International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

# SECURE FILE STORING AND AUDITING

Dr.K.L. NEELA M.E., Ph.D.[1],

SIVAPRAKASH S [2], VISWESWARAN R[3]

Department of Computer Science and Engineering University College Of Engineering, Thirukkuvalai

(A constituent College of Anna University::Chennai and Approved by AICTE, New Delhi)

---

**ABSTRACT**

Consensus method and highly programmable smart contracts automatically perform secure data sharing for dynamic audits. One of the most significant developing technologies at the moment is blockchain technology. With regard to traditional cloud audit data in central storage, vulnerability, tampering, incomplete transmission, and unsafe data flow, its characteristics of decentralization, non-tampering, traceability, and high security may effectively address these issues. Blockchain technology can speed up data compilation and enhance the accuracy of audit findings. According to this concept, a Consortium blockchain is created for various auditors, and only users who have identity compliance through the node admission method are permitted to exchange cloud auditing data. For diverse audit data with varying ownership structures and levels of sensitivity, off-chain asynchronous safe storage is employed. In the proposed task, the auditor is required to start a new transaction for each verification, conduct the transaction, and incorporate the verification-related data into the transaction. The user can check the transaction's creation time to see when the auditor performed the verification after the transaction was added to the blockchain.

**KEYWORDS**: File Storage in Cloud, Blockchain Creation, Data Auditing, Third Party Authentication, TPA Verification.

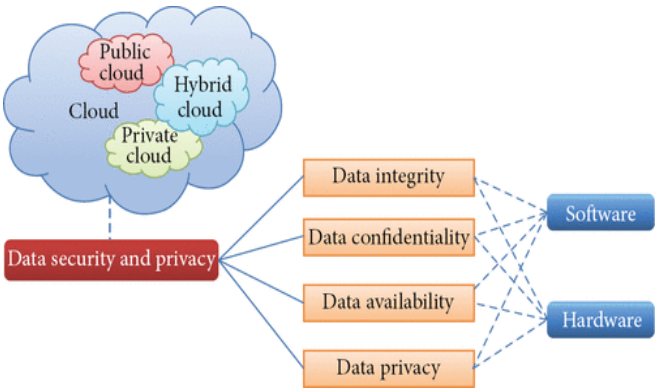
## 1. INTRODUCTION

As organizations increasingly rely on cloud storage and computing, ensuring the integrity and security of their files becomes more critical than ever. Cyber attacks, human error, and hardware failures can all lead to data corruption or loss, which can have severe consequences for businesses and their customers. To mitigate these risks, many organizations are turning to file integrity checking and monitoring solutions. These tools use various techniques to verify the integrity of files stored in the cloud, such as hash functions, digital signatures, and checksums. They can also detect unauthorized changes or access to files and provide alerts or notifications to security personnel. This report will explore the challenges of file integrity checking and monitoring in the cloud, the different techniques and technologies used to ensure data integrity, and the benefits and limitations of various solutions. We will also discuss the regulatory and compliance requirements that organizations must adhere to, and how file integrity checking and monitoring can help meet these requirements.

### Data Integrity

Data integrity is one of the most critical elements in any information system. Generally, data integrity means protecting data from unauthorized deletion, modification, or fabrication. Managing entity's admittance and rights to specific enterprise resources ensures that valuable data and services are not abused, misappropriated, or stolen. Data integrity is easily achieved in a standalone system with a single database. Data integrity in the standalone system is maintained via database constraints and transactions, which is usually finished by a database management system (DBMS). Transactions should follow ACID (atomicity, consistency, isolation, and durability) properties to ensure data integrity. Most databases support ACID transactions and can preserve data integrity.

Authorization is used to control the access of data. It is the mechanism by which a system determines what level of access a particular authenticated user should have to secure resources controlled by the system.




## 2. RELATED WORK

Li, Yannan, et.al,…[1] proposed a key updating and authenticator-evolving mechanism with zero-knowledge privacy of the stored files for secure cloud data auditing, which incorporates zero knowledge proof systems, proxy re-signatures and homomorphic linear authenticators. Here instantiate proposal with the state-of-the-art Shacham-Waters auditing scheme. When the cloud user needs to update his key, instead of downloading the entire file and re-generating all the authenticators, the user can simply download one single file tag, work out a re-signing key with the new private key and upload the new file tag together with some verification information to the cloud server, in which the user undertakes the least amount of the workload in the updating phase. Three kinds of entities are involved in the scenario, namely cloud users or data owners, the cloud server and a third party auditor (TPA). A cloud user generates data files and stores large amount of data on the remote cloud server without keeping a local copy. TPA can be an organization managed by the government, which has expertise and capabilities that cloud users do not have and is trusted to check the integrity of the hosted data on behalf of cloud users upon request. TPA is responsible for checking the integrity of the cloud data on behalf the cloud users in case that they have no time, resources or feasibility to monitor their data, and returns the auditing report to the cloud user.

Shen, et.al,…[2] implemented an efficient public auditing protocol with global and sampling blockless verification as well as batch auditing, where data dynamics are substantially more efficiently supported than is the case with the state of the art. Note that, the novel dynamic structure in this protocol consists of a doubly linked info table and a location array. The doubly linked info table (DLIT) is a two-dimensional data structure employed by the TPA to store data information concerning auditing, differing from the one-dimensional Index Hash Table (IHT). Data information in the DLIT is divided into two types: file information and block information. The left part is the file information, including user ID and file ID. In previous works by other researchers, the file information simply consists of the file ID, therein making the length of the file ID long. Moreover, when the total number of files increases over time, it becomes more difficult to make each file ID unique. Hence, here the concatenation of the user ID and file ID to identify each file, making the unique identifier much easier to find. For instance, if the file ID is 4 bits, 16 identifiers can be generated. However, even if a user ID of only one bit is added into the process of identifier generation, the number of identifiers can be doubled to 32. In the real world, both the file ID bit and the user ID bit will be larger, and the number of identifiers will consequently be increased. The right part is the block information, including the current version number and the time stamp, which are generated when a given block is uploaded or updated. With such a double linking data structure, the insertion and deletion of a file or data block will no longer cause a change in other records in the DLIT. Moreover, the advantages of the DLIT will be reflected in batch operations at lower costs when searching for a certain element.

Shen, et.al,…[3] presented a new paradigm named remote data possession checking with privacy-preserving authenticators for cloud storage. In this new paradigm, both cloud service provider and the public verifier do not have access to the real authenticators (signatures) for cloud data. Meanwhile, the integrity of cloud data is still able to be efficiently checked. It is potentially useful in some special situations where electronic checks and contracts are outsourced. To securely protect the privacy of the authenticator, we design a new authenticator called Homomorphic Invisible Authenticator (HIA), which protects the privacy of authenticator and supports the blockless verification. Based on HIA, we construct the first remote data possession checking scheme with privacy-preserving authenticators for cloud storage. There are four types of entities involved in the framework: the cloud user, the cloud, the third party auditor (TPA) and the trusted authority (TA). The cloud user owns large amount of data that will be outsourced into the cloud. The cloud provides enormous storage space for the

cloud user, which is supervised by cloud service providers (CSPs). The TPA, who has considerable computation and communication capability, is delegated by the cloud user to check the data possession of the cloud. The TA is an organization trusted by both the cloud and the cloud user, which can be a public institution or a non-government organization. The public key of the TA is used by the cloud user to generate privacy-preserving authenticators for his outsourced data. Under rational requirements, the TA can recover the real authenticators from the privacy-preserving ones.

Wenting, Md Towfiqul, et.al,…[4] implemented a cloud storage auditing scheme for group users, which greatly reduces the computation burden on the user side. In this scheme, the Third Party Medium (TPM) to perform time-consuming operations on behalf of users. The TPM is in charge of generating authenticators for users and verifying data integrity on behalf of users. In order to protect the data privacy against the TPM, the blind data using simple operations in the phase of data uploading and data auditing. The user does not need to perform time-consuming decryption operations when using cloud data. The TPM cannot get the real data from the blinded ones of users in the phase of data uploading, and cannot derive the real data from the cloud's response in the phase of auditing. In a group, there are multiple group users and one original user who is the original owner of data and can create shared data to the cloud. After the original user uploads data to the cloud, other users in the group can access these shared cloud data. The original data owner can play the role of the group manager. When the user wants to upload data to the cloud, he needs to blind the data, and then sends them to the TPM. After receiving the blinded data from the user, the TPM will generate the corresponding authenticators for the blinded data, and then uploads the blinded data and the corresponding authenticators to the cloud together. The cloud recovers the real data and the real authenticators from the blinded data and the corresponding authenticators, and stores them. When the TPM wants to verify the integrity of cloud data, he will send an auditing challenge to the cloud. After receiving this challenge, the cloud will respond to this TPM with a proof of data possession. And then, the TPM will check the correctness of the proof to verify the integrity of cloud data.

Suguna, M., et.al,…[5] implemented a third party trusted verifier is introduced in the proposed work which maintains dynamic metadata stored locally for the verification process. Bilinear mapping is used to ensure verification without retrieving the original data called blockless process. The verification proof generated using the proposed method is a small signature, which reduces the auditing overhead at the client side compared to existing solution. A data auditing protocol is proposed for ensuring privacy of outsourced data at semi trusted storage by verifying data intactness through blockless verification. The third party trusted verifier (TV) is assigned the verification process in charge of the client thereby reducing the computation and storage overhead at the client side. In order to ensure the privacy of data against TV, blockless process is ensured where the challenge involves a metadata proof from the server which is verified locally by TV without the actual data. Even when the TV tries to get the data from encrypted file, the proof construction using bilinear map ensures the data privacy. In provable data possession schemes, the auditing process ensures the data intactness through challenge response method. The usage of a trusted third party verifier ensures low computation and storage overhead for the client. But, the privacy of data becomes a question as the verifier can derive the data from his history of challenge and proof in hand. Using blockless verification, the proposed P-DAP protocol ensures privacy of user's data against the auditor verification process. Also the user signing process in every data dynamic operation ensures the authenticity of updates. The storage server is also benefited by proving it's trustworthiness by responding with the proof as the challenge information is sent by the verifier.

## 3. EXISTING SYSTEM

In public verification schemes, the user establishes a verification period (i.e., the frequency with which the auditor conducts the verification) following data outsourcing. The auditor then confirms the integrity of the outsourced data at the appropriate time. In reality, the auditor creates a verification report with several verification outcomes (corresponding to multiple epochs, or periods). If the verification result is "Reject" for any period, it signifies that the data may be corrupted, and the auditor must immediately notify the user. Otherwise, at the end of each epoch, the auditor creates a verification log and gives it to the user. The user can designate the auditor to execute the verification with any period as needed since the auditor can check the data integrity without requiring the user's involvement. Although cloud auditing is somewhat more convenient for auditors, internal and external data security issues remain. Currently, the "cloud audit

platform" is in charge of centrally managing the collection, storage, transmission, distribution, and analysis of audit data. The public key infrastructure (PKI) is the foundation for the majority of public verification schemes, and it requires the auditor to manage the user's certificate in order to select the appropriate public key for verification. As a result, these schemes struggle with the expensive and time-consuming certificate management issue, which includes certificate revocation, storage, distribution, and verification.

The first certificateless public data integrity verification technique (CPVPA) that thwarts nefarious and tardy auditors was implemented in this proposed effort. Utilising blockchain technology, which offers a distributed and tamper-proof method of conducting transactions without a central authority (i.e., bank), is the main concept behind CPVPA. In CPVPA, the auditor is needed to initiate a new transaction after each verification, complete the transaction, and include the verification-related data into the transaction. The user can check the transaction's creation time to see when the auditor performed the verification after the transaction was added to the blockchain.

## 4. PROPOSED METHODOLOGY

The process of confirming the accuracy, integrity, and security of data kept on a cloud platform based on a blockchain is known as data auditing in a blockchain environment. Blockchain technology is a popular option for cloud storage solutions because it offers safe and open record-keeping. However, there is still a chance of data loss or tampering even with blockchain's built-in security mechanisms.

Blockchain cloud data auditing is inspecting the data to make sure it hasn't been altered or tampered with. This is accomplished by examining the hashes and cryptographic signatures of the data blocks kept on the blockchain. Blockchain cloud data auditing also entails keeping track of who has access to the data and making sure it is appropriately protected from unauthorized access. Compliance is a key component of data auditing in the blockchain cloud. Organizations must make sure that their blockchain cloud solutions adhere to all applicable standards given the proliferation of rules governing data privacy and security. Data auditing assists in locating any non-compliance problems and offers advice on how to resolve them. Overall, blockchain cloud data auditing is essential for guaranteeing the security, correctness, and integrity of data stored on a blockchain-based cloud platform. By assisting in risk mitigation and ensuring regulatory compliance, it raises the credibility and confidence of the blockchain cloud service. It emphasizes that the more users a blockchain system has, the higher security guarantee it can offer.

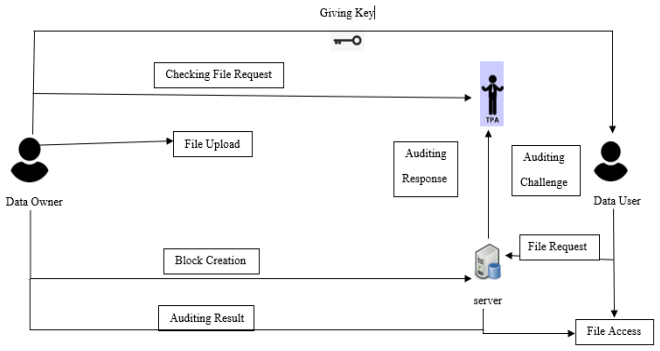



Fig 2: Blockchain Data Auditing Design

As a result, rather than building CPVPA on a brand-new blockchain technology, we choose one that is already established and widely used. The suggested plan also encourages public audits employing a TPA (Third Party Auditor) to assist clients with limited resources. The suggested method is more effective than the current schemes that are made to support deduplication and public auditing simultaneously, and it complies with all essential security requirements.

## 5.IMPLEMENTATION

### Block Chain Storage

To provide secure data storage and sharing, the data storage scheme makes use of cloud storage technology based on blockchain. Create a local cloud in this module and offer reasonably priced, ample storage services. The primary transactions in the medical blockchain are data storage and access control. The ability to store all medical data on the blockchain would be ideal. Once users have cloud storage, they can upload and exchange data there. Utilising block chain technology, cloud storage can be implemented in this task with excellent security. The cloud server, which offers cloud storage services, is controlled by the cloud service provider. It has a lot of computing power in addition to having a lot of storage capacity.

**User Enrolment**

A user must first register with a system in order to access it, after which he must be authenticated in order to request a server. In a basic authentication process, a user must present some credentials, such as a user ID, to prove that the user is the rightful owner of the user ID. Data owners could upload the file on the cloud, where it will be encrypted once it is stored.

**Data Block Creation**

A blockchain is a digital notion that is used to store data. Because these blocks are connected together, their data is immutable. When a data block is linked to the other blocks, its data cannot be modified again. It will always be publicly accessible to everyone who wants to see it, just as it was when it was uploaded to the blockchain. Blockchain technology functions are dependable for use in a hashing crypto method, which aids in the creation of an acceptable and strong hashing code and its conversion from a bit of fixed size data to character strings. Each transaction proposed in a blockchain is hashed together before being pushed into a block, and hash pointers connect each block to the next block for holding previous hash data as it is undeniable. As a result, any changes in the blockchain transaction of hashing function will result in a different hash string of character and affect all the involved blocks.

**Data Auditing**

TPA benefits the user. It delivers the verification findings to the user and the cloud server as soon as feasible and detects data corruption. TPA's communication with other entities is authenticated. The user determines the verification timeframe. TPA first extracts the hash values of ‘successive blocks that are the latest ones confirmed on the blockchain, where ‘denotes the number of blocks deep used to confirm a transaction and sends the challenging message to the cloud server at a point in time when the data integrity should be verified. When the cloud server receives the challenging message, it computes the relevant proof. TPA validates the proof to ensure the data's integrity. If the checking fails, TPA informs the user that the data may be corrupted.

**Data Sharing**

The storage server is the most significant module in the data sharing concept. A massive amount of data is stored in storage data. This information is safely stored on a storage server. It also stores encrypted data as well as the key used for data encryption. When a user needs data, he sends a request to the storage server. There are two keys that are utilized for encryption and decryption. It is possible to share data in a secure manner.

**Blockchain Technology:**

Transaction initiation: A transaction gets started by a user who wishes to send digital assets to another user on the blockchain. The transaction comprises details such as the value of assets being transferred and the address of the receiver.

Verification: The transaction is broadcast to all nodes (or validators) on the blockchain network, who validate it. This includes verifying the user's digital signature to validate their identity as well as confirming the user has enough assets to finish the transaction.

Block creation: A new block is formed when a particular amount of validated transactions have been collected (commonly referred to as a block). This block contains a hash of the preceding block, resulting in the formation of a chain of blocks (thus the word blockchain).

Consensus: The network nodes collaborate to establish agreement on the validity of the new block. This frequently entails determining which nodes can add the new block to the chain using a consensus mechanism, such as Proof of Work (PoW) or Proof of Stake (PoS).

Block addition: After reaching consensus, the new block is added to the blockchain, and the transactions within it are regarded confirmed and final.

Mining (optional): In PoW-based blockchains, mining is a method in which nodes compete to solve a challenging mathematical issue in order to earn a reward for including a new block to the chain, incentivizing nodes to participate in the network and protect the blockchain.

Node synchronization: Each node on the network updates its copy of the blockchain to include the new block, guaranteeing that all nodes have an accurate and up-to-date copy of the blockchain.

## 6. CONCLUSIONS

A certificateless public verification technique, named CPVPA, is offered here against the procrastinating auditor. This proposed solution makes use of on-chain currencies, with auditor verification integrated into an on-chain currency transaction on the blockchain. Furthermore, the suggested system does not have a certificate management issue. When compared to existing schemes, the security study shows that CPVPA gives the strongest security guarantee.

## REFERENCES

[1] J. Yu, K. Wang, D. Zeng, C. Zhu, and S. Guo, "Privacy-preserving data aggregation computing in cyber-physical social systems," ACM Transactions on Cyber-Physical Systems, vol. 3, no. 1, p. 8, 2018.

[2] H. Ren, H. Li, Y. Dai, K. Yang, and X. Lin, "Querying in internet of things with privacy preserving: Challenges, solutions and opportunities," IEEE Network, vol. 32, no. 6, pp. 144–151, 2018.

[3] J. Li, H. Ye,W.Wang,W. Lou, Y. T. Hou, J. Liu, and R. Lu, "Efficient and secure outsourcing of differentially private data publication," in Proc. ESORICS, 2018, pp. 187–206.

[4] L. Zhong, Q. Wu, J. Xie, J. Li, and B. Qin, "A secure versatile light payment system based on blockchain," Future Generation Computer Systems, vol. 93, pp. 327–337, 2019.

[5] G. Xu, H. Li, Y. Dai, K. Yang, and X. Lin, "Enabling efficient and geometric range query with access control over encrypted spatial data," IEEE Trans. Information Forensics and Security, vol. 14, no. 4, pp. 870–885, 2019.

[6] K. Yang, K. Zhang, X. Jia, M. A. Hasan, and X. Shen, "Privacy preserving attribute-keyword based data publish-subscribe service on cloud platforms," Information Sciences, vol. 387, pp. 116–131, 2017.

[7] W. Shen, B. Yin, X. Cao, Y. Cheng, and X. Shen, "A distributed secure outsourcing scheme for solving linear algebraic equations in ad hoc clouds," IEEE Trans. Cloud Computing, to appear, doi:10.1109/TCC.2016.2647718.

[8] H. Yang, X. Wang, C. Yang, X. Cong, and Y. Zhang, "Securing content-centric networks with content-based encryption," Journal of Network and Computer Applications, vol. 128, pp. 21–32, 2019.

[9] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enabling public verifiability and data dynamics for storage security in cloud computing," in Proc. of ESORICS, 2009, pp. 355–370.

[10] X. Zhang, H. Wang, and C. Xu, "Identity-based key-exposure resilient cloud storage public auditing scheme from lattices," Information Sciences, vol. 472, pp. 223–234, 2018.

[11] K.Wang, J. Yu, X. Liu, and S. Guo, "A pre-authentication approach to proxy re-encryption in big data context," IEEE Transactions on Big Data, 2017, to appear, doi. 10.1109/TBDATA.2017.2702176.

[12] J. Ni, K. Zhang, Y. Yu, X. Lin, and X. Shen, "Providing task allocation and secure deduplication for mobile crowd sensing via fog computing," IEEE Transactions on Dependable and Secure Computing, to appear, doi. 10.1109/TDSC.2018.2791432.4. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. 3rd edn. Springer-Verlag, Berlin Heidelberg New York (1996)