A Survey on Identity-Based, Blockchain-Assisted, and Adaptive Data Integrity Auditing for Cloud Storage

Gauri Bobade[1], Prasad Koyande[2], Pradeep Shirke[3], Sunil Dodake[4]
[1,2,3,4]*Vidyalankar Polytechnic*

Abstract— **In the past decade, the advent of cloud storage has changed how individuals and businesses handle their data; however, outsourcing critical data to third-party physical infrastructure raises concerns about integrity, privacy, reliability and accountability of that data. Methods such as Remote Data Integrity Checking (RDIC), Provable Data Possession (PDP), Proof of Retrievability (PoR) and privacy-preserving public auditing have all developed as popular means for verifying integrity of offsite data without having to download the complete file. Initially, Research on auditing of offsite data was based on traditional public-key methods. As research continued, it shifted to identity-based methods which reduced the burden of certificates. The next avenue of research is blockchain-aided auditing systems that allow for transparency in the audit process as well as provide the ability to prove that an audit occurred and was not tampered with. Finally, adaptations of existing auditing methods have resulted in new, faster techniques for detecting breaches of integrity through intelligent anomaly detection. This paper synthesizes the awakening research results to create a unified view of all past research. In addition, this paper proposes a classification system that categorizes auditing protocols into four levels: traditional auditing protocols (with or without identity), blockchain-aided audit protocols, adaptive audit protocols and intelligent audit protocols. The auditing methods will be compared and contrasted on six major categories: privacy, verifiability, support for dynamic data, decentralised audit procedures, overhead costs and audit transparency and audit responsiveness to attacks. The paper concludes with discussing several unresolved research challenges which include developing cloud auditing methods that assume a 'zero trust' model e.g., multi-cloud environments, methods for explainable anomaly detection, developing a means for interoperable auditing layers and developing a standardised means for the evaluation of all types of auditing systems across the entire body of literature. This effort is intended to provide a sound basis for performing trustworthy intelligent and scalable data integrity verification in a cloud environment in the near future.**

Keyword— Cloud storage, data integrity auditing, RDIC, PDP, PoR, identity-based cryptography, blockchain auditing, anomaly detection, zero-trust cloud security, multi-cloud auditing.

I. INTRODUCTION

Large-scale data storage, enterprise information management, and content delivery have transitioned to using cloud computing as the primary infrastructure due to its cost-effectiveness, scalability, and elasticity. However, once data is stored by a cloud provider, the data owner no longer has direct, physical control over the storage medium, and must rely upon the honesty and capabilities of the cloud provider to ensure that the data will remain intact over time. Problems arise from this trust assumption because a cloud provider could be: honest but curious; selectively malicious; operationally careless; or economically motivated to hide storage loss; delay maintenance; and/or secretly destroy data that is not accessed often.

To address this problem, cryptographic auditing methods were developed by the research community so that a client or public verifier can check the integrity of their outsourced data without requiring them to download the entirety of the document. Existing models include: Provable Data Possession (PDP); Proof of Retrievability (PoR); and privacy-preserving public audit; all of which together define and support a challenge-response paradigm for efficiently verifying outsourced data.

Eventually, these methods were extended to accommodate dynamic data, identity-based verification, decentralized management of evidence, and adaptive monitoring facilitated by behavioral intelligence.

The literature is fragmented; while some papers focus on formal cryptographic guarantees, others concentrate on using blockchains to provide visibility into transactions, and recent papers address topics related to AI-based anomaly detection or zero trust approaches to decision-making. There is relatively little integrated discussion about how these various topics relate to each other, where they overlap, and where there are still technical gaps that need to be filled in order for next generation cloud auditing frameworks to be developed.[6][7][9] This survey will provide an overview of the existing body of knowledge, organize it into a unified taxonomy, and identify the architectural and research transitions in the development of cloud integrity auditing through history.

II. BACKGROUND AND CORE CONCEPTS

Auditing data integrity in cloud storage has many definitions based on the process and scope of verification. Most references give descriptions of verifying that an external data source has kept data complete, accurate, and retrievable without having to download all the data. In order for a data owner or their representative (the verifier) to perform an audit on data, the data owner will first define a file in terms of metadata or tags, and then a verifier will issue a challenge to the cloud provider to verify the data via a challenge-response process whereby the provider provides proof and the verifier checks the proof against public or private information.[2][3]

Many related terms frequently appear in the literature, and include:

- PDP (Provable Data Possession) verifies that a provider is able to provide the data from a client, even if the provider cannot prove that the complete data can be retrieved.[3]
- PoR (Proof of Retrievability) validates that a complete file can be retrieved, including possible errors due to incomplete files or missing data using structures that add error correction or redundancy.[3]
- RDIC (Remote Data Integrity Checking) provides a wide definition for any type of protocol that allows for verifying remotely stored cloud data to maintain efficiency in communication and privacy.[1]
- Public Auditing allows a verifier that is not associated with a data provider to provide data integrity testing without gaining access to the actual files associated with the test.[2]
- Identity-Based Auditing uses identity to create keys from an identity so that the overhead associated with Public Key Infrastructure (PKI) and managing PKI certificates is minimized.[1][14]
- Blockchain-Assisted Auditing uses distributed ledger technology as a secure and traceable storage solution for the proofs, logs, and audit-produced metadata.[4][12][3]
- Adaptive Auditing creates a change in the verification method based on factors such as prior history of violations, risk associated with the provider, or external risks.[5][10]

III. CLOUD INTEGRITY AUDITING DEVELOPMENT

The development of cloud integrity auditing can be explained through four general phases of cloud-based integrity auditing, from an original focus on cryptographic verification to a more distributed, behaviour-based and policy-driven method of auditing.

The first phase of cloud integrity auditing, the traditional cryptographic auditing phase, used the proofs of data possession (PDP), proofs of retrievability (PoR) and privacy-preserving public auditing (PPPA) methods. The primary goal of this phase was the generation of efficient proofs for the integrity of data stored in the cloud, as well as low communication cost for delivering that proof.

Phase two introduced support for dynamic and identity-based schemes, improving the ability to update data stored in the cloud and reduce the number of certificates associated with those updates, simplifying the trust infrastructure required of their use.

The third phase incorporated the use of blockchain as a method to enhance the transparency, non-repudiation, distributed trust, and tamper-evident logging of cloud audits.

The fourth, and current phase, of cloud integrity auditing integrates techniques for cloud audits with anomaly detection, zero-trust policy models, and adaptive scheduling that allows the audit to adjust its intensity based upon observed risk rather than at predetermined intervals.

These phases illustrate a shift away from traditional audit verification methods, which previously considered audits simply as static mathematical checks, to newer audit verification techniques, which view the audit in the context of a larger cyber-trust ecosystem incorporating long-term evidence, behaviour learning, and policy adaptation.

In summary, because of this evolution of auditing methodologies, future auditing research will be more interdisciplinary combining cryptography, distributed systems, machine learning, and cloud security operations.

IV. TAXONOMY OF EXISTING APPROCHES

This survey groups the literature into four major categories.

A. Traditional Public Auding

Conventional public audit systems provided a starting point for out sourced data verification. Their advantages are: they are clearly defined for challenge and response, offer the potential for strong privacy preserving audit verifications, and provide cost effective communication (as compared to full file retrieval). On the other hand, many of these systems require either a trusted third-party auditor or a static verification environment, and do not contain decentralized evidence or the ability for intelligent adaptations.

B. Identity-Based and Certificateless Auditing

Identity-based auditing systems provide a way to deploy PKI and validate certificates without the typical burden involved with both. Key management is simplified in cloud environments (where identity-based auditing schemes are an attractive option) through deriving cryptographic keys from the user's identity or from systems to which the user belongs. Thus identity-based systems allow public verification, while still providing a level of privacy.

However, identity-based auditing schemes still rely on trusted identity authorities, may not support decentralized trust, and in many cases have static audit behavior that does not adapt to the evolving risk of service providers.

C. Blockchain-Assisted Auditing

Blockchain-assisted audit methods use distributed ledgers for the storage of audit trail information, proof digests, and challenge metadata; they do so in order to improve the tamper-resistance, traceability, and accountability of the audit process. Blockchain-assisted audit methods are intended to alleviate one of the major weaknesses of centralized audit systems – the ease with which the historical record can be altered or disavowed. However, the act of using blockchain will not automatically create improved audit detection capabilities. In the absence of adaptive logic, using blockchain simply increases the amount of evidence preserved at the expense of increased storage requirements and transaction overhead.

D. Adaptive and Intelligent Auditing

Adaptive auditing incorporates context-sensitive decision-making into the integrity-checking process. For example, anomaly detection, zero-trust scoring, and behavioral monitoring can dictate when and how audits are performed.[5][10][11] This area of research is still in its infancy compared to traditional cryptographic methods, making its potential for faster abuse detection and more effective allocation of audit resources especially exciting. Because of these benefits, adaptive auditing will likely be especially beneficial in multi-cloud and heterogeneous environments.[9][5]

V. COMPARATIVE ANALYSIS

The review shows that none of the categories examined function best across all dimensions. Traditional and identity-based schemes perform well on efficiency and cryptographic clarity, whereas blockchain solutions support the integrity and accountability of evidence; however, responsive and operationally intelligent auditing is enhanced through the use of adaptive techniques.[1][4][5] Therefore, the overall direction of research in this area points to increased convergence, or combining assurance resulting from cryptographic rigor with non-replaceable evidence and intelligent policy controls.

Approach Category	Privacy Preservation	Public Verifiability	Dynamic Data Support	Decentralized Trust	Audit Transparency	Adaptive Behavior	Typical Limitation

Traditional public auditing	Strong [2]	Strong [2]	Moderate [13]	Weak [2]	Limited [2]	None [2]	Trusted auditor, static policy
Identity-based auditing	Strong [1]	Strong [1]	Moderate [1]	Weak to moderate [1]	Limited [1]	None [1]	Dependence on PKG, static scheduling
Blockchain-assisted auditing	Moderate to strong [4]	Strong [4]	Moderate [12]	Strong [4]	Strong [4][7]	Limited [4]	Storage/transaction overhead
Adaptive intelligent auditing	Variable [5]	Emerging [5]	Potentially strong [11]	Moderate to strong [5]	Strong when combined with blockchain [5][7]	Strong [5][10]	Evaluation, immaturity, explainability

VI. IDENTITY-BASED RDIC IN DETAIL

Identity-based RDIC forms one of the primary classes of literature relating to adaptive auditing because these approaches help to eliminate or reduce deployment barriers that would otherwise exist due to the PKI-based verification processes.[14] The process by which public identities of users and entities are used to derive keys eliminates the need for significant time and costs associated with distributing and verifying certificates, thereby providing a much simpler means of creating and managing identity-based systems. This makes identity-based RDIC particularly attractive in cloud architectures, where there may be many users, services and auditors spanning possibly numerous organizations that have a requirement to work together.

The primary benefit gained from the use of identity-based RDIC systems is that they allow for the maintenance of privacy-preserving public verification in conjunction with a more efficient trust infrastructure.[1][15] However, the introduction of this kind of system also requires new design considerations. For instance, the private key generator of an identity-based RDIC system becomes an important trust anchor; however, this design approach does not inherently solve issues such as the need for audit transparency, distributed non-repudiation, or dynamic behavior adaptation.[1][14] This explains the trend for someone working on developing ID-based RDIC systems to continue treating them as a solid cryptographic foundation to build upon rather than to do away with altogether.

VII. BLOCKCHAIN FOR CLOUD INTEGRITY AUDITING

The need for auditing capabilities for cloud data to provide evidence of integrity and account for audit activities motivated the use of blockchain technology as the basis for providing this capability.[4][12][7] Auditable cloud data produced using typical architectures allows untrusted auditors or colluded auditors to suppress evidence of failed results from proofs even if they are mathematically sound and to remove evidence of previous failures from the audit history.[4][12] The use of a distributed ledger for the purpose of recording and storing audit activities and the results of verification of records and supplying the digest of proofs provides the ability to protect all audit-related data from alterations and facilitates audit trail reconstruction using only the distributed ledger without using other verification methods.[4][12] The most significant advantages of blockchain technology in the auditability of cloud data are transparency, non-repudiation, and decentralized trust.[4][7][8] However, there are also significant disadvantages to using blockchain technology for cloud auditability, such as the costs associated with the additional storage requirements, logging complexity, execution overhead of smart contracts, and the scalability challenges associated with high frequency audit workloads.[4][8] As a result, the literature suggests treating the blockchain as an evidence and coordination layer to help validate cloud data rather than as the sole source of evidence for cloud auditability.[12][3].

VIII. ADAPTIVE AUDITING AND ANOMALY DETECTION

Anomaly-detecting and zero-trust auditing approaches represent an outstanding paradigm shift in how cloud environments are verified for integrity.[5][10] Rather than performing rigid audits with fixed schedules against each cloud provider and its data set, these adaptive audit schemes work to allocate auditing resources based on contextual risk factors such as prior audit failure, anomalous latency, suspicious access activity, repeated near-miss audit failures, and inconsistencies amongst multiple cloud providers in the audited period.[10][11][16]

This trend has multiple benefits: (a) it reduces the time it takes to detect stealthy or prolonged integrity violations;[5][10](b) as resources are focused more effectively on higher-risk cloud providers, less auditing effort is wasted on the lower-risk cloud providers;[11][9](c) it provides a natural extension to the zero-trust security model because trust is no longer a permanent attribute of any given entity based on behaviour within the cloud environment but is continually updated via a behaviour-based approach to audit every entity within the cloud environment.[5] The primary impediment is the lack of rigor in evaluating the various adaptive auditing methodologies; such prescriptive methodologies are missing standard data sets, do not have a community of practice with respect to performance benchmarking, and do not provide any expectation of replicability, making direct comparisons between auditing methodologies nearly impossible.[9][10]

IX. OPEN CHALLENGES

Several research challenges remain unresolved across the literature.

A. Zero-Trust Multi-Cloud Auditing

A majority of the existing auditing frameworks continue to be evaluated under simplified environments consisting of only one cloud provider, yet in the real world, it is very common to find cloud deployments that use more than one provider and that employ multiple different levels of trust amongst those providers.[12][8] The next generation of auditing applications must support the ability to correlate integrity across multiple providers, to jointly model risk across multiple architectures, and to provide for an interoperable auditing framework to manage the audits of multiple providers.

B. Explainable Anomaly Detection

In order for auditors to trust that adaptive auditing is providing them with accurate information regarding the integrity of cloud data, they require clarity about how the anomaly score determines a given data set and the audit challenges posed to a particular cloud provider.[5][16] To that end, there exists a need for explainable anomaly detection methodologies that enable auditors to understand the processes through which the challenges have been raised to the level of density, thereby increasing the overall risk rating assigned to a cloud provider.

C. Standardized Evaluation Benchmarks

Benchmarking in the literature has consistently yielded inconsistencies in performance metrics, datasets, and attack scenarios, hindering further research.[9][6] A common set of benchmarks would allow for improved methodological rigor and reproducibility.

D. Lightweight Blockchain Integration

Blockchain technology can improve the auditability and trustworthiness of data produced from cloud platforms but due to the limited resources of a high volume cloud auditing, the overhead associated with blockchain usage can outweigh the benefits.[4][8] Future work should consider implementing lightweight logging processes in combination with selective, optimized on-chain commitment processes, as well as establishing a hierarchical method for storing audit data.

E. Integration of Cryptography and Intelligence

Most of the papers reviewed still treat cryptographic verification processes as a separate concern from AI-based monitoring processes.[5][10] A large area of opportunity exists in creating frameworks in which the existence of anomaly-related evidence can directly alter or influence audit policy, while cryptographic proofs can preserve formal proof guarantees.

X. FUTURE RESEARCH DIRECTIONS

The next generation of research focused on cloud integrity auditing will most likely evolve towards

integrated, zero-trust, dynamic models. Identity-based models will continue to be used for reason that their deployment does not present complexities, and blockchain will remain relevant for the preservation of evidence and decentralization of trust.[1][4] The most promising area to develop closed-loop architectures where risk ratings for anomaly events will define audit frequency, challenge density and rules for escalation.[5][11] This type of architecture will be particularly useful in multi-cloud, edge-cloud and distributed enterprise settings because of the heterogeneous and dynamic nature of their attack surfaces.[9][16] Another potential area for development is to develop evaluation frameworks and survey-derived taxonomies that provide both a connection between the aspects of cryptographic security, system overhead, and AI explainability. The reason this will be significant is that future research done in relation to auditing will not merely be based on provable security characteristics but will also be based on aspects related to operational realism, transparency in trust, and the ability to deploy systems at scale.[9][6][8]

XI. DISCUSSION

The body of literature indicates that cloud data integrity auditing has evolved from being strictly an area of cryptography and into that of Systems, including Distributed Trust, Policy Governance, Audit Evidence Management, Anomaly Detection and Operational Response.[4][5][3] Traditional Audit Methods are still the basis of Cloud Data Integrity Auditing because they provide solid models of Challenge and Response Verifications; however, they are becoming less effective on their own in today's Cloud-based Ecosystems.[2][1] Identity-based solutions make deployment easier, Blockchains provide improved Transparency and Non-repudiation, and Adaptive Solutions improve Responsiveness and Resource Allocation.[1][4][5]

A well designed, next-generation framework will likely include the best of all three properties instead of focusing on just one. As such, research into this domain is important now; the domain has reached sufficient maturity so that systematic comparisons can be made, but has not matured sufficiently for a well-defined taxonomic structure and synthesis to make a significant contribution.[6][7][8]

XII. CONCLUSION

This survey has outlined the evolution of cloud data integrity auditing from classical public audits and RDIC to Identity-Based Verification (IBV), to Blockchain-based Transparency (B-T) to Adaptive Anomaly-Aware Models (AAAM). It has proposed a taxonomic structure, compared/mapped the major categories across Technical and Operational dimensions, and identified the current gaps/challenges associated with zero trust multi-cloud auditing, explainable anomaly detection, interoperable evidence management, and standardized evaluation. Based on the findings from this survey, the future of assurance in Cloud Integrity will be in the convergence of Efficient Cryptographic Verification, Immutable Distributed Evidence and Adaptive Intelligence Driven Policy Control.[5][4][3]

REFERENCES

[1] Yong Yu, Man Ho Au, Giuseppe Ateniese, Xinyi Huang, Willy Susilo, Yuanshun Dai, and Geyong Min, "Identity-Based Remote Data Integrity Checking With Perfect Data Privacy Preserving for Cloud Storage," IEEE Transactions on Information Forensics and Security, vol. 12, no. 4, pp. 767–778, 2017.

[2] Cong Wang, Sherman S. M. Chow, Qian Wang, Kui Ren, and Wenjing Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage," IEEE Transactions on Computers, vol. 62, no. 2, pp. 362–375, 2013.

[3] Blockchain-based Public Auditing for Big Data in Cloud Storage, 2020.

[4] Minglei Song, Zhaohui Hua, Yihong Zheng, Yuliang Zhou, and Xiaofeng Jia, "Blockchain-Based Deduplication and Integrity Auditing Over Encrypted Cloud Storage," IEEE Transactions on Dependable and Secure Computing, 2023.

[5] A Comprehensive Survey on Edge Data Integrity Verification, 2024.

[6] A Methodical Literature Survey on Block Chain-based Public Auditing in Cloud, 2023.

[7] A Methodical Literature Survey on Block Chain-based Public Auditing in Cloud, IEEE, 2023.

[8] A Survey on Blockchain-based Integrity Auditing for Cloud, 2025.

[9] A Comprehensive Survey on Edge Data Integrity Verification, 2024.

[10] An Improved Mechanism to Maintain Data Integrity and Anomaly Detection in Cloud Environments, 2026.
[11] A Proposed Method for Real-Time Automatic Cloud Storage, 2025.
[12] Enhancing anomaly detection and prevention in Internet of Things ...," 2025.
[13] 13. Kun Yang and Xiaohua Jia, "An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing," IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 9, pp. 1717–1726, 2013.
[14] Enhancing Cloud Data Security with Identity-Based Remote Data Integrity Checking," 2024.
[15] Sathyabama A. R. and Jeevaa Katiravan, "Enhancing anomaly detection and prevention in Internet of Things (IoT) using deep neural networks and blockchain based cyber security," Scientific ReportsScientific Reports, vol. 15, no. 1, article 22369, Jul. 2025, doi: 10.1038/s41598-025-04164-4.
[16] Enhancing anomaly detection and prevention in Internet of Things, 2025.