DYNAMIC AND EFFICIENT BLOCKCHAIN BASED CLOUD STORAGE SYSTEM

#1**ERUGURALLA .SATISH BABU,** Assistant Professor,
Dept of CSE,
VIVEKANANDA INSTITUTE OF TECHNOLOGY &SCIENCE, KARIMANGAR

ABSTRACT: Academic and applied research in several fields is improving due to Blockchain's promise and capacity for progress. Blockchain technology is still in its infancy, but it is already being touted as a game-changing answer to many of today's information system problems, including centralization, identification, trust, character, data ownership, and data-driven decision making. The diversity and quantity of digital information created by robots and humans has increased worldwide. The search for the best cloud data storage and processing solution has been aided by blockchain technology. Blockchain technology could protect cloud storage, according to this article.

Key words- Blockchain, cloud computing, data provenance, blockchain-as-a-service, and blockchain service model

1. INTRODUCTION

Due to its practicality and accessibility, cloud computing has becoming increasingly popular. As data grows exponentially, traditional data management systems are struggling to keep up. A network (internet or intranet), a front-end platform (mobile device or client), and a back-end platform make up the classic cloud storage concept (storage or a server). Researchers are improving cloud storage and application due to data growth [1].

Cloud computing is used by corporations and the military to store data. Incompatibility and security risks increase when hardware and software are distributed across many cloud computing platforms. The safety of data stored, transmitted, and sent between clouds is a serious concern [2]. Cloud computing provides instantaneous access to shared computing resources such data storage, server processing power, and software and service delivery through pay-as-you-go pricing [3, 4]. Management or the service provider can easily remove these assets.

Outsourcing computation is one of cloud computing's main benefits. Cloud storage can benefit smaller devices [5, 6] due to pay-as-you-go pricing. Users can rent and pay for cloud services like storage or utility calculations. Cloud computing is more flexible than conventional data storage [7]. Due to limited storage space, information is saved remotely. Due to privacy rules, you cannot trust the service provider and must keep the classifier and data confidential [8, 9]. Academics are enthusiastic about cloud data processing and storage. Heterogeneity is also researchers' biggest data storage challenge. These data storage possibilities are called "big data" or "large-scale data." The cloud infrastructure and blockchain system were adjusted accordingly. Developers are integrating these two technologies [11] to improve program performance. Therefore, blockchain is a secure network protocol for data distribution and storage across numerous computers. This technology helps us move and store massive amounts of data. It saves money and improves efficiency and precision [12].

2. BACKGROUND WORK

BLOCKCHAIN TECHNOLOGY FOR CLOUD STORAGE

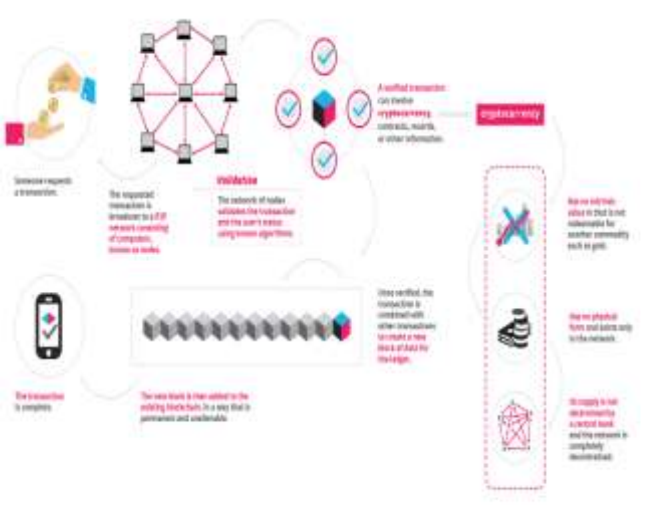
Blockchain technology will consolidate client-server computing. Data processing may be distributed, but it's handled centrally. A hybrid cloud architecture stores data in both public and private clouds. The blockchain creates a distributed storage market. Cloud storage's blockchain data structure is complicated. Block Geeks' info graphic provides additional information.

Multiple new businesses could be founded on a "blockchain storage marketplace" where "hosts sell their additional storage capacity and renters buy this extra capacity and upload files." After data is encrypted and split into "shards," the blockchain distributes them "smartly among hundreds of nodes in dozens of sites."

Blockchain technology may benefit cloud data storage. No one is in charge.

By distributing data across numerous datacenters in different areas, Amazon S3 offers redundancy, yet each datacenter is a possible single point of failure. Blockchain is decentralized because your data is spread among decentralized nodes worldwide. Data is safer during transit and storage. Blockchain could improve users' anonymity because of this. Users' private keys distribute encrypted data across a network of nodes.

Only you (or the data owner) can decrypt the nodes' data. Security is not a concern because the host node can only see a piece of your file. Cost reduction is the last and most important benefit of this adjustment. Amazon S3 is much more expensive than blockchain storage, which costs $2 per terabyte each month.


Someone requests a transaction.
The requested transaction is broadcast to a P2P network consisting of computers, known as nodes.
Validation
The network of nodes validates the transaction and the user's status using known algorithms.
A verified transaction can involve cryptocurrency, contracts, records, or other information.
cryptocurrency
Has no intrinsic value in that it is not redeemable for another commodity such as gold.
Has no physical form and exists only in the network.
Its supply is not determined by a central bank and the network is completely decentralized.
Once verified, the transaction is combined with other transactions to create a new block of data for the ledger.
The new block is then added to the existing blockchain, in a way that is permanent and unalterable.
The transaction is complete.

Blockgeeks

Decentralized cloud computing solutions include Filecoin, MaidSafe, Siacoin, and Storj. Their marketing brochures promise to be better than centralized cloud storage in reliability, security, and cost. Tutorial videos are included in most of them.

All of these sites work because they rent all worldwide hard drive space. Renting out a spare bedroom can actually make you money. Like Uber manages its fleet, these solutions enhance hard disk storage capacity (i.e., more frequently being driven by those who require a car as opposed to the car owners).

These decentralized cloud storage providers' prices fluctuate based on supply and demand. No matter how much storage capacity is given, large companies charge a flat rate.

The big suppliers had to pay their workforce and build enormous facilities around the world, hoping to make a profit. Blockchain technology may use current server infrastructure, therefore distributed solutions don't require a large upfront investment. These cost cuts benefit customers and businesses.

3. EXISTING TECHNIQUES OF BLOCKCHAIN FOR CLOUD STORAGE

Here, learn about this SLR's blockchain technology. Based on a few research, blockchain technology used in cloud storage has been categorized as follows:

- Cryptography Technique (CT)
- Data Deduplication Scheme (DDS)
- Data Integrity Checking Technique (DIC)
- Storage Efficiency Technique (SET)
- Bitcoin Technology (BT)
- Blockchain-based Cloud Storage (BCS)

This SLR uses acronyms. Fine-grained access control, also known as data auditing, searchable encryption, verifiable computation, reliable distributed deduplication, safe data deduplication, bitcoin transactions, non-repudiation assurance, and FAC, tracks individual users' actions on a system. Hyperledger Fabric, Hash Technology, Digital Signature, Merkle Tree, and Ethereum's Proof of Work, Payment Method, PW Access Control System, Deletion Scheme, and Encryption Scheme are symbolized as HF, HT, DS, MT, and ET.

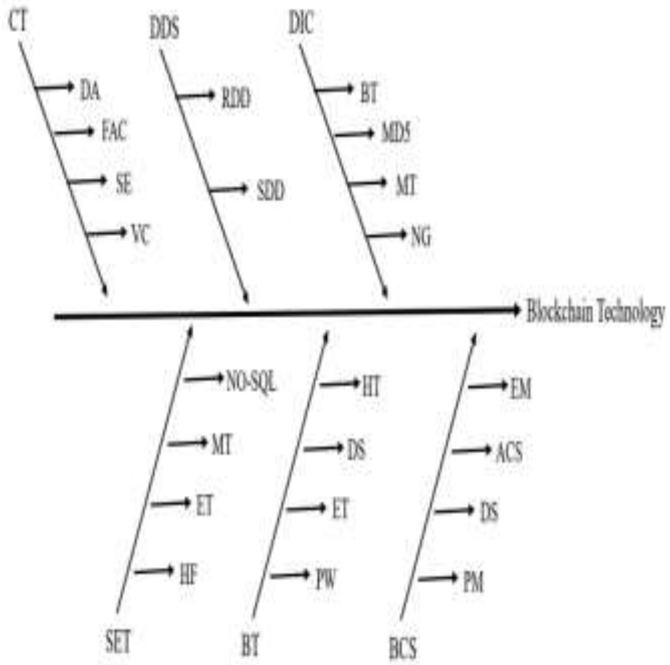


Fig. Blockchain methods selected in this SLR

Blockchain-based Cryptographic Technique for Cloud Storage Service

Users can use the technology to securely transport data to cloud nodes [55]. This diagram shows the encrypted storage structure's three components.

Data Processor (DP): Processing of information before sending it to the cloud.

Data Verifier (DV): Verification of the damaged information stored in the cloud.

Token Generator (TG): For saving the documents of the clients on the cloud, the token generator generates the token for each user.

Figure represents:

(1) Master "y" information processor sets the information before sending it to the cloud;

(2) Master "x" needs the permission of Master "y" for scanning a keyword;

(3) Master "y" generates a token for a keyword and send it back to Master "x";

(4) Master x receives the token and send it to the cloud;

(5) To locate the appropriate encoded documents, the cloud utilizes the token and send the resultant documents back to Master x. (?) At any time, Master y's data verifier can confirm the integrity of the information.

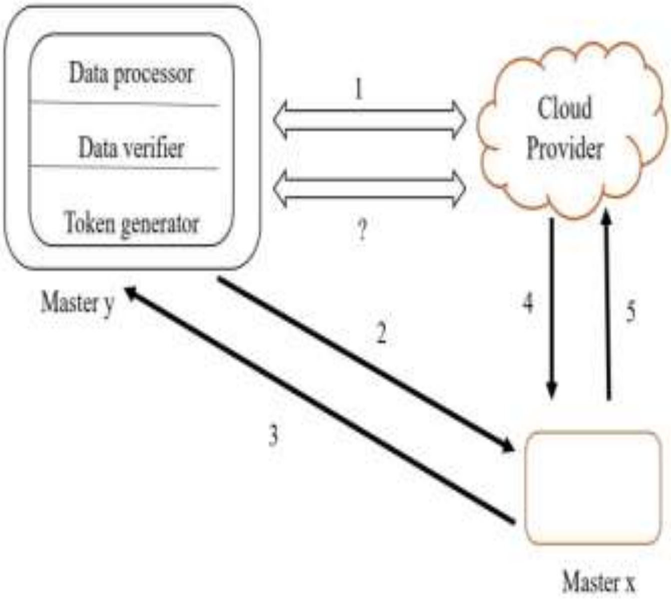


Fig. The architecture of cryptographic cloud storage.

Blockchain-based Data Deduplication Scheme for Cloud Storage Service

Data deduplication frees up space in cloud storage. This method reduces storage and keeps only one copy of data. Thus, it can minimize hardware costs and improve data efficiency [56], but data dependability may worsen. The blockchain tracks each server's location and distributes files across several servers without duplication. Blockchain technology and data deduplication help protect critical system data. For systems that store data across multiple locations, it's useful. To deliver services, connected service providers (CSPs) and data owners must be blockchain nodes. Every replica and transaction must be recorded on a blockchain for data veracity [57]. Based on data size, disk position, and duplication, the data deduplication method is separated into components.

The three types of data deduplication are data unit, location, and disk placement. Below are duplicate clusters. Data unit deduplication can also be divided into file-level and chunk-level. The two files' hash values are compared during deduplication. If the hash values match, both copies are discarded. As part of chunk-level deduplication, the file is split into pieces of various sizes. Next, review duplicated content. Source and target deduplication comprise location-based deduplication. The receiver executes target deduplication and discards duplicates after receiving the files from the client. Deduplication on the storage device does not affect the client. The deduplication method is hidden from users. Before transmission, source data is deduplicated. This deduplication increased network traffic by using client resources. Disk deduplication can be forward- or backward-referencing.

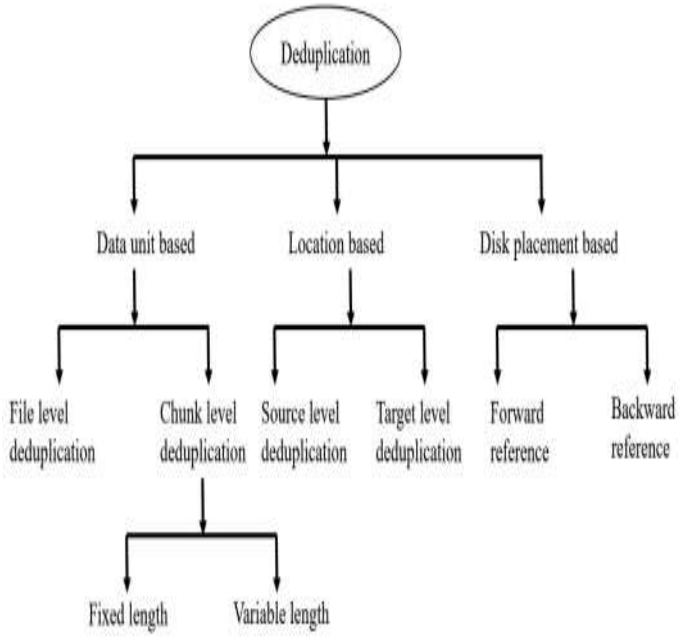


Fig. Classification of deduplication

Blockchain-based Data Integrity Checking Technique for Cloud Storage Service

Decentralized data integrity checks are possible with blockchain-based cloud storage. Three fundamental pillars support integrity checking. Data user, CSS, and blockchain are the three parties. Blockchain technology powers Data Integrity Service (DIS). Before using the DIS, nodes must initialize the blockchain client [59]. Blockchain data verification can employ Merkle trees. Pre-processing and validation comprise this process.

Pre-processing phase: Before building a hash Merkle tree, the data consumer fragments it. CSS and the user agree that a hash Merkle tree is valid, and the user owns its root. Public Merkle trees and user data are stored in CSS. When a user wants access to their CSS data, the service will supply the address.

Verification phase: CSS uses the client's challenge number to prioritize shard verification. Using the challenge number and shard, a hash function calculates the hash digest. CSS forward digests are supporting statistics for the blockchain. A new hash root will be generated by the smart contract and compared. The information is consistent if the hash roots are identical. Data quality has degraded if this is not true. The blockchain will provide the output to the client after verification [60].

Storage Efficiency Technique for Cloud-based Design

NoSQL databases are the best distant data storage option for cloud computing. NoSQL databases are stored and managed on MongoDB, Hadoop, Graph Databases, Column-oriented Databases, Document Databases, and Key-value stores [61]. Cloud storage used plain text, which wasn't good for archiving. Cloud storage and operating system data storage are hampered by this. However, a relational database management system is incompatible with the Map Reduce approach, which manages huge volumes of data [62]. Distributed, scalable, and failsafe, BigchainDB is a cloud database. A NoSQL RethinkDB database underpins BigchainDB, a reliable cloud-based database management system. Like blockchains, it has hashed blocks, transactions, voting, and immutable records [63].

Blockchain-based Bitcoin Technology for Cloud Storage

A decentralized, peer-to-peer payment network supports bitcoin, a virtual currency. Encrypted channels create and send bitcoin [75]. By eliminating banks from financial transactions, Bitcoin revolutionized finance. Bitcoin's publicly acknowledged and shared ledger of transactions is based on the blockchain's cryptographic technologies used by its highly secure and advanced peer-to-peer (P2P) protocols. Bitcoin also safeguards user privacy [76]. Therefore, Blockchain or similar digital money technology eliminates the need for customers or middlemen to trust each other. A decentralized system is reliable by nature.

Blockchain-based Cloud Storage Access Control System

Blockchain technology can track unwanted access to cloud data. In cloud storage, this is the only way to secure critical data. Users can utilize a blockchain-based access control [67] system with a cryptographic approach based on characteristics and dynamic properties to get entry. Key generation, distribution, and revocation are recorded and verified using blockchain, a distributed ledger system.

Permissions are managed using blockchain in Reference [68]. Access control involves identifying, validating, and approving users. It describes how granularly a client's access to a framework activity can be monitored. After verifying a user's identity and cryptographic credentials, the new architecture will allow users to access their EHR over common blockchain data pools. Identity-based validation verifies the client's identity. Businesses can engage without giving customers power over their promises by minimizing job duplication and flexibility. An access control system designed with smart contracts is more stable, adaptable, and useful in securing sensitive data. The blockchain and smart contracts are utilized in a role-based access control (RBAC-SC) system to display the trust relationship crucial to RBAC and handle the challenge-response process that certifies a user's roles [69].

4. BLOCKCHAIN STORAGE PLATFORMS

SWARM

Our blockchain storage method keeps as much of the Ethereum public ledger decentralized and redundant. Its main purpose is to ease blockchain data interchange and storage, as well as decentralized application code and data.

Swarm's distributed file storage offers the following benefits.

- It can withstand attacks that disable all services.
- No rest now.
- Hidden inconsistencies.
- People can barter their possessions for money because the incentives are strong.
- Not muzzled.

IPFS

The Interplanetary Data System is discussed here. IPFS, a file-sharing network, modernizes cross-border data and knowledge exchange.

IPFS benefits:

- It helps organize data.
- It remembers these files' evolution.
- Its self-authenticating file system (SFS). SFS, a decentralized file-transfer system, eliminates the need to re-authenticate before transmitting or receiving data.
- IPFS provides safe transaction access and transparent storage.

SIA

Sia's main benefit is widespread access to low-cost, high-efficiency data systems. It does this independently.

Sia's traits:

- It sends files to global hosts in 30 parts. Thus, no host is likely to become a critical vulnerability. Redundancy and availability are improved.
- Smart contracts transfer data. This allows Sia blockchain-tracked cryptographic SLAs.
- Intermediaries are unneeded. Siacoin can pay hosts and guests.

STORJ

Decentralized cloud storage is popular because all data is encrypted from upload to download. It wants to end censorship, data tracking, and service outages.

Storj is amazing because:

- It supports server space sales using a Distributed Hash Table (DHT). A distributed network of nodes transmits offers from both sides.
- A node must utilize "publish-subscribe" to have the system sign and broadcast an unfinished contract. Interested nodes can sign promising contracts.

MAIDSAFE

This decentralized, user-managed network securely stores data for all users (SAFE).

SAFE network components:

- Always available, consistent, and easy to get. Secure addressing and routing enable data retrieval.
- Each file's encryption key is unique. Thus, their vaults are secure and decryption impossible.
- Server administrators own content, not creators.

5. CONCLUSION

This paper examines the technical obstacles of blockchain-based cloud computing. This project involves service, security, and performance technology. This survey covers the BaaS service model, blockchain-enabled cloud access control, data provenance, searchable encryptions, data deduplication, smart contract-based cloud apps, blockchain-powered offloading, blockchain hardware development, and blockchain-relate. The study's primary findings guide future blockchain-powered cloud datacenter reengineering projects.

REFERENCES

- Zhihua Xia, Xinhui Wang, Liangao Zhang, Zhan Qin, Xingming Sun, and Kui Ren. 2016. A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. IEEE Trans. Info. Forensics Secur. 11, 11 (2016), 2594–2608. https://doi.org/10.1109/TIFS.2016.2590944
- Xueping Liang, Sachin Shetty, Deepak Tosh, Charles Kamhoua, Kevin Kwiat, and Laurent Njilla. 2017. Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability. In Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing. 468–477. https://doi.org/10.1109/CCGRID.2017.8
- David S. Linthicum. 2010. Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide. Pearson Education.
- Rajnish Choubey, Rajshree Dubey, and Joy Bhattacharjee. 2011. A survey on cloud computing security, challenges, and threats. Int. J. Comput. Sci. Engineer. 3 (2011), 1227–1231.
- Bin Feng, Xinzhu Ma, Cheng Guo, Hui Shi, Zhangjie Fu, and Tie Qiu. 2016. An efficient protocol with bidirectional verification for storage security in cloud computing. IEEE Access 4 (2016), 7899–7911. https://doi.org/10.1109/ ACCESS.2016.2621005
- Luis M.Vaquero,Luis Rodero-Merino, Juan Caceres, Maik Lindner. A Break in the Clouds: Toward a Cloud Definition. ACM SIGCOMM Computer Communication Review,2009,39(1):50-55.
- Wu Jiyi,Ping Lingdi,Pan Xuezeng.Cloud Computing: Concept and Platform, Telecommunications Science, 12:23-30, 2009.
- Jonathan Strickl and. How Cloud Storage Works[OL], http://communication.howstuffworks.com/cloudstorage.htm, 2010.
- K. Gai, Y. Wu, L. Zhu, M. Qiu, and M. Shen, “Privacy-preserving energy trading using consortium blockchain in smart grid,” IEEE Trans. on Industrial Informatics, vol. 15, no. 6, pp. 3548–3558, 2019.
- Eyal, A. Gencer, E. Sirer, and R. V. Renesse, “Bitcoin-ng: A scalable blockchain protocol,” in 13th {USENIX} Sym. on Networked Systems Design and Implementation, 2016, pp. 45–59.
- L. Zhu, Y. Wu, K. Gai, and K.-K. R. Choo, “Controllable and trustworthy blockchain-based cloud data management,” Future Generation Computer Systems, vol. 91, pp. 527 – 535, 2019.