	$j\lvert k\rvert$	$j\lvert Hash\rvert$
[105]	Not Considerable	$j\lvert k\rvert$	Not Applicable (Private Auditing)
[134]	Not Considerable	$c\lvert s\rvert + \lvert p\rvert$	$\lvert p\rvert + 2\lvert q\rvert$
[97]	Not Considerable	$K(\lvert p\rvert + \lvert q\rvert)$	$2p.(k + q)$
[135]	Not Considerable	$c(\lvert p\rvert + \lvert n\rvert)$	$(s + 1)\lvert p\rvert$
[94]	Not Considerable	$\lvert hash\rvert + j\lvert id\rvert + j\lvert k\rvert$	$\lvert k\rvert(j + 1) + \lvert c\rvert$

Table 6 Comparison of Computational Overhead between DO, CSP, and TPA During Auditing Phase

Ref.	Data Owner	Cloud Service Provider	Third Party Auditor
[88]	$jHash + (j * k)Exp + jAdd + jExp$	$Hash + kMulExp + kAdd + Exp + (k + 1)Mul$	$Hash + kHash + kMulExp + 3Exp + Mul + 2Pair + Mul$
[89]	$2jHash + 4jExp + 2jAdd + 2jMul + 2jMul$	$Hash + Mul + Exp + Mul + Add + Exp$	$KHash + 2Exp + (k + 1)Mul + Mul + Exp$
[90]	$jHash + jExp + jMul + Add$	$Exp + (k - 1)Mul + (k - 1)Add + kExp$	$4Pair + 2(k - 1)Add + 2Mul + 2Exp + (k + 1)Mul + KHash + (k + 1)Exp$
[133]	$jHash + 2jExp + jAdd + jMul + jMul$	$j + 2Exp + (j + 1)Mul + (k + 1)Exp + kMul$	$4Pair + (k + 2)Exp + (j + 2)Exp$
[78]	Add	$n(2Exp + Mul + Hash$	$KHash + 2Hash + 2(k + 1)Mul + (2k + 3)Exp + 2Pair + (k - 1)Add + kMul$
[87]	$10 * j(Add + Shift + Sub + MixC) + j * Hash$	$j\|k\|$	$j * Hash + Com$
[105]	$j(Hash + Mul + Enc) + Exp$	$(j - 1)Mul + 2Exp$	Not Applicable
[91]	$4Encrypt + 4Add$	$2Decrypt$	$2Decrypt + Encrypt + Add + Comp$
[97]	$jHash + jMul + jAdd + jInv$	$Hash + 2Add + Mul + Inv + 4Mul$	$Mul + 2Pair + Add$
[135]	$jHash + 2jExp + jAdd + jMul + jMul$	$j * k(Add + Mul + jExp_{G1} + Mul$	$(j + k + 1)Mul + 2Pair + (j + k)Exp$
[94]	$4jMul + jHash + Exp$	$4Mul + Exp$	$kAdd$

and RTPA. Here, Pair denotes bilinear pairing operatons, Hash denotes hash function, Mul denotes multiplication operation, ADD denotes addition operation, Exp denotes exponential operation,Inv denotes inverser operation,Encrypt denotes encryption operation, decrypt denotes decryption operation,and Sub denotes subtraction operation etc.

c) **Storage overhead**: Entire file or block files, metadata, signature, and replica blocks are required to be stored at cloud storage and at client side depending on the policy of system models. Cloud user storage overhead should be little during auditing verification to save extra storage overhead [36].

d) **Cost overhead**: It denotes the summarized cost of communication overhead, computational overhead, and storage overhead.

e) **Data Dynamic Analysis**: Stored data in cloud storage is not always static. Sometimes, alternation of data, deletion of data or addition of new data with old one are basic functions that come into the practical picture due to the dynamic demanding nature of clients. Therefore, data integrity verification should be done after all dynamic operations on stored data. In [93], the insertion, deletion and updation time of increasing data blocks are less than [123] due to less depth of the authenticated structure of the dynamic data integrity auditing scheme.

Comparative analysis of data integrity strategies

integrity checking scheme

This section presents a comparative study and comparison of data integrity strategies. Table 7 shows a comparative analysis of the data integrity strategy of cloud storage for expected design methods with limitations. Zang et al. [88] introduced a random masking technique in public audibility scheme during the computation of proof information generation time. Due to the linear relationship between the data block and proof information, malicious

Table 7 Comparative Analysis of the data integrity strategy of cloud storage

Ref.	Objectives	Limitations
[88]	Public auditing, resist all external adversary, protect data from a malicious auditor	Due to the missing of data storing acknowledge verification, the reputation of the Cloud server may be destroyed
[89]	Public data integrity, error localization, replica level auditing, dynamic update	Due to missing of data storing acknowledge verification, the reputation of CS may be destroyed
[90]	Data integrity auditing, sensitive data hiding	Due to missing of audit message verification scheme, TPA can deceive user about audit message
[85]	Data auditing, privacy-preserving	Audit report needs to verify otherwise TPA may be malicious TPA
[61]	Data integrity, resist replay attack and MITC attack	Data privacy issue because after repeatedly passed challenging phase, CSP becomes capable of getting original data block
[87]	Public auditing, data integrity	Audit message verification scheme need to be presented otherwise TPA may be malicious
[86]	Data integrity for static data resist from the external adversary	The author assumes that TPA is a trusted one but practically not possible
[91]	Public auditing, data integrity, dynamic data operation	Acknowledgment message about insert, modification and deletion of data needs to verify otherwise CS may be malicious CS
[136]	Public auditing, dynamic big graph data operation	During verifying time of dynamic graph operations, data privacy is not properly maintained
[93]	Dynamic update, data integrity auditing, reply forgery and reply attack	An audit message verification scheme needs to be present otherwise TPA may be malicious TPA
[78]	Public auditing, data integrity	Audit message and acknowledge message verification scheme needs to be present otherwise TPA and cloud may be malicious
[97]	Public auditing, reduce computational overhead, resist adaptive chosen-message attack	Validation results need to be verified otherwise TPA may be malicious
[137]	Data integrity, privacy-preserving	An audit message verification scheme needs to be present otherwise TPA may be malicious TPA
[122]	Data integrity, resist forge attack	No effective and secure data integrity scheme is present to support the data deduplication process of fog and cloud node
[126]	Dynamic auditing, dynamic data operation, resist reply attack and replace attack	BLS signature is not suitable for a big data environment
[125]	Certificateless public verification	Searching time over encrypted outsourced data in blockchain system takes much time
[124]	Zero-knowledge public auditing, privacy-preserving	Not applicable for large scale big data and TPA don't have the capability of auditing multiple user's data simultaneously

adversaries are capable of inert the effectiveness of the SWP scheme. In the SWP scheme, CSP generates proof information and sends it to TPA for verification. There may be an uncertain situation arise when CSP is intruded on by an external and malicious adversary that can alter every data block's information. To hoax TPA and pass the verification test, a malicious adversary can eavesdrop challenge message and break off the proof message. Therefore, in the SWP scheme, we assume that TPA is the trustworthy element. But practically, it is not possible. To defend against external malicious adversaries without a protective channel, the author proposed here a nonlinear disturbance code as a random masking technique to alter the linear relationship into a nonlinear relationship between data blocks and proof messages. The author applied a BLS hash signature on each block to help the verifier for random block verification. These public audibility verification techniques assure boundless, effective, stateless auditor and soundness criteria with two limitations are that due to the missing data storing acknowledge verification, the reputation of the Cloud services may be destroyed and this scheme is applicable for only static data.

M Thangavel et al. [89] proposed a novel auditing framework, which protects cloud storage from malicious attacks. This technique is based on a ternary and replica-based ternary hash tree which ensures dynamically block updating, data correctness with error localization operation, data insertion, and data deletion operations. W. Shen et al. [90] introduced identity-based data auditing scheme to hide sensitive information at the block level for securing cloud storage during data sharing time. Using this scheme, sanitizer sanitizes data blocks containing sensitive information. Chameleon hash and an unforgeable chameleon hash signature do not provide blockless auditing and require high computational overhead. Hence, this PKG-based signature method assures blockless verification and reduces computational overhead. These public audibility verification techniques assure auditing soundness, private key correctness, and sensitive information hiding one limitation is that due to missing audit messages, TPA can deceive users about data verification. S.Mohanty et al. [85] introduce a confidentiality-preserving auditing scheme by which cloud users can easily verify the risk of the used service from the audit report which is maintained by TPA. This scheme has two benefits. First, it helps to check the integrity of cloud users' data. Second, it verifies the TPA's authentication and repudiation. In this scheme, the author proposed a system model which supports the basic criteria of cloud security auditing, confidentiality, and availability. HMAC-MD5 technique is used on metadata to maintain data privacy on the TPA side. Chen et al. [61] proposed MAC oriented data integrity technique based on the metadata verification method which reinforces auditing correctness. These technique helps to protect stored data in cloud storage from MitM and replay attacks. But this scheme needs to improve because, after some repeated pass of challenge-proof messages, CSP will have the ability to get actual block elements of the user's confidential data.

S. Hiremath et al. [87] established a public blockless data integrity scheme that secures fixed time to check data of variable size files. For data encryption, the author uses the AES algorithm and SHA-2 algorithm for the data auditing scheme. The author uses the concept of random masking and Homomorphic Linear Authenticator (HLA) techniques to ensure stored data confidentiality during auditing time. But this scheme is only applicable for static data stored in cloud storage. Hence, it needs to expand for dynamic data operations. T. Subha. et al. [86] introduced the idea of public auditability to check the correctness of stored data in cloud storage and assume that TPA is a trusty entity. Data privacy mechanisms like Knox and Oruta have been proposed here to grow the security level at cloud storage and resist active adversary attacks. The author uses the Merkle hash tree to encrypt data block elements. B.Shao et al. [93] established a hierarchical Multiple Branches Tree(HMBT) which secures users' data auditing correctness, fulfills the crypto criteria of data privacy, and gives protection against forgery and replay attacks. The scheme is used a special hash function to give BLS signature on block elements and helps in public auditing.DCDV is a concept based on a hierarchical time tree and Merkle hash tree. Simultaneous execution of access control and data auditing mechanism rarely happens in attribute-based cryptography. Hence, Dual Control and Data Variable(DCDV) data integrity scheme is proposed in [132]. This scheme ensures the solution of the private data leakage problem by the user's secret key and assures the correctness of the auditing scheme. A PDP technique is proposed for data integrity verification scheme that supports dynamic data update operations, reduces communication overhead for fine-grained dynamic update of Bigdata increases the protection level of stored data at cloud storage, and resists collusion resistance attacks and batch auditing [114]. Another novel public auditing scheme based on an identity-based cryptographical idea ensures low computational overhead from revocated users during the possession of all file blocks. It fulfills the crypto criteria of soundness, correctness, security, and efficiency of revoke users [78].

Some research works introduced BLS cryptographical signature which has the shortest length among all available signatures [88]. This signature is based on a special hash function that is probabilistic, not deterministic. Also, it has more overhead of power exponential and hash calculation. To overcome signature efficiency and computational overhead, a new signature ZSS is proposed [97]. This integrity scheme supports crypto criteria like privacy protection, public auditing correctness, and resisting message and forgery. An attribute-based data auditing scheme is proposed in [137] which proved data correctness and soundness based on discrete logarithm and Diffie-Hellman key exchange algorithm. This scheme maintains the privacy of confidential data of cloud users and resists collusion in blocks during auditing verification time. attacks. ID-based remote data auditing scheme(ID-PUIC) is introduced here which secures efficiency, security, and flexibility with the help of the Diffie-Hellman problem [98]. It also supports ID-based proxy data upload operation when users are restricted to access public cloud servers. It shows a lower computation cost of server and TPA than [107]. Both researches works [105, 126] have worked on public checking of data intactness of outsourced data and reducing communication and computational cost of the verifier. These also support dynamic data auditing, blockless verification, and privacy preservation.

Future trends in data integrity approaches

As further research work, we are discussing here the future direction of the data integrity scheme to enlarge the scope of cloud data security for research process continuity. New emerging trends in data integrity schemes are listed below.In [39], authors have already discussed and shown evolutionary trends of data integrity schemes through a timeline representation from 2007 to 2015 which presented possible scopes of data integrity strategy. Hence, we show a visual representation of all probable trends of the integrity scheme from 2016 to 2022 in the timeline infographic template, Fig. 2.

a) **Blockchain data-based integrity** : Blockchain technology is decentralized, peer-peer technology. It supports scalable and distributed environments in which all the data are treated as transparent blocks that contain cryptographic hash information of the previous block, and timestamps to resist any alteration of a single data block without modifying all the subsequent linked blocks. This feature of this technology improves the performance of cloud storage and maintains the trust of data owners by increasing data privacy through the Merkle tree concept. In [138], a distributed virtual agent model is proposed through mobile agent technology to maintain the reliability of cloud data and to ensure trust verification of cloud data via multi-tenant. In [139], a blockchain-based generic framework is proposed to increase the security of the provenance data in cloud storage which is important for accessing log information of cloud data securely. In [140–142], all research works have the same intention of using blockchain technology to enhance data privacy and maintain data integrity in cloud storage.In Table 8, this article show use Blockchain technology to overcome some issues of cloud storage.

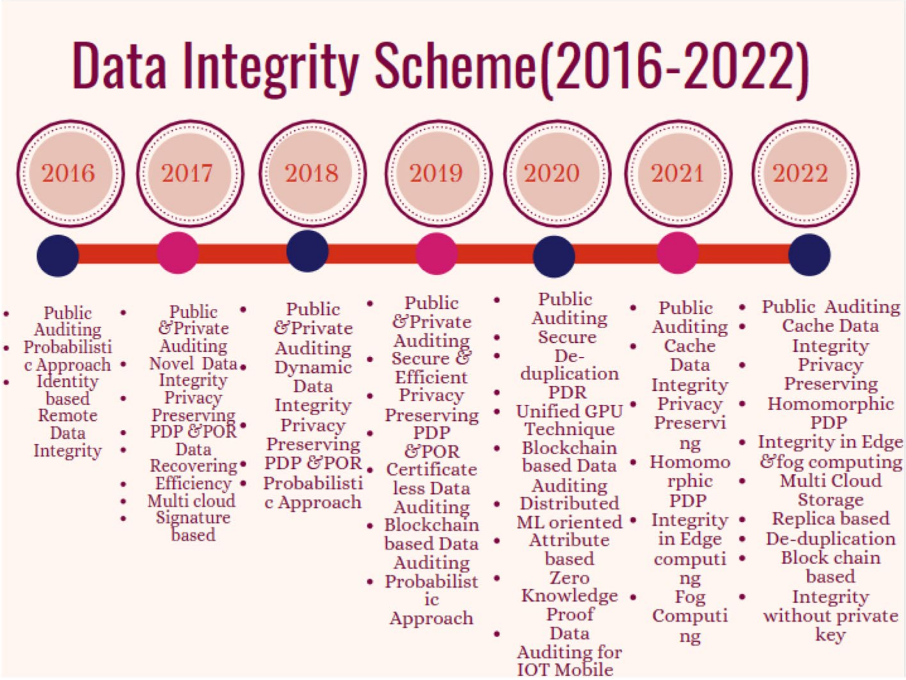

Fig. 2 Timeline Infographic of Data Integrity

Table 8 Cloud Computing with Blockchain Technology and its merit with regards to storage level data integrity strategies

Ref.	Issues of Cloud Storage	Merit of Blockchain Technology	Achivements
[100]	In multi-cloud storage environment, majority of the comparable schemes depend on reliable org. like the CSP and the centralised TPA, related and it might be challenging to pinpoint malevolent service providers in the wake of service disputes	used to detect service disputes and accurately identify dishonest service providers, blockchain technology is utilised to record the interactions between users, service providers, and organizers utilized,during the data auditing process	batch verification at a cheap cost without a TPA
[143]	Several TPAs generate challenges for multi-cloud storage, sent to CS to verify data custody. TPAs may dishonestly exploit auditing protocols or collude with CS.	With the usage of blockchain technology, CS might be able to deduce the challenge messages, and there's a chance that user data might be disclosed to the TPA while the audit is being conducted	This ensures decentralized, private audits, allowing public result verification for users
[144]	App development requires data sharing and storage. Functional encryption(FE) solves public-key encryption drawbacks, but requires expensive bilinear pairings.	Cryptocurrency built on the blockchain that allows users to pay third parties when their outsourced decryption is successfully completed	The payment in an FE with outsourced decryption scheme is achieved
[133]	In order to measure cloud data of virtual machines (VMs), two critical concerns in safe IaaS cloud storage are integrity evaluation and decision making.	A two-layer blockchain network can be used to create a revisable user-defined policy-based encryption mechanism and to construct a one-to-one relationship between a user, a node, and a virtual machine.	Enhance data integrity level and aids in controlling the scope of approved verifiers in a flexible manner
[145]	Vast storage proofs and/or vast auditor states are prerequisites for the application of Dynamic Proof-of-Storage techniques designed for traditional cloud storage to Distributed Systems	Static proof-of-success systems promise compact proofs; dynamic on-blockchain auditing protocols can provide concretely tiny auditor states	Index information management is accomplished by optimisation strategies.
[146]	In a multicloud storage environment, controlling scalability, data governance,non-tampering, trustworthiness, and transparency are two challenges.	A novel strategy for the security of huge data storage that makes use of highway protocol and blockchain technology to create new blocks that address problems with baseline models	dynamic control over sharing data manipulation is achieved.

b) **Data integrity in fog computing** : Generally, privacy protection schemes are able to resist completely insider attacks in cloud storage. In [147], a fog computing-based TLS framework is proposed to maintain the privacy of data in Fog servers. The extension part of cloud computing is fog computing which was firstly introduced in 2011 [148]. The three advantages of fog computing are towering real-time, low latency, and broader range geographical distribution which is embedded with cloud computing to ensure the privacy of data in fog servers which is a powerful supplement to maintain data privacy preservation in cloud storage.

c) **Distributed Machine Learning Oriented Data Integrity** : In artificial intelligence, maintaining the integrity of training data in the distributed machine learning environment is a rapidly growing challenge due to network forge attacks. In [136], distributed machine learning-oriented data integrity verification scheme (DML-DIV) is introduced to assure training data intactness and to secure training data model. PDP sampling auditing algorithm is adopted here to resist tampering attacks and forge attacks. Discrete logarithm problem (DLP) is introduced in the DML-DIV scheme to ensure privacy preservation of training data during TPA's challenge verification time. To reduce key escrow problem and certificate cost, identity-based cryptography and key generation technology is proposed here.

d) **Data Integrity in Edge Computing** : Edge computing is an extensional part of distributed computing. Cache data integrity is a new concept in edge computing developed based on cloud computing which serves optimized data retrieval latency on edge servers and gives centralized problems of cloud storage server.Edge data integrity(EDI) concept is first proposed to effectively handle auditing of vendor apps' cache data on edge servers which is a challenging issue in dynamic, distributed, and volatile edge environments described In [149]. Research work proposed here EDI-V model using variable Merkle hash tree (VMHT) structure to maintain cache data auditing on a large scale server through generating integrity of replica data of it. In [150], the EDI-S model is introduced to verify the integrity of edge data and to localize the corrupted data on edge servers by generating digital signatures of each edge's replica.

Conclusion

With the continuously enlarging popularity of attractive and optimized cost-based cloud services, it is inconvenient to make sure data owners the intactness of outsourced data in cloud storage environments has become a disaster security challenge. We have tried to highlight several issues and the corresponding solution approaches for cloud data integrity which will provide a visualization as well as clear directions to researchers. The current state of the art in this mentioned research field will provide extra milestones in several areas like cloud-based sensitive health care, secured financial service, managing social media flat-forms, etc. In this paper, we have discussed phases of data integrity, characteristics of data integrity scheme, classification of data integrity strategy based on the type of proposal, nature of data and type of verification schemes, and desired design challenges of data integrity strategy based on performance overhead. We have also identified issues in physical cloud storage and attacks on storage-level services along with mitigating solutions. Lastly, we have established here a timeline infographic visual representation of a variety of data integrity schemes and future aspects of data integrity strategy to explore all the security directions of cloud storage.

Acknowledgements
Not applicable.

Authors' contributions
Paromita Goswami, Neetu Faujdar and Somen Debnath invented the proposed methodology and wrote the main manuscript text, Ajay Kumar Khan prepared tables and Ghyanshyam Singh prepared figures, Ghyanshyam Singh and Ajay Kumar Singh is also written literature and all authrs reviewed the whole manuscript.

Funding
Not applicable.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 6 May 2023 Accepted: 30 January 2024
Published online: 15 February 2024

References

1. Buyya R, Broberg J, Goscinski AM (2010) Cloud computing: Principles and paradigms, vol 87. Wiley
2. Mell P, Grance T, et al (2011) The nist definition of cloud computing
3. Wu C, Buyya R, Ramamohanarao K (2019) Cloud pricing models: Taxonomy, survey, and interdisciplinary challenges. ACM Comput Surv (CSUR) 52(6):1–36
4. Dimitri N (2020) Pricing cloud iaas computing services. J Cloud Comput 9(1):1–11

5. Roy SS, Garai C, Dasgupta R (2015) Performance analysis of parallel cbar in mapreduce environment. In: 2015 International Conference on Computing, Communication and Security (ICCCS). IEEE, pp 1–7
6. Singhal S, Sharma A (2020) Load balancing algorithm in cloud computing using mutation based pso algorithm. In: Advances in Computing and Data Sciences: 4th International Conference. Springer, pp 224–233
7. Luong NC, Wang P, Niyato D, Wen Y, Han Z (2017) Resource management in cloud networking using economic analysis and pricing models: A survey. IEEE Commun Surv Tutorials 19(2):954–1001
8. Goswami P, Roy SS, Dasgupta R (2017) Design of an architectural framework for providing quality cloud services. In: International Conference on Grid, Cloud, & Cluster Computing. pp 17–23
9. Anuradha V, Sumathi D (2014) A survey on resource allocation strategies in cloud computing. In: International Conference on Information Communication and Embedded Systems (ICICES2014). IEEE, pp 1–7
10. Magalhaes D, Calheiros RN, Buyya R, Gomes DG (2015) Workload modeling for resource usage analysis and simulation in cloud computing. Comput Electr Eng 47:69–81
11. Singhal S, Sharma A (2021) Mutative aco based load balancing in cloud computing. Eng Lett 29(4)
12. Chandramohan D, Vengattaraman T, Dhavachelvan P, Baskaran R, Venkatachalapathy V (2014) Fewss-framework to evaluate the service suitability and privacy in a distributed web service environment. Int J Model Simul Sci Comput 5(01):1350016
13. Klosterboer L (2011) ITIL capacity management. Pearson Education
14. Majumdar A, Roy SS, Dasgupta R (2017) Job migration policy in a structured cloud framework. In: 2017 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, pp 1529–1534
15. Singhal S, Sharma A (2021) A job scheduling algorithm based on rock hyrax optimization in cloud computing, vol 103. Springer, pp 2115–2142
16. Dong Y, Sun L, Liu D, Feng M, Miao T (2018) A survey on data integrity checking in cloud. In: 2018 1st International Cognitive Cities Conference (IC3). IEEE, pp 109–113
17. Bian G, Fu Y, Shao B, Zhang F (2022) Data integrity audit based on data blinding for cloud and fog environment. IEEE Access 10:39743–39751. https://doi.org/10.1109/ACCESS.2022.3166536
18. Iqbal A, Saham H (2014) Data integrity issues in cloud servers. Int J Comput Sci Issues (IJCSI) 11(3):118
19. Caronni G, Waldvogel M (2003) Establishing trust in distributed storage providers. In: Proceedings Third International Conference on Peer-to-Peer Computing (P2P2003). IEEE, pp 128–133
20. Ogiso S, Mohri M, Shiraishi Y (2020) Transparent provable data possession scheme for cloud storage. In: 2020 International Symposium on Networks, Computers and Communications (ISNCC). IEEE, pp 1–5
21. Masood R, Pandey N, Rana Q (2020) Dht-pdp: A distributed hash table based provable data possession mechanism in cloud storage. In: 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). IEEE, pp 275–279
22. Bian G, Chang J (2020) Certificateless provable data possession protocol for the multiple copies and clouds case. IEEE Access 8:102958–102970
23. Zhang X, Wang X, Gu D, Xue J, Tang W (2022) Conditional anonymous certificateless public auditing scheme supporting data dynamics for cloud storage systems. IEEE Trans Netw Serv Manag 19(4):5333–5347. https://doi.org/10.1109/TNSM.2022.3189650
24. Li J, Yan H, Zhang Y (2021) Certificateless public integrity checking of group shared data on cloud storage. IEEE Trans Serv Comput 14(1):71–81. https://doi.org/10.1109/TSC.2018.2789893
25. Yuan Y, Zhang J, Xu W (2020) Dynamic multiple-replica provable data possession in cloud storage system. IEEE Access 8:120778–120784
26. Juels A, Kaliski Jr BS (2007) Pors: Proofs of retrievability for large files. In: Proceedings of the 14th ACM conference on Computer and communications security. ACM, pp 584–597
27. González-Manzano L, Orfila A (2015) An efficient confidentiality-preserving proof of ownership for deduplication. J Netw Comput Appl 50:49–59
28. Yu CM, Chen CY, Chao HC (2015) Proof of ownership in deduplicated cloud storage with mobile device efficiency. IEEE Netw 29(2):51–55
29. Di Pietro R, Sorniotti A (2012) Boosting efficiency and security in proof of ownership for deduplication. In: Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security. ACM, pp 81–82
30. Gentry C (2009) Fully homomorphic encryption using ideal lattices. In: Proceedings of the forty-first annual ACM symposium on Theory of computing. ACM, pp 169–178
31. Enoch SY, Hong JB, Kim DS (2018) Time independent security analysis for dynamic networks using graphical security models. In: 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, pp 588–595
32. Kumar S, Singh SK, Singh AK, Tiwari S, Singh RS (2018) Privacy preserving security using biometrics in cloud computing. Multimed Tools Appl 77(9):11017–11039
33. Sirohi P, Agarwal A (2015) Cloud computing data storage security framework relating to data integrity, privacy and trust. In: 2015 1st international conference on next generation computing technologies (NGCT). IEEE, pp 115–118
34. Prasad D, Singh BR, Akuthota M, Sangeetha M (2014) An etiquette approach for public audit and preserve data at cloud. Int J Comput Trends Technol (IJCTT) 16
35. Skibitzki B (2021) How zebra technologies manages security & risk using security command center. https://cloud.google.com/blog/products/identity-security/how-zebra-technologies
36. Li A, Chen Y, Yan Z, Zhou X, Shimizu S (2020) A survey on integrity auditing for data storage in the cloud: from single copy to multiple replicas. IEEE Trans Big Data 8(5):1428–1442.
37. Tan CB, Hijazi MHA, Lim Y, Gani A (2018) A survey on proof of retrievability for cloud data integrity and availability: Cloud storage state-of-the-art, issues, solutions and future trends. J Netw Comput Appl 110:75–86
38. Pujar SR, Chaudhari SS, Aparna R (2020) Survey on data integrity and verification for cloud storage. In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, pp 1–7
39. Zafar F, Khan A, Malik SUR, Ahmed M, Anjum A, Khan MI, Javed N, Alam M, Jamil F (2017) A survey of cloud computing data integrity schemes: Design challenges, taxonomy and future trends. Comput Secur 65:29–49
40. Debnath S, Bhuyan B (2019) Large universe attribute based encryption enabled secured data access control for cloud storage with computation outsourcing. Multiagent Grid Syst 15(2):99–119
41. Hsien WF, Yang CC, Hwang MS (2016) A survey of public auditing for secure data storage in cloud computing. Int J Netw Secur 18(1):133–142
42. Zhou L, Fu A, Yu S, Su M, Kuang B (2018) Data integrity verification of the outsourced big data in the cloud environment: A survey. J Netw Comput Appl 122:1–15
43. Liu CW, Hsien WF, Yang CC, Hwang MS (2016) A survey of public auditing for shared data storage with user revocation in cloud computing. Int J Netw Secur 18(4):650–666
44. Garg N, Bawa S (2016) Comparative analysis of cloud data integrity auditing protocols. J Netw Comput Appl 66:17–32
45. Sutradhar MR, Sultana N, Dey H, Arif H (2018) A new version of kerberos authentication protocol using ecc and threshold cryptography for cloud security. In: 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). IEEE, pp 239–244
46. Patel SC, Singh RS, Jaiswal S (2015) Secure and privacy enhanced authentication framework for cloud computing. In: 2015 2nd International Conference on Electronics and Communication Systems (ICECS). IEEE, pp 1631–1634
47. Hong H, Sun Z, Xia Y (2017) Achieving secure and fine-grained data authentication in cloud computing using attribute based proxy signature. In: 2017 4th International Conference on Information Science and Control Engineering (ICISCE). IEEE, pp 130–134
48. Wang W, Ren L, Chen L, Ding Y (2019) Intrusion detection and security calculation in industrial cloud storage based on an improved dynamic immune algorithm. Inf Sci 501:543–557
49. Yan Q, Yu FR, Gong Q, Li J (2015) Software-defined networking (sdn) and distributed denial of service (ddos) attacks in cloud computing

environments: A survey, some research issues, and challenges. IEEE Commun Surv Tutor 18(1):602–622

50. Dong S, Abbas K, Jain R (2019) A survey on distributed denial of service (ddos) attacks in sdn and cloud computing environments. IEEE Access 7:80813–80828
51. Thirumallai C, Mekala MS, Perumal V, Rizwan P, Gandomi AH (2020) Machine learning inspired phishing detection (pd) for efficient classification and secure storage distribution (ssd) for cloud-iot application. In: 2020 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, pp 202–210
52. Mary BF, Amalarethinam DG (2017) Data security enhancement in public cloud storage using data obfuscation and steganography. In: 2017 World Congress on Computing and Communication Technologies (WCCCT). IEEE, pp 181–184
53. Nakouri I, Hamdi M, Kim TH (2017) A new biometric-based security framework for cloud storage. In: 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC). IEEE, pp 390–395
54. Meddeb-Makhlouf A, Zarai F, et al (2018) Distributed firewall and controller for mobile cloud computing. In: 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA). IEEE/ACS, pp 1–9
55. Fu Y, Au MH, Du R, Hu H, Li D (2020) Cloud password shield: A secure cloud-based firewall against ddos on authentication servers. In: 2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS). IEEE, pp 1209–1210
56. Zeidler C, Asghar MR (2018) Authstore: Password-based authentication and encrypted data storage in untrusted environments. In: 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, pp 996–1001
57. Erdem E, Sandıkkaya MT (2018) Otpaas-one time password as a service. IEEE Trans Inf Forensic Secur 14(3):743–756
58. Chandramohan D, Vengattaraman T, Rajaguru D, Baskaran R, Dhavachelvan P (2013) Emppc-an evolutionary model based privacy preserving technique for cloud digital data storage. In: 2013 3rd IEEE International Advance Computing Conference (IACC). IEEE, pp 89–95
59. Bakas A, Dang HV, Michalas A, Zalitko A (2020) The cloud we share: Access control on symmetrically encrypted data in untrusted clouds. IEEE Access 8:210462–210477
60. Rukavitsyn AN, Borisenko KA, Holod II, Shorov AV (2017) The method of ensuring confidentiality and integrity data in cloud computing. In: 2017 XX IEEE International Conference on Soft Computing and Measurements (SCM). IEEE, pp 272–274
61. Chen Y, Li L, Chen Z (2017) An approach to verifying data integrity for cloud storage. In: 2017 13th International Conference on Computational Intelligence and Security (CIS). IEEE, pp 582–585
62. Alneyadi S, Sithirasenan E, Muthukkumarasamy V () A survey on data leakage prevention systems. J Netw Comput Appl 62:137–152
63. Baloch FS, Muhammad TA, Waqas L, Mehmet B, Muhammad AN, Gönül Cömertpay, Nergiz Çoban et al (2023) "Recent advancements in the breeding of sorghum crop: current status and future strategies for marker-assisted breeding." Frontiers in Genetics 14:1150616.
64. Rakotondravony N, Taubmann B, Mandarawi W, Weishäupl E, Xu P, Kolosnjaji B, Protsenko M, De Meer H, Reiser HP (2017) Classifying malware attacks in iaas cloud environments. J Cloud Comput 6(1):1–12
65. Perez-Botero D, Szefer J, Lee RB (2013) Characterizing hypervisor vulnerabilities in cloud computing servers. In: Proceedings of the 2013 international workshop on Security in cloud computing. ACM, pp 3–10
66. Tunc C, Hariri S, Merzouki M, Mahmoudi C, De Vaulx FJ, Chbili J, Bohn R, Battou A (2017) Cloud security automation framework. In: 2017 IEEE 2nd International Workshops on Foundations and Applications of Self Systems. IEEE, pp 307–312
67. Maithili K, Vinothkumar V, Latha P (2018) Analyzing the security mechanisms to prevent unauthorized access in cloud and network security. J Comput Theor Nanosci 15(6–7):2059–2063
68. Somasundaram TS, Prabha V, Arumugam M (2012) Scalability issues in cloud computing. In: 2012 Fourth International Conference on Advanced Computing (ICoAC). IEEE, pp 1–5
69. Yousafzai A, Gani A, Noor RM, Sookhak M, Talebian H, Shiraz M, Khan MK (2017) Cloud resource allocation schemes: review, taxonomy, and opportunities. Knowl Inf Syst 50(2):347–381
70. Natu M, Ghosh RK, Shyamsundar RK, Ranjan R (2016) Holistic performance monitoring of hybrid clouds: Complexities and future directions. IEEE Cloud Comput 3(1):72–81
71. Mahajan A, Sharma S (2015) The malicious insiders threat in the cloud. Int J Eng Res Gen Sci 3(2):245–256
72. Liao X, Alrwais S, Yuan K, Xing L, Wang X, Hao S, Beyah R (2018) Cloud repository as a malicious service: challenge, identification and implication. Cybersecurity 1(1):1–18
73. Singh A, Chatterjee K (2017) Cloud security issues and challenges: A survey. J Netw Comput Appl 79:88–115
74. Daniel E, Durga S, Seetha S (2019) Panoramic view of cloud storage security attacks: an insight and security approaches. In: 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, pp 1029–1034
75. Devi BK, Subbulakshmi T (2017) Ddos attack detection and mitigation techniques in cloud computing environment. In: 2017 International Conference on Intelligent Sustainable Systems (ICISS). IEEE, pp 512–517
76. Yusop ZM, Abawajy J (2014) Analysis of insiders attack mitigation strategies. Procedia-Soc Behav Sci 129:581–591
77. Song H, Li J, Li H (2021) A cloud secure storage mechanism based on data dispersion and encryption. IEEE Access 9:63745–63751. https://doi.org/10.1109/ACCESS.2021.3075340
78. Zhang Y, Yu J, Hao R, Wang C, Ren K (2020) Enabling efficient user revocation in identity-based cloud storage auditing for shared big data. IEEE Trans Dependable Secure Comput 17(3):608–619. https://doi.org/10.1109/TDSC.2018.2829880
79. Zuo C, Shao J, Liu JK, Wei G, Ling Y (2018) Fine-grained two-factor protection mechanism for data sharing in cloud storage. IEEE Trans Inf Forensic Secur 13(1):186–196. https://doi.org/10.1109/TIFS.2017.2746000
80. Cui H, Deng RH, Li Y, Wu G (2019) Attribute-based storage supporting secure deduplication of encrypted data in cloud. IEEE Trans Big Data 5(3):330–342. https://doi.org/10.1109/TBDATA.2017.2656120
81. Sun S, Ma H, Song Z, Zhang R (2022) Webcloud: Web-based cloud storage for secure data sharing across platforms. IEEE Trans Dependable Secure Comput 19(3):1871–1884. https://doi.org/10.1109/TDSC.2020.3040784
82. Cheng K, Wang L, Shen Y, Wang H, Wang Y, Jiang X, Zhong H (2021) Secure kk-nn query on encrypted cloud data with multiple keys. IEEE Trans Big Data 7(4):689–702. https://doi.org/10.1109/TBDATA.2017.2707552
83. Wang B, Li B, Li H (2014) Oruta: Privacy-preserving public auditing for shared data in the cloud. IEEE Trans Cloud Comput 2(1):43–56
84. Indhumathil T, Aarthy N, Devi VD, Samyuktha V (2017) Third-party auditing for cloud service providers in multicloud environment. In: 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM). IEEE, pp 347–352
85. Mohanty S, Pattnaik PK, Kumar R (2018) Confidentiality preserving auditing for cloud computing environment. In: 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE). IEEE, pp 1–4
86. Subha T, Jayashri S (2017) Efficient privacy preserving integrity checking model for cloud data storage security. In: 2016 Eighth International Conference on Advanced Computing (ICoAC). IEEE, pp 55–60
87. Hiremath S, Kunte S (2017) A novel data auditing approach to achieve data privacy and data integrity in cloud computing. In: 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT). IEEE, pp 306–310
88. Zhang Y, Xu C, Li H, Liang X (2016) Cryptographic public verification of data integrity for cloud storage systems. IEEE Cloud Comput 3(5):44–52
89. Thangavel M, Varalakshmi P (2019) Enabling ternary hash tree based integrity verification for secure cloud data storage. IEEE Trans Knowl Data Eng 32(12):2351–2362
90. Shen W, Qin J, Yu J, Hao R, Hu J (2018) Enabling identity-based integrity auditing and data sharing with sensitive information hiding for secure cloud storage. IEEE Trans Inf Forensic Secur 14(2):331–346
91. Singh P, Saroj SK (2020) A secure data dynamics and public auditing scheme for cloud storage. In: 2020 6th International Conference on

Advanced Computing and Communication Systems (ICACCS). IEEE, pp 695–700

92. Ni J, Yu Y, Mu Y, Xia Q (2013) On the security of an efficient dynamic auditing protocol in cloud storage. IEEE Trans Parallel Distrib Syst 25(10):2760–2761
93. Shao B, Bian G, Wang Y, Su S, Guo C (2018) Dynamic data integrity auditing method supporting privacy protection in vehicular cloud environment. IEEE Access 6:43785–43797
94. Shen J, Liu D, He D, Huang X, Xiang Y (2017) Algebraic signatures-based data integrity auditing for efficient data dynamics in cloud computing. IEEE Trans Sustain Comput 5(2):161–173
95. Wang B, Li H, Liu X, Li F, Li X (2014) Efficient public verification on the integrity of multi-owner data in the cloud. J Commun Netw 16(6):592–599
96. Yu Y, Li Y, Yang B, Susilo W, Yang G, Bai J (2017) Attribute-based cloud data integrity auditing for secure outsourced storage. IEEE Trans Emerg Top Comput 8(2):377–390
97. Zhu H, Yuan Y, Chen Y, Zha Y, Xi W, Jia B, Xin Y (2019) A secure and efficient data integrity verification scheme for cloud-iot based on short signature. IEEE Access 7:90036–90044
98. Wang H, He D, Tang S (2016) Identity-based proxy-oriented data uploading and remote data integrity checking in public cloud. IEEE Trans Inf Forensic Secur 11(6):1165–1176
99. Thakur AS, Gupta P (2014) Framework to improve data integrity in multi cloud environment
100. Zhang C, Xu Y, Hu Y, Wu J, Ren J, Zhang Y (2021) A blockchain-based multi-cloud storage data auditing scheme to locate faults. IEEE Trans Cloud Comput 10(4):2252–2263.
101. Subha T, Jayashri S (2014) Data integrity verification in hybrid cloud using ttpa. In: Networks and communications (NetCom2013). Springer, pp 149–159
102. Mao J, Zhang Y, Li P, Li T, Wu Q, Liu J (2017) A position-aware merkle tree for dynamic cloud data integrity verification. Soft Comput 21(8):2151–2164
103. Han S, Liu S, Chen K, Gu D (2014) Proofs of retrievability based on mrd codes. In: International Conference on Information Security Practice and Experience. Springer, pp 330–345
104. Kaaniche N, El Moustaine E, Laurent M (2014) A novel zero-knowledge scheme for proof of data possession in cloud storage applications. In: 2014 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. IEEE, pp 522–531
105. Khedr WI, Khater HM, Mohamed ER (2019) Cryptographic accumulator-based scheme for critical data integrity verification in cloud storage. IEEE Access 7:65635–65651
106. Khatri TS, Jethava G (2013) Improving dynamic data integrity verification in cloud computing. In: 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). IEEE, pp 1–6
107. Wang H (2012) Proxy provable data possession in public clouds. IEEE Trans Serv Comput 6(4):551–559
108. Apolinário F, Pardal M, Correia M (2018) S-audit: Efficient data integrity verification for cloud storage. In: 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing and Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, pp 465–474
109. Li Y, Fu A, Yu Y, Zhang G (2017) Ipor: An efficient ida-based proof of retrievability scheme for cloud storage systems. In: 2017 IEEE International Conference on Communications (ICC). IEEE, pp 1–6
110. Shacham H, Waters B (2008) Compact proofs of retrievability. In: International conference on the theory and application of cryptology and information security. Springer, pp 90–107
111. Erway CC, Küpçü A, Papamanthou C, Tamassia R (2015) Dynamic provable data possession. ACM Trans Inf Syst Secur (TISSEC) 17(4):1–29
112. He D, Kumar N, Wang H, Wang L, Choo KKR (2017) Privacy-preserving certificateless provable data possession scheme for big data storage on cloud. Appl Math Comput 314:31–43
113. Wang B, Li B, Li H, Li F (2013) Certificateless public auditing for data integrity in the cloud. In: 2013 IEEE conference on communications and network security (CNS). IEEE, pp 136–144
114. Liu C, Chen J, Yang LT, Zhang X, Yang C, Ranjan R, Kotagiri R (2013) Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates. IEEE Trans Parallel Distrib Syst 25(9):2234–2244
115. Fu A, Li Y, Yu S, Yu Y, Zhang G (2018) Dipor: An ida-based dynamic proof of retrievability scheme for cloud storage systems. J Netw Comput Appl 104:97–106
116. Xu J, Chang EC (2012) Towards efficient proofs of retrievability. In: Proceedings of the 7th ACM symposium on information, computer and communications security. pp 79–80
117. Lu Y, Hu F (2019) Secure dynamic big graph data: Scalable, low-cost remote data integrity checking. IEEE Access 7:12888–12900
118. Ateniese G, Di Pietro R, Mancini LV, Tsudik G (2008) Scalable and efficient provable data possession. In: Proceedings of the 4th international conference on Security and privacy in communication netowrks. ACM, pp 1–10
119. Tian H, Chen Y, Chang CC, Jiang H, Huang Y, Chen Y, Liu J (2015) Dynamic-hash-table based public auditing for secure cloud storage. IEEE Trans Serv Comput 10(5):701–714
120. He D, Zeadally S, Wu L (2015) Certificateless public auditing scheme for cloud-assisted wireless body area networks. IEEE Syst J 12(1):64–73
121. Yoosuf MS, Anitha R (2022). LDuAP: lightweight dual auditing protocol to verify data integrity in cloud storage servers. J Ambient Intell Humanized Comput 13(8):3787–3805.
122. Tian H, Nan F, Chang CC, Huang Y, Lu J, Du Y (2019) Privacy-preserving public auditing for secure data storage in fog-to-cloud computing. J Netw Comput Appl 127:59–69
123. Singh AP, Pasupuleti SK (2016) Optimized public auditing and data dynamics for data storage security in cloud computing. Procedia Comput Sci 93:751–759
124. Wang C, Chow SS, Wang Q, Ren K, Lou W (2011) Privacy-preserving public auditing for secure cloud storage. IEEE Trans Comput 62(2):362–375
125. Zhang Y, Xu C, Lin X, Shen XS (2019) Blockchain-based public integrity verification for cloud storage against procrastinating auditors. IEEE Trans Cloud Comput 9(3):923–937.
126. Shen J, Shen J, Chen X, Huang X, Susilo W (2017) An efficient public auditing protocol with novel dynamic structure for cloud data. IEEE Trans Inf Forensic Secur 12(10):2402–2415
127. Oualha N, Leneutre J, Roudier Y (2012) Verifying remote data integrity in peer-to-peer data storage: A comprehensive survey of protocols. Peer-to-Peer Netw Appl 5(3):231–243
128. Xu Z, Wu L, Khan MK, Choo KKR, He D (2017) A secure and efficient public auditing scheme using rsa algorithm for cloud storage. J Supercomput 73(12):5285–5309
129. Sookhak M, Gani A, Talebian H, Akhunzada A, Khan SU, Buyya R, Zomaya AY (2015) Remote data auditing in cloud computing environments: a survey, taxonomy, and open issues. ACM Comput Surv (CSUR) 47(4):1–34
130. Mohammed A, Vasumathi D (2019) Locality parameters for privacy preserving protocol and detection of malicious third-party auditors in cloud computing. In: International Conference on Intelligent Computing and Communication. Springer, pp 67–76
131. Carroll M, Van Der Merwe A, Kotze P (2011) Secure cloud computing: Benefits, risks and controls. In: 2011 Information Security for South Africa. IEEE, pp 1–9
132. Zhang Q, Wang S, Zhang D, Wang J, Zhang Y (2019) Time and attribute based dual access control and data integrity verifiable scheme in cloud computing applications. IEEE Access 7:137594–137607
133. Li Y, Yu Y, Yang B, Min G, Wu H (2018) Privacy preserving cloud data auditing with efficient key update. Futur Gener Comput Syst 78:789–798
134. Shen W, Qin J, Yu J, Hao R, Hu J, Ma J (2021) Data integrity auditing without private key storage for secure cloud storage. IEEE Trans Cloud Comput 9(4):1408–1421. https://doi.org/10.1109/TCC.2019.2921553
135. Garg N, Bawa S, Kumar N (2020) An efficient data integrity auditing protocol for cloud computing. Futur Gener Comput Syst 109:306–316
136. Zhao XP, Jiang R (2020) Distributed machine learning oriented data integrity verification scheme in cloud computing environment. IEEE Access 8:26372–26384. https://doi.org/10.1109/ACCESS.2020.2971519
137. Yu Y, Au MH, Ateniese G, Huang X, Susilo W, Dai Y, Min G (2016) Identity-based remote data integrity checking with perfect data privacy preserving for cloud storage. IEEE Trans Inf Forensic Secur 12(4):767–778

138. Wei P, Wang D, Zhao Y, Tyagi SKS, Kumar N (2020) Blockchain data-based cloud data integrity protection mechanism. Futur Gener Comput Syst 102:902–911
139. Sifah EB, Xia Q, Agyekum KOBO, Xia H, Smahi A, Gao J (2021) A blockchain approach to ensuring provenance to outsourced cloud data in a sharing ecosystem. IEEE Syst J 16(1):1673–1684.
140. Huang P, Fan K, Yang H, Zhang K, Li H, Yang Y (2020) A collaborative auditing blockchain for trustworthy data integrity in cloud storage system. IEEE Access 8:94780–94794
141. Pise R, Patil S (2021) Enhancing security of data in cloud storage using decentralized blockchain. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE, pp 161–167
142. Sharma P, Jindal R, Borah MD (2019) Blockchain-based integrity protection system for cloud storage. In: 2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON). IEEE, pp 1–5
143. Miao Y, Huang Q, Xiao M, Li H (2020) Decentralized and privacy-preserving public auditing for cloud storage based on blockchain. IEEE Access 8:139813–139826. https://doi.org/10.1109/ACCESS.2020.3013153
144. Cui H, Wan Z, Wei X, Nepal S, Yi X (2020) Pay as you decrypt: Decryption outsourcing for functional encryption using blockchain. IEEE Trans Inf Forensic Secur 15:3227–3238. https://doi.org/10.1109/TIFS.2020.2973864
145. Duan H, Du Y, Zheng L, Wang C, Au MH, Wang Q (2023) Towards practical auditing of dynamic data in decentralized storage. IEEE Trans Dependable Secure Comput 20(1):708–723. https://doi.org/10.1109/TDSC.2022.3142611
146. Sasikumar A, Ravi L, Kotecha K, Abraham A, Devarajan M, Vairavasundaram S (2023) A secure big data storage framework based on blockchain consensus mechanism with flexible finality. IEEE Access 11:56712–56725. https://doi.org/10.1109/ACCESS.2023.3282322
147. Wang T, Zhou J, Chen X, Wang G, Liu A, Liu Y (2018) A three-layer privacy preserving cloud storage scheme based on computational intelligence in fog computing. IEEE Trans Emerg Top Comput Intell 2(1):3–12
148. Bonomi F, Milito R, Zhu J, Addepalli S (2012) Fog computing and its role in the internet of things. In: Proceedings of the first edition of the MCC workshop on Mobile cloud computing. pp 13–16
149. Li B, He Q, Chen F, Jin H, Xiang Y, Yang Y (2020) Auditing cache data integrity in the edge computing environment. IEEE Trans Parallel Distrib Syst 32(5):1210–1223.
150. Li B, He Q, Chen F, Jin H, Xiang Y, Yang Y (2021) Inspecting edge data integrity with aggregated signature in distributed edge computing environment. IEEE Trans Cloud Comput 10(4):2691–2703.

Publisher's Note