Journal of Harbin Engineering University
ISSN: 1006-7043
Vol 44 No. 8
August 2023

A Comprehensive Review of Data Auditing Techniques in Cloud

D.Sophia Navis Mary[PT] , Dr. S.Gopinathan
Dept. of Computer Science,University of Madaras
Ethiraj College for Women(Autonomous),Chennai

Abstract- In today's world the cloud storage is prevalent and an influential business model communication used by every user in the internet. As the envisioned next-generation of computing architecture for small, medium and large enterprises with the benefits of massive storage, cost and scalability. The cloud service providers depend on Third Party Auditors to share the data securely. Data auditing schemes are the integral part of the cloud storage to provide secure data sharing. In order to enable safe data sharing between data owners and users. This paper describes how various techniques are essential to ensuring the trustworthiness and compliance of cloud data.
Keywords: Data auditing, cloud computing , data sharing. Data owners, Third Party Auditors

I. Introduction

As cloud computing continues to evolve, advancements in data auditing techniques will be crucial in maintaining trust and confidence in cloud data storage and management. Techniques for data auditing are crucial for guaranteeing the accuracy and security of cloud data. The techniques discussed in this article, including Provable Data Possession, Homomorphic Auditing, Zero-Knowledge Proofs, Blockchain-based Auditing.

In these schemes cryptographic constructs play a vital role to guarantee the safety of the data. There are several data auditing techniques used cryptographic primitives including:

Digital Signatures: Digital signatures use cryptographic algorithms to create a unique and tamper-proof identifier for a dataset. The owner digitally signs the data using their private key, and anyone has the capability of verifying the signature by utilizing the public key of the owner. The signature ensures data integrity and authenticity.

Blockchain Technology: Techniques for data auditing are crucial for guaranteeing the accuracy and security of cloud data. Blockchain technology serves as a decentralized and distributed ledger system that facilitates the recording of transactions across multiple computers.Data can be kept in a blockchain and ownership can be established and audited through the consensus mechanism and cryptographic hashing algorithms. The immutable nature of blockchain ensures the integrity and transparency of ownership records.

Timestamping: Timestamping involves associating a specific date and time with a dataset. By using trusted timestamping services or utilizing blockchain-based timestamping, data owners can prove the existence of their data at a specific point in time. This proof helps establish ownership and prevent data tampering.

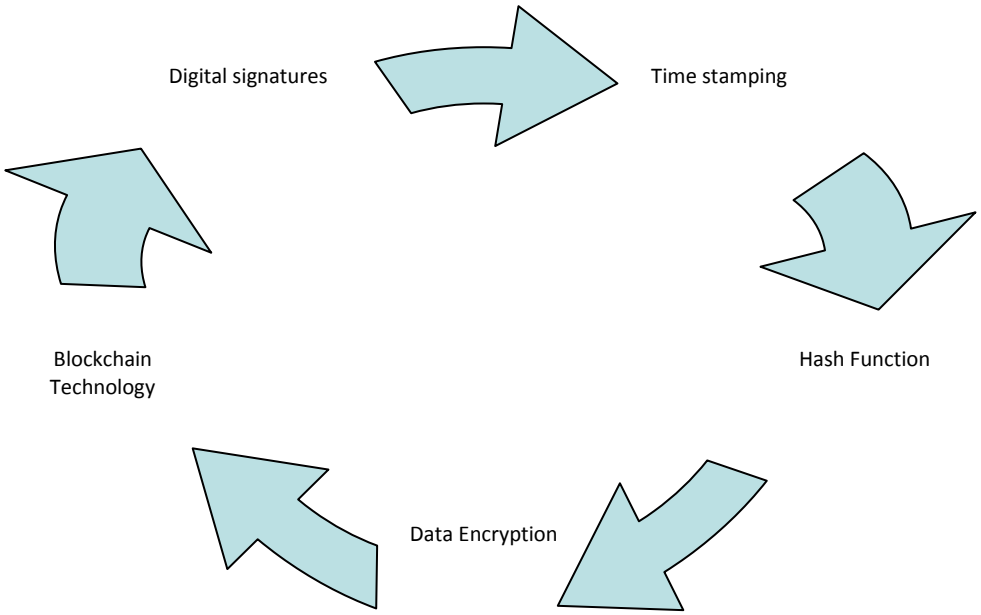

Fig 1. Primitives of Cryptographic systems

Data Hashing: Hash functions generate a unique string of characters (hash) based on the content of a file. By means of a comparative analysis between the hash value of the initial dataset and the calculated hash of the scrutinized data, proprietors of data are able to confirm the soundness of the information and establish their rightful possession of it.

Data Encryption: Encrypting data using a private key confirms that the owner of the data, can access and decrypt the data. This method provides a level of ownership proof by demonstrating exclusive access to the encrypted information.

Metadata and Audit Logs: Maintaining comprehensive metadata and audit logs can help establish proof of ownership by documenting the creation, modification, and access history of data. These records can be used to track ownership changes and establish a data trail for auditing purposes**.**

II. Data auditing Model

The data auditing process model in cloud encompasses of Cloud service provider, The individual or organization responsible for the possession of the information, alongside a qualified and independent assessor, commonly referred to as a third-party auditor, is a critical facet in data management and user. A person or an organisation has data that will be put in storage in the cloud. A user is someone who manipulates or views data. The service provider offers cloud storage so that users can save their data. The trusted third-party auditor who, upon request, routinely audits the remote data in the cloud. The statistics were also monitored and validated by the third party auditor, if he knows the public key used in the verification scheme.

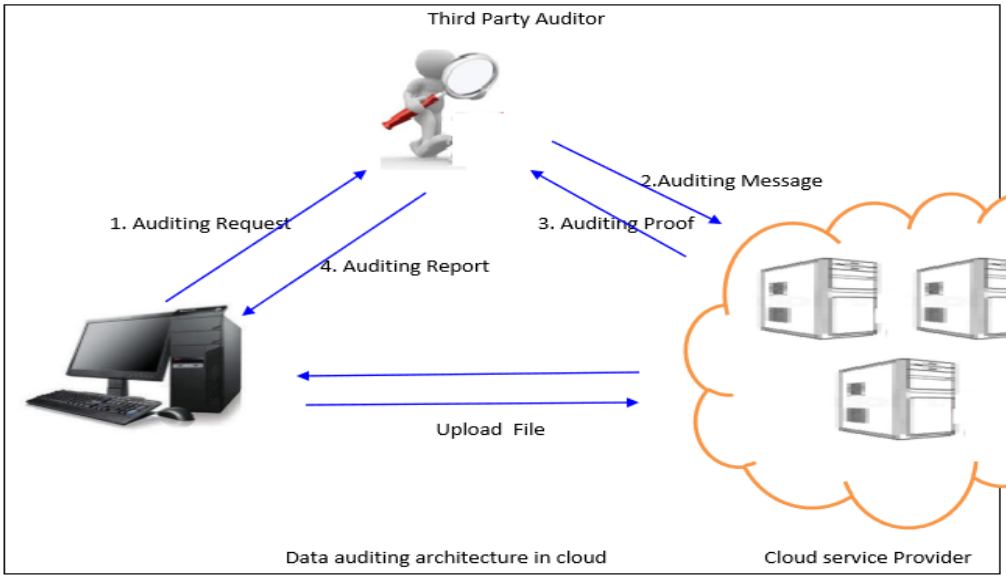

Fig 2 : Data Auditing Process System Model

III. Provable Data Possession (PDP)

PDP is a frequently used auditing approach that allows users of cloud services to verify the correctness of their data without completely retrieving it. The Provable Data Possession scheme is explained by step by step

Algorithm FileUpload(File,Metadata)

1. Split the file F into n blocks (f1,f2,..,fn)
2. Generate Metadata for the blocks in the file by pre-processing f1',f2',f3',..fn' stored at client
3. Upload the file into server after pre-processing
4. Remove the local copy from the client

Algorithm FileVerify(Challenge)

1. The random challenge raised by the client to verify the file.
2. calculate the answer to the challenge of the client by the server
3. Send the response to the client
4. Verify using Metadata stored locally and assure the response

PDP[1] relies on cryptographic proofs and challenge-response mechanisms towards safety that the cloud provider is storing the data as promised. By generating and verifying short integrity proofs, users can detect any unauthorized modifications or data loss. Ateniese et al. [2] suggested a PDP model where The file F is divided into blocks (B1, B2,..., Bm) by the data owner, who also produces metadata (a tag) for each block that can be used for identify the data block. The file is subsequently delivered to a remote/cloud server for storage. When the verifier issues challenges, the cloud server reacts and offers evidence that the material has not been altered or partially wiped from the storage. As stated in [2], the setup and verification portions of the PDP method rely on an effective symmetric-key algorithm.

Homomorphic Auditing enables cloud users to delegate data auditing tasks to the cloud service provider without disclosing the actual data content. It leverages homomorphic encryption schemes that allows calculations to be performed on encrypted data. By encrypting the data and providing auditing functions on encrypted data, cloud users or organizations have the ability to guarantee the confidentiality and authenticity of their information. Linear authenticators, which are verification metadata constructed from file blocks, serve the purpose of validating an accurately computed linear combination of said blocks by verifying only the aggregated tag/authenticator. This is done in

order to persuade the verifier of the combination's Erway, A. Kupcu et al. [3] describes dynamic provable data possession (DPDP), is a publicly verifiable scheme supports dynamic data modification in the cloud storage.

Distributed Provable Data Possession extends the traditional PDP technique provision to data auditing in a distributed and multi-cloud environment. DPDP allows data to be distributed across multiple cloud providers while ensuring data integrity. It employs cryptographic techniques and challenge-response protocols to verify the possession of data by multiple cloud servers.

Zhu. et al.[5] scheme enabling the validation of the dynamic information stored in the cloud by a third-party auditor (TPA) on behalf of the client. In order To ensure the veracity of the dynamic data housed within the cloud on behalf of the client, it is necessary to confirm its accuracy. the TPA eliminates the client's involvement by assessing whether his data is indeed intact when stored in the cloud. Since cloud computing services go beyond backup data alone, support for data manipulations in file such as block updating, delete and adding a data block.

Wang, H. Li, et al [4] in their work, Where TPA may perform several auditing duties concurrently assignments. In order to successfully provide public auditability, the homomorphic authenticator does not require access to the data blocks themselves. This system supports auditability and dynamic data operations. To aid in the effective handling of various auditing duties, the approach extending bilinear aggregate signature, wherein the Third-Party Auditor (TPA) can execute numerous auditing tasks simultaneously, is also scrutinized. The provision of public auditability in an effective manner without requiring physical access to the blocks is also explored. This approach employs homomorphic authenticators coupled with individual data blocks to furnish unassailable information.

Zero-Knowledge Proofs (ZKP) provides a powerful technique for data auditing in cloud environments while maintaining data confidentiality. ZKPs allows the prover's task is to persuade the verifier of the veracity of a given assertion without imparting any supplementary information. Data owners can publish proofs regarding their data without actually sharing the data by using Zero Knowledge Proofs. Discrete logarithm ZKPs With the presented approach [5] proposes a key-updating and authenticator-evolving technique that incorporates zero knowledge and authenticator updates to ensure secure cloud data auditing through zero knowledge evidence of the stored files. Without revealing the real data, it can be used to demonstrate that data storage and processes are correct.

authenticity.

IV. Proof of Retrievability (PoR)

Proof of Retrievability (PoR) is an approach to examine the availability and integrity of data stored in the remote . It ensures that the data remains intact and can be retrieved successfully when needed. PoR provides assurance to data owners that their data has not been tampered with or lost in the cloud environment. PoR involves the generation of a compact proof that can be used to verify the integrity of the stored data. This proof serves as evidence that the data has not been modified or corrupted since its original storage. It also verifies that the data can be retrieved successfully without any loss.

The process of implementing PoR typically involves the following steps:

Data Preparation: The data owner splits the original data into smaller blocks or chunks, which are then encrypted for confidentiality and integrity. Each chunk is associated with a unique identifier.

Data Outsourcing: The encrypted chunks of data are dispatched to the provider of cloud storage for preservation. In addition to the data, the possessor of the data forwards a group of metadata, including the unique identifiers and cryptographic keys required for verification.

Challenges and Proofs: To ensure the data's retrievability and integrity, the proprietor of the data regularly prompts the cloud-based storage provider to furnish verification of the data fragments. still intact and retrievable. These challenges can be in the form of random requests for specific data blocks.

Proof Generation: Upon receiving a challenge, the cloud storage provider performs the necessary

computations to generate a proof that demonstrates the integrity and availability of the requested data Proof Verification: The data owner authenticates the received proof to ensure its validity. This involves using the metadata and cryptographic keys provided during the data outsourcing phase. By comparing the proof with the expected results, the proprietor of data can determine whether the stored data remains intact and can be retrieved successfully.

V. Blockchain Based Auditing

Blockchain technology has gained significant attention in recent years, and it can also be utilized for data auditing in cloud environments. Blockchain-based auditing ensures transparency, immutability, and tamper-resistance of the auditing process by storing audit logs and information on a distributed ledger. It enables users of the cloud to check the accuracy of their data on their own and identify any unauthorised changes performed by the cloud provider. The data auditing does not include the third party auditor.

The authors have developed a scheme[20] with the intention of mitigating the cost of computation and communication for integrity verification. The scheme involves recording minimal verification tags on the blockchain by data owners, which are then utilized to produce a proof through the building of a Merkle Hash Tree with the said hashtags.

block. This proof is then sent back to the data owner.

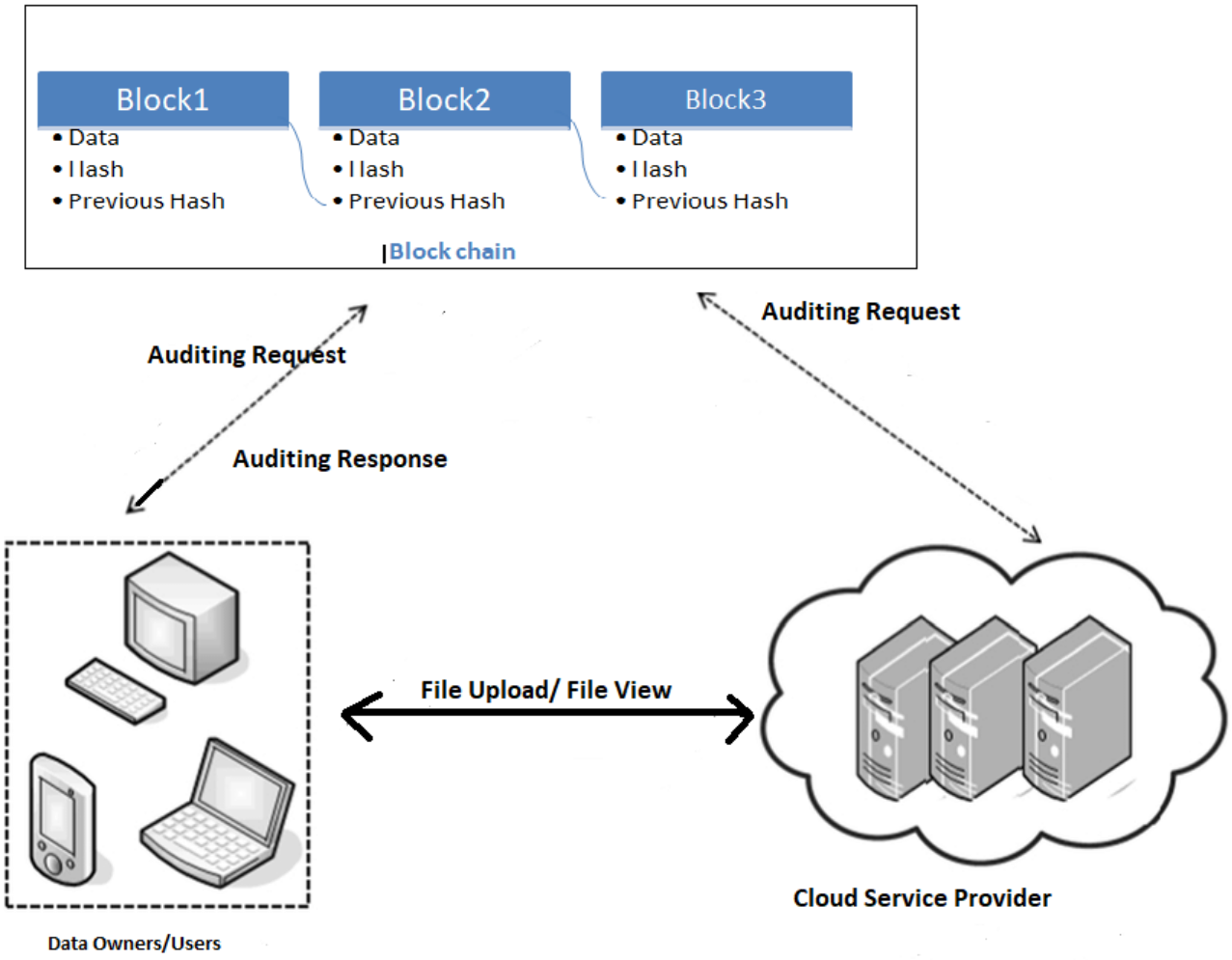

Fig 3. Block chain Based Data auditing Model

In data auditing scheme[21] the data owner relies on a third party authority to complete the process of data verification in a data auditing scheme is not require for IoT devices. .A decentralised and dynamic data environment is provided by blockchain-based smart contract services. And the framework that enables both data owners and data consumers to independently validate certain pieces of data without depending on a single external auditor.

Conclusion

Cloud data security is primarily concerned with the three factors of availability, integrity, and secrecy. A comprehensive investigation into data integrity verification methods that are more trustworthy and sophisticated as the number of clients grows. If the TPA is compromised or unable to function, data auditing systems cannot function. Additionally, the TPA is unable to handle several users' simultaneous demands for extensive auditing.

References:

[1] Giuseppe Ateniese, et al "Provable Data Possession at Untrusted Stores" 2007,IACR Cryptol. ePrint Arch.pg202

[2] G. Ateniese, R.D. Pietro, L. V. Mancini, G. Tsudik ,"Scalable and Efficient Provable Data Possession", Proc.4th International Conference on Security and Privacy in Communication Networks, 2008.

[3] C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia, (2009) "Dynamic Provable Data Possession," Proceedings of ACM CCS 2009, pp. 213–222.

[4] C. Wang, Q. Wang, K. Ren, and W. Lou, (2011)"Towards Secure and Dependable Storage Services in Cloud Computing," IEEE Transactions on Services Computing, vol. 5, no. 2, pp. 220–232, 2011.

[5] Y. Zhu, G.-J. Ahn, H. Hu, S. S. Yau, H. G. An, and S. Chen, "Dynamic Audit Services for Outsourced Storage in Clouds," IEEE Transactions on Services Computing,

[6] Yu, Y., Li, Y., Au, M. Ho., Susilo, W., Choo, K. & Zhang, X. (2016) "Public Cloud Data Auditing with Practical Key Update and Zero Knowledge Privacy" Australia Information Security and Privacy - 21st Australasian Conference, ACISP 2016, Proceedings

[7] A. Juels and B. Kaliski. PORs: Proofs of retrievability for large files. ACM CCS'07

[8] AyadF.Barsoum and M.AnwarHasan, Provable Possession and Replication of Data over Cloud Servers - Centre For Applied Cryptographic Research (CACR), University of Waterloo, Report 2010/32, 2010, http://

[9] www.cacr.math.uwaterloo.ca/techreports/2010/cacr2010-32.pdf

[10] Blundo C, Cimato S, De Capitani di Vimercati S, De Santis A, Foresti S, Paraboschi S, et al. Efficient key management forenforcing access control in outsourced scenarios. In: Gritzalis D, Lopez J, editors. Emerging challenges for security, privacyand trust, IFIP advances in information and communication technology, 297. Boston: Springer; 2009. p. 364–75.

[11] Deyan Chen and Hong Zhao. Data security and privacy protection issues in cloud computing. In Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on, volume 1, pages 647–651, 2012.

[12] Erway, Chris, et al. "Dynamic provable data possession." Proceedings of the 16th ACM conference on Computer and communications security. Acm, 2009.

[13] Dynamic remote data auditing for securing big data storage in cloud computing2013.

[14] Sookhak, Mehdi, et al. "Remote data auditing in cloud computing environments: a survey, taxonomy, and open issues." ACM Computing Surveys (CSUR) 47.4 (2015): 65.

[15] Sookhak, Mehdi. Dynamic remote data auditing for securing big data storage in cloud computing. Diss. University of Malaya, 2015.

[16] Wang, Q., Wang, C., Ren, K., Lou, W., & Li, J. (2010). Enabling public auditability and data dynamics for storage security in cloud computing. *IEEE transactions on parallel and distributed systems*, *22*(5), 847-859.

[17] Wang, C., Ren, K., Lou, W., & Li, J. (2010). Toward publicly auditable secure cloud data storage services. *IEEE network*, *24*(4), 19-24.

[18] Sookhak, M., Talebian, H., Ahmed, E., Gani, A., & Khan, M. K. (2014). A review on remote data auditing in single cloud server: Taxonomy and open issues. *Journal of Network and Computer Applications*, *43*, 121-141.

[19] Tian, H., Chen, Y., Chang, C. C., Jiang, H., Huang, Y., Chen, Y., & Liu, J. (2015). Dynamic-hash-table based public auditing for secure cloud storage. *IEEE Transactions on Services Computing*, *10*(5), 701-714

[20] Jiaxin Li, Jigang Wu, Guiyuan Jiang, Thambipillai Srikanthan, Blockchain-based public auditing for big data in cloud storage, Information Processing & Management,Volume 57, Issue 6,2020,

[21] K. Gangadevi, R.R. Devi A survey on data integrity verification schemes using blockchain technology in cloud computing environment IOP Conference Series: Materials Science and Engineering, vol. 1110, IOP Publishing (2021), Article 012011

[22] B. Liu, X.L. Yu, S. Chen, X. Xu, L. Zhu Blockchain based data integrity service framework for IoT data 2017 IEEE International Conference on Web Services (ICWS), IEEE (2017), pp. 468-475

[23] J. Li, J. Wu, G. Jiang, T. Srikanthan Blockchain-based public auditing for big data in cloud storage Inf. Process. Manag., 57 (6) (2020), Article 102382

Author Detail

D.Sophia Navis Mary is working as assistant professor and part time PhD scholar in the Department of Computer Science at University of Madras. Her research area includes network security, information security and she is currently working as an Assistant professor in Ethiraj college for Women

.**S.Gopinathan** is working as a Professor and Head of the Department in the Department of Computer Science, University of Madras. He received Ph.D from University of Madras and M.Sc (Computer science) from Bharathiar University. His research area includes data mining, software engineering, digital image processing. He is peer reviewer for various International journals and Technical committee member for various national and International conferences. He is a Member of Board of studies in various Universities.