Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 10 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i10.43305

PAPER

Optimal Method Combining Blockchain and Multi-Agent System to Ensure Data Integrity and Deduplication in the Cloud Environment

Mohamed El Ghazouani[1](✉), Abdelouafi Ikidid[2], Charafeddine Ait Zaouiat[1], Layla Aziz[1], Moulay Youssef Ichahane[3], Latifa Er-Rajy[4]

[1]ESIM, Polydisciplinary Faculty of Sidi Bennour, Chouaîb Doukkali University, El Jadida, Morocco

[2]Laboratory of System Analysis, Information Processing and Industrial Management, EST Salé, Mohammed V University, Salé, Morocco

[3]Information Technology Laboratory, National School of Applied Sciences, Chouaîb Doukkali University, El Jadida, Morocco

[4]Computer Science Department, Laboratory of Information Systems Engineering, Cadi Ayyad University, Marrakesh, Morocco

mohamed.elghazouani63@gmail.com

ABSTRACT

Cloud computing (CC) refers to the transmission, storage, and processing of any type of information at a location that is not owned or controlled by the information owner. This information can be stored and accessed anytime, from anywhere, and using any device. However, several challenges threaten the deployment of cloud computing technology within organizations, including user privacy, data integrity, and managing the retention of the vast amount of data that will be stored and analyzed in this environment. In recent years, various data integrity checking systems and data storage mechanisms have been proposed to address these diverse challenges. Nevertheless, some protocols do not guarantee data confidentiality against auditors during the auditing process, while others do not provide batch auditing. Additionally, the majority of these protocols do not support the deduplication technique. Therefore, in this research work, the security features of blockchain technology and those of multi-agent systems (MAS) have prompted us to propose models that enable both data deduplication, saving storage space, and data integrity auditing, efficiently verifying the accuracy of data outsourced to CC environments. Implementation findings indicate that the suggested system is highly feasible and applicable in real-world scenarios.

KEYWORDS

cloud computing (CC), security, data integrity, deduplication, MHT, blockchain, multi-agent systems (MAS)

1 INTRODUCTION

Many companies store massive amounts of data locally, which leads to several challenges, such as processing, security concerns, and the need to invest in hardware and human resources. The key is to externalize the data to the cloud.

Cloud computing (CC) refers to the use of networks of remotely located servers, commonly accessible via the Internet, to provide data storage, management,

El Ghazouani, M., Ikidid, A., Zaouiat, C.A., Aziz, L., Ichahane, M.Y., Er-Rajy, L. (2024). Optimal Method Combining Blockchain and Multi-Agent System to Ensure Data Integrity and Deduplication in the Cloud Environment. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(10), pp. 90–105. https://doi.org/10.3991/ijim.v18i10.43305

Article submitted 2023-07-20. Revision uploaded 2023-12-29. Final acceptance 2024-02-17.

and processing. For clients, CC provides access to numerous technologies while reducing barriers to entry, such as technical expertise and cost. In general, the cloud services market is categorized into three main service models: infrastructure, platforms, and software. According to business needs and security concerns, cloud clients can choose between private, public, or hybrid cloud deployment models.

In 2021, four billion users were connected to the cloud compared to 2.4 billion in 2013 [1]. Thus, CC generated nearly US $400 billion in revenue in 2021 [2].

Besides, 60% of businesses use cloud infrastructure to store confidential data. While 84% of IT professionals were concerned about the security of the cloud during the widespread telecommuting in 2020 due to the COVID-19 pandemic [2]. In addition, approximately 75% of our digital universe is a replica; in other words, only 25% is unique. This presents an opportunity for cost-saving technologies such as data compression and deduplication.

Once data is stored on cloud storage servers, the owner forfeits control of it. While this technology provides many benefits, it also poses security concerns, particularly those related to data integrity. To safeguard the authenticity of the externalized data, the owner must activate audit processes. Furthermore, due to the evolution of vast amounts of data, much of which is redundant, storage efficiency is another crucial requirement that must be ensured through technologies like deduplication. To this end, we have proposed a model that aims to implement a new blockchain-based architecture ensuring data integrity verification and deduplication in the cloud environment.

Nevertheless, implementing this technique on the cloud side is complex because large amounts of data must be deduplicated quickly before being stored on the cloud servers. To address these concerns, we propose a new system that utilizes a multi-agent system, where multiple intelligent agents collaborate to establish a flexible server-side deduplication system. In this study, we will specifically explore the deduplication technique and data integrity as crucial aspects for cloud adoption. The organization of this paper is as follows: Section II outlines the convergence of cloud computing and blockchain. The various related works are discussed in Section III. Section IV depicts the deduplication technique and the multi-agent system. Section V describes our proposed system. Section VI includes performance evaluation and result analysis. Finally, a conclusion is presented in Section VII.

2 CONVERGENCE OF CLOUD COMPUTING AND BLOCKCHAIN

2.1 Blockchain: Definition and types

Blockchain technology is an innovative digital ledger system that operates in a distributed and decentralized manner. It enables secure, transparent, and immutable recording of transactions without the need for intermediaries. In a blockchain network, transactions are sequentially logged and organized into a chain of blocks. Each block contains a cryptographic hash of the preceding block, creating an unbreakable chain of blocks. This feature makes it difficult for any unauthorized modification of the data stored on the blockchain because any change made to one block would require consensus from the entire network. This creates a high level of trust and transparency, making it an ideal technology for various applications, including supply chain management, voting systems, and financial transactions [3]. Blockchain is classified into three main types:

- **Public blockchain:** Anyone could verify the operation and also contribute to the consensus-building process. Ethereum and Bitcoin are both public blockchains [4].

- **Private blockchain:** Direct access to data is strictly limited and restricted to a specific list of entities. The private blockchain is often utilized by companies or groups of people who prefer not to make their payment information available to everyone [4].
- **Consortium blockchain:** Also known as federated blockchain, this innovative system caters to the requirements of companies that need both public and private blockchain functionality. Hyperledger and R3CEV are both consortium blockchains [5].

2.2 Applications of blockchain

Many fields and industries are already using blockchain technology or plan to do so in the coming years. Among these sectors are insurance, banking, logistics, energy, real estate, aeronautics, health, and education. Table 1 shows a variety of application areas and examples of applications in which blockchain has been used [6].

Table 1. Applications of blockchain

Field of Application	Application Examples
Identity management	– Public key management [7] – Public DNS [8] – Pretty Good Privacy (PGP) [9]
Authorization	– Fair Access [10]
Storage System	– A software solution for decentralized file storage, utilizing blockchain technology "Meta disk" [11]
Financial and Commercial Services	– Fintech [12]
Government services	– Kudos [13]
Other fields	– Energy saving [14] – Software license [15] – Security: an anti-malware [16]

Blockchain technology is utilized in various industries, which is not surprising given its wide array of functions, including transparency, high resistance to failures, tamper-proof systems, auditability, and self-regulation. Blockchain is able to effectively guarantee data integrity, authenticity, and auditability, compared to other privacy-enhancing technologies (PET).

2.3 Decentralizing cloud data via blockchain

Data integrity threats are of paramount importance because as data tampering can maliciously impact critical business decisions. This problem is particularly common in CC environments, where data owners cannot control basic features such as physical data storage and access control. Blockchain has recently emerged as an intriguing technology that, among other things, offers compelling data integrity properties. Using blockchain to address data integrity threats seems to be an effective choice. The authors in [17] highlight the need to transition to a decentralized architecture to ensure the sustainability of storage, retrieval, and file sharing on a decentralized network. In addition, the authors in [18] demonstrate the effectiveness of using a blockchain-based database to maintain data integrity in CC environments.

In this study, we propose an elegant model that will maintain data integrity and provide data deduplication through the use of blockchain databases in a transparent and secure manner. Figure 1 illustrates the benefits that arise from the adoption of blockchain in cloud computing [19].

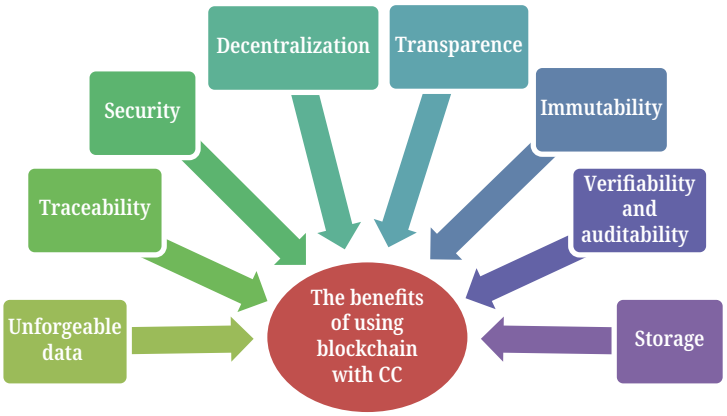

Fig. 1. The benefits of using blockchain with CC

In the previous paragraphs, we have outlined the threats and obligations associated with security measures for CC systems, as well as the requirements for storage efficiency to ensure a reliable and secure system. In this research work, we do not address all the defined challenges. Instead, we focus on two specific requirements that are crucial in the context of our research: data integrity and deduplication techniques. However, it should be noted that our model could be extended to support any proposed improvement. We support this statement by the fact that our proposed blockchain-based solution enhances data storage while preserving the integrity of the stored data.

3 RELATED WORKS

Recently, data integrity and deduplication have garnered significant attention, leading to the proposal of various approaches and protocols. Tian et al. [20] utilized a dynamic hash table (DHT) protocol that enables both public and dynamic auditing functionalities. Lee KM et al. [21] introduced a novel system for verifying the integrity of remotely stored data. M. Shaik Saleem and M. Murali [22] proposed a method for conducting public audits of data integrity in the cloud while preserving privacy. Y. Yannan Li Y., Yong Yu et al. [23] proposed a data integrity auditing method based on fuzzy identities aimed at enhancing reliability in cloud storage systems. Zhu H. et al. [24] proposed an integrity verification scheme that is secure and efficient for cloud-IoT environments, utilizing short signature techniques. Zhao M. et al. [25] proposed a scheme that ensures privacy while auditing remote data integrity in cloud environments, with the assistance of third-party auditors (TPAs). Yu et al. [26] introduced a novel scheme for verifying the integrity of cloud data in secure outsourced storage through attribute-based mechanisms. Tong W. et al. [27] suggested a new method for ensuring privacy while verifying data integrity in secure mobile edge storage environments. Ji Y. et al. [28] proposed a privacy-preserving approach for flexible remote data integrity checking in cloud storage using identity-based mechanisms. Yang C., Song B. et al. [29] proposed a streamlined approach to data integrity auditing that allows for verifiable data updates in secure cloud storage systems. Wang et al. [30] have proposed a scheme aimed at ensuring both security and efficiency when auditing the integrity of data stored in cloud storage. These works explore various techniques, protocols, and algorithms to ensure the integrity of data externalized to CC environments and provide auditing mechanisms to detect any tampering or unauthorized modifications. The major disadvantage of the listed protocols mentioned above is that they do not support data deduplication.

Li et al. [31] suggested a secure storage scheme with deduplication that can support keyword search by using the converged encryption technique to encrypt

data before outsourcing. Miao et al. [32] proposed a secure deduplication protocol assisted by multiple servers in the CC. Zhang et al. [33] introduced a scheme for content-defined asymmetric extremum segmentation in backup storage systems to facilitate data deduplication. Fan Y., Lin X., et al. [34] proposed a scheme that ensures secure privacy while enabling deduplication in a cloud environment.

The strength of this work lies in its thorough exploration of various techniques and protocols to ensure data integrity in CC environments. The authors have compiled a diverse set of approaches, ranging from dynamic hash table protocols to privacy-preserving methods, to address the critical issue of data integrity. The inclusion of these works highlights a broad spectrum of research efforts, providing readers with a comprehensive understanding of the state of the art.

However, the absence of support for data deduplication in the listed protocols is a significant weakness. While the papers discuss various methods to ensure data integrity, they do not address the increasingly important aspect of data deduplication in cloud storage systems. This limitation could influence the overall efficiency and optimization of storage resources in cloud environments. The latter part of the literature introduces works specifically focused on data deduplication, providing a more balanced perspective. Nevertheless, the absence of an integrated approach that seamlessly combines data integrity and deduplication remains a gap. This research paper could strive to bridge this divide and offer a comprehensive solution that addresses both aspects cohesively, ensuring a more robust and holistic approach to data management in cloud computing.

4 DEDUPLICATION TECHNIQUE AND MULTI-AGENT SYSTEM

4.1 Deduplication process

Deduplication is a great solution that ensures data storage efficiency. This implies that there are no multiple copies of the same data being stored. This technique can be applied at the file or file block level (a chunk of a file).

Deduplication allows for:

- Minimize storage space by storing only unique data.
- Avoid the need to invest in specialized storage equipment (reducing infrastructure costs).
- Minimizing network loading by transferring less data can free up more bandwidth for other tasks.

4.2 Chunking algorithms

Chunking is a challenging technique and a key aspect of the deduplication procedure. It can be achieved using various algorithms:

- File-level chunking (FLC): The process involves dividing the file into chunks and calculating the hash value of each chunk. When a duplicate is identified, only one version of the file is retained; otherwise, the file is treated as new data [35].
- Fixed-size chunking (FSC): To ensure consistent handling of data, it is necessary to divide all files into fixed-sized blocks, such as 8 KB, using a hash function to compute a unique hash value for each block. When comparing hashes, if an identical hash is discovered, the file block is identified as redundant and treated accordingly. Conversely, if no matching hash is found, the chunk along with its corresponding hash value will be stored.

- Variable-size chunking (VSC): The data can be fragmented into multiple chunks of different sizes, which are determined based on the content of the files. This approach solves the problems associated with fixed block sizes.
- Content-aware chunking (CAC): If the chunking approach includes the file data stream (file format), the deduplication method in this case saves more space. This algorithm defines boundaries more naturally than other algorithms, leading to increased efficiency.

The VSC algorithm and the CAC algorithm essentially consume more CPU resources [36]. In our proposed model, we utilize the FSC algorithm, where the file is divided into blocks of equal size. Therefore, the FSC algorithm outperforms other algorithms in terms of the speed of executing the deduplication process, making it extensively utilized in various storage systems.

4.3 Multi-agent system

A multi-agent system (MAS) is a collection of intelligent agents that utilize resources and knowledge to interact with each other or their environment in order to solve a problem or achieve a goal.

The agent enjoys the following features [37]:

- **Autonomy:** The agent is able to act independently without any human or other intervention; it controls the possible actions to be taken;
- **Reactivity:** The agent perceives its environment and reacts rapidly to changes;
- **Proactivity:** The agent is able to demonstrate proactive behavior by taking initiatives.
- **Social:** The agent has the capability to engage in communication with other agents [38].

Recently, many authors have suggested the utilization of MAS in CC platforms, incorporating knowledge repositories to effectively handle the data storage of cloud clients. The main challenges that introduce complexity into CC include managing cloud resources, communication, and accounting for resource and/or service usage by each user [39].

5 DESCRIPTION OF THE PROPOSED SYSTEM

Our proposed model leverages blockchain technology to ensure data integrity and deduplication. With reference to the requirements described in Section 2, this section presents how our proposal successfully addresses them. We call our proposed system B-DID for blockchain-based data integrity and deduplication. Therefore, we aim to achieve the following objectives in our proposal:

- Server-side deduplication eliminates duplicate data on the cloud service provider (CSP) side, thereby reducing the amount of stored data.
- Confidentiality ensures the privacy of the data maintained by the auditor during the auditing process.
- Public auditing allows an external auditor to verify the integrity of data.
- Batch auditing ensures that the auditor can run multiple auditing tasks simultaneously, submitted by different users with unique profiles.
- Data integrity ensures that the CSP cannot manipulate or bypass the audit process while maintaining the data unchanged.

5.1 Storage phase (B-DID)

Whenever a client attempts to upload a file, the cloud provider verifies whether the complete file or specific chunks of it exist on the cloud storage server. The goal is to minimize the amount of stored data through deduplication techniques. Each block in our method's blockchain structure contains metadata about a user's file. The block comprises the following information: user ID (U_{id}), file ID (F_{id}), version number (v), number of file blocks (N), timestamp, Merkle root, and the hash of the preceding block. Figure 2 illustrates the structure of our proposed system.

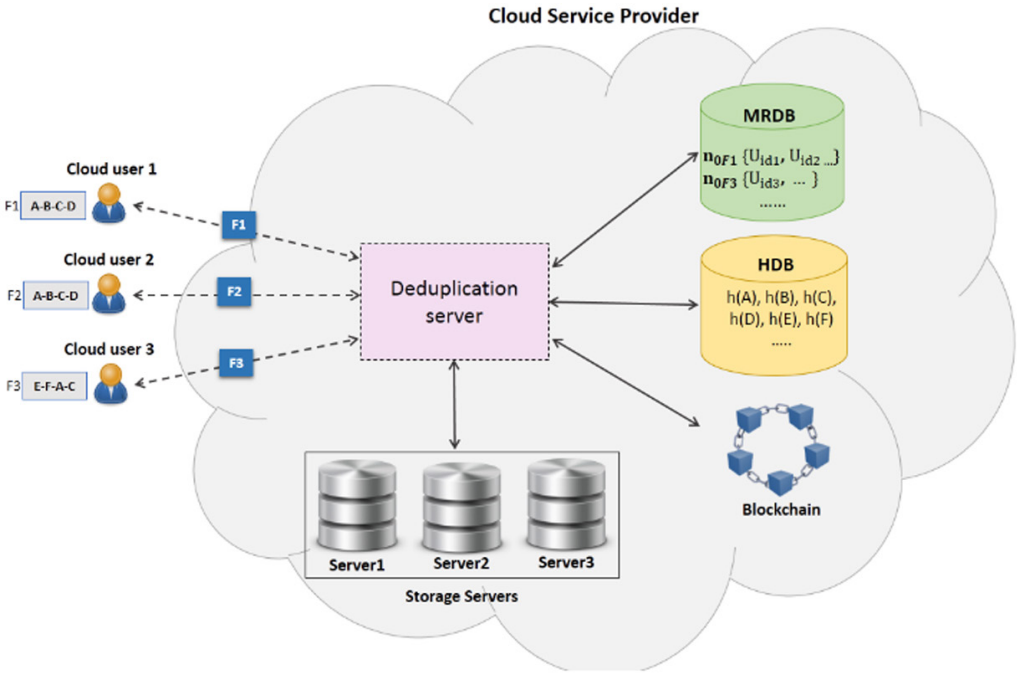

Fig. 2. Architecture of the proposed system

The fixed-size chunking (FSC) module and the MD5 hash algorithm are used to facilitate the deduplication process. The hashes of chunks and the Merkle root are calculated and then compared with those stored in a local Merkle root database (MRDB) and a hash database (HDB), which were generated in the preceding operations, to identify duplicate files or chunks. The deduplication process consists of two steps:

Case 1. In the event that the generated Merkle root, denoted as n_0, does not match any existing root stored in the MRDB, the new root is stored locally in the MRDB along with the user's UID for use in subsequent storage operations. Afterward, the generated hashes are compared with those stored in the HDB. This comparison of hashes will require less time than comparing file blocks. Two possibilities are available: If there are no matching chunk hashes, the calculated hashes of all chunks will be stored in the HDB. Subsequently, all chunks will be externalized to the cloud storage servers. Otherwise, if there are identical hashes, they will be ignored, and only the unique chunk hashes will be stored in the HDB. Only the unique chunks will be externalized to the CSS, reducing storage space usage. Subsequently, a new block is created in the blockchain, and the respective user retains a reference to the block linked to their file.

Case 2. If the generated Merkle root n_0 is identical to a root that was previously stored in the MRDB and the actual U_{id} is listed in the list of user IDs linked to that n_0,

the data owner receives a notification indicating that they have previously stored an identical file in a preceding storage transaction. Otherwise, if the current UID is not found in the list of user IDs associated with this root, the new UID will be added to this list, and a new block will be generated in the blockchain. In the event that there is no need to upload the data to the CSP for storage, it saves bandwidth and storage space.

The root is crucial because it depends on the hashes of all the constituent chunks of files as its foundation. It thus guarantees optimal and secure data content checking.

5.2 Using MAS to ensure the deduplication process (BMAS-DID)

The deduplication process involves a set of steps that can be performed in parallel, especially in cases where thousands of users are simultaneously accessing massive data storage. Given this observation and through this subsection, we propose a reliable and secure system. We introduce the use of MAS to handle server-side deduplication, which will reduce the volume of data stored on CSPs' servers. This new system is driven by managing the deduplication process using a set of intelligent agents. It operates with a distribution, collaboration, and communication mode known as BMAS-DID which stands for blockchain, and MAS is applied for data integrity and deduplication. In addition, this new method utilizes the blockchain data structure, similar to the previous one, as a database to store the metadata of cloud users' files.

Several intelligent agents have been developed to manage deduplication on the CSP side (see Figure 3).

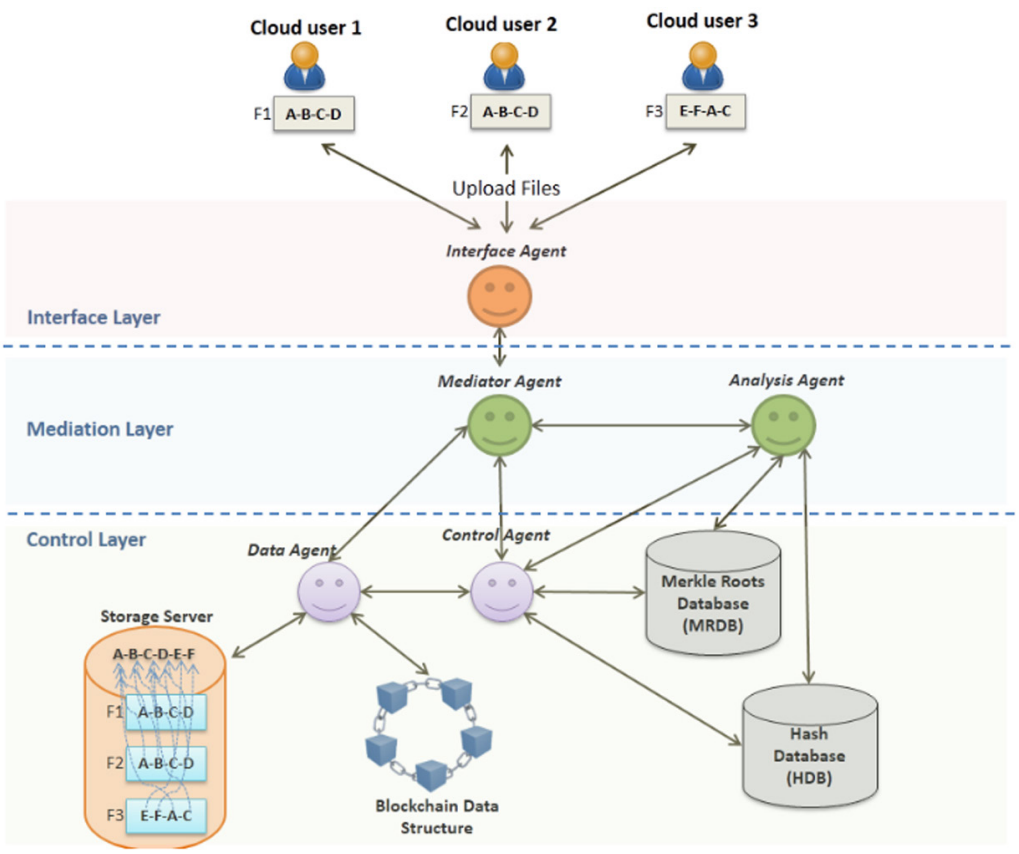

Fig. 3. Architecture of the approach based on multi-agent systems

- **Interface agent:** Its role is to receive files and then transmit them to the MA so that other agents can use them.
- **Mediator agent:** Manages the communication between the different agents.
- **Analysis agent:** Checks whether both the Merkle root and the client ID are present in the MRDB to identify duplicate files. It also performs block-level deduplication by comparing the computed hashes with others stored in the HDB to identify duplicate file blocks.
- **Control agent:** Calculates the Merkle root and transmits it to the Analysis Agent for comparison. Additionally, it preserves the chunk hashes within the HDB. The Merkle root and the client ID are them stored within the MRDB for use in subsequent storage operations.
- **Data agent:** Constructs an additional block within the blockchain, containing the corresponding metadata of the file. The system them stores the chunks that correspond to a file in the CSP storage servers.

Multi-agent systems are characterized by adaptation, cooperation, and distribution. These systems excel at handling various tasks by distributing them among a group of intelligent agents, thereby reduce the execution time of the deduplication process.

5.3 Auditing phase

A TPA is an externally competent entity entrusted by the client to verify the integrity of data held by the CSP. However, the TPA can also pose a risk to the data owner. Thus, preventing data leaks is one of the key challenges in the data auditing process and in preserving data confidentiality. Figure 4 illustrates the public auditing system, depicting the interaction among the three entities (client, CSP, and TPA).

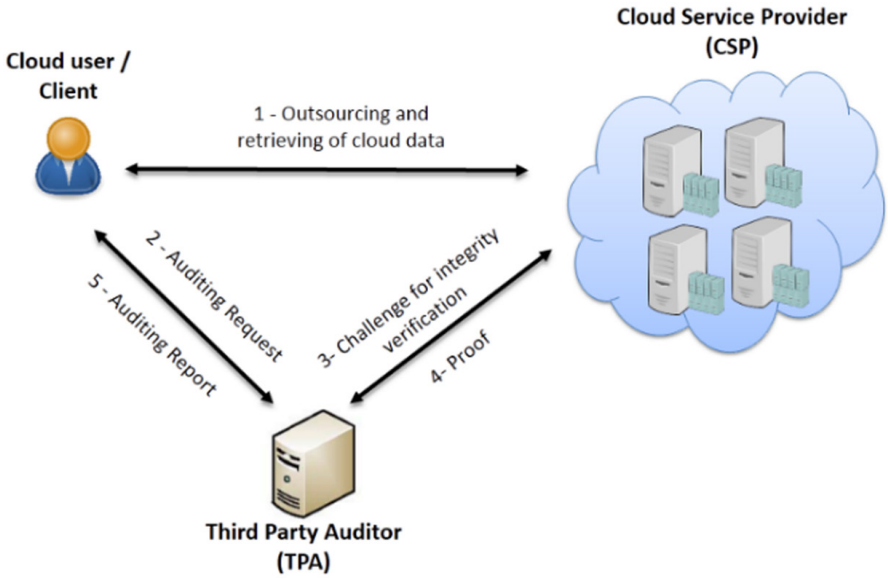

Fig. 4. Relationship between the participating entities in a public auditing model

Our system utilizes blockchain to conduct the verification activities transparently and efficiently. The primary purpose of using blockchain is to allow the TPA to verify data integrity without requiring access to any user data. Therefore, the utilization of

blockchain presents itself as a novel trust model. The algorithm for the audit phase is outlined below:

Algorithm: Auditing Phase

```
1: Begin
2:      OWN sends leaves and Merkle Root to TPA
3:      TPA computes the seed  r = h^P(n_0)
4:      TPA derives leaf numbers in each P chunk as :
5:              for j in 0..P − 1: l_j = G(r,j)
6:      TPA sends leaf numbers {lj} to CSP
7:      CSP provides the appropriate sibling information to TPA
8:      TPA computes the new Merkle root n'_0
9:              checks if n_0 = n'_0
10:     if they match then the file is verified
11:     TPA sends the auditing result to OWN
12: End
```

Due to the algorithm employed, the TPA remains unaware of the client's data, ensuring data confidentiality for auditors. The TPA can run several auditing tasks simultaneously, as received from various users. When the TPA receives multiple audit inquiries for the same document from various clients, it can be more efficient to consolidate them into a single audit task to verify data integrity rather than treating them as separate and independent tasks.

6 PERFORMANCES EVALUATION AND RESULT ANALYSIS

6.1 Implementation and simulation platforms (OpenStack and JADE)

For the implementation of the proposed systems, we utilized the following techniques: the PHP and Java languages, the JADE development platform, and the OpenStack cloud platform. Several implementation and simulation tools are developed in cloud environments. These tools allow us to evaluate and validate hypotheses in a controlled environment where results can be reproduced. In addition, they allow testing services in a repetitive and controllable context before the deployment phase on a real cloud.

All of our experiments were conducted on the OpenStack virtual cloud platform. This technology was chosen because it is built on a modular architecture composed of several interconnected projects that enable control over various resources of the virtual machines, such as computing power, storage, and network. Multi-agent development environments are essential to improving the success of multi-agent technology. MAS platforms enable developers to design and implement their own applications. These simulation platforms are equipped with basic functions for specifying agents and the environment.

6.2 Implementation and evaluation of the B-DID model

The application is built on the FSC algorithm. It accepts a file and a fixed file block size as input. Then, it allows the calculation of fixed-size file blocks from the

beginning of the file and stores them in a temporary folder. We used the MD5 hash algorithm, which generates a unique 32-character string.

The CSP generates file blocks and the corresponding Merkle root of the received files to store them on the storage servers and create a new block in the blockchain. Simply put, the main objective is to decide whether to store the entire file, specific parts of the file, or nothing at all. After the Merkle Root is generated, we compare the Merkle Root n_0 with the Merkle Roots stored in the MRDB from previous storage operations to identify duplicate files. If there is no match between these Merkle Roots, another comparison operation is performed between the generated file block hashes and those stored in the HDB to detect duplicate file blocks. To illustrate the efficiency of our proposal, we conducted several experiments where the file size was increased by 200 MB. We used large files to showcase the advantages of the deduplication technique. Figure 5 illustrates a graph showing the computation time in seconds versus the size of the input files in MB.

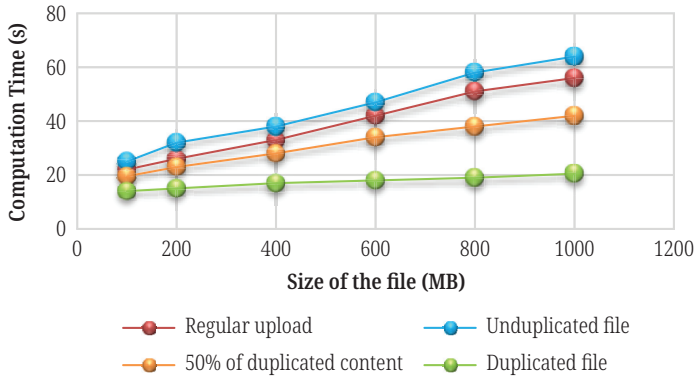

Fig. 5. Evaluation of the time taken in storage tasks for files of varying sizes

The empirical analysis reveals a noteworthy trend where the computational time required for storing a replicated file is consistently lower compared to similar processes in alternative scenarios. The reduction in execution time becomes more pronounced with each additional replica of the file, indicating a tangible correlation between replication and decreased computational overhead. This trend underscores the effectiveness of deduplication techniques in minimizing storage space on servers. As the number of replicas increases, the associated volume of data blocks diminishes proportionally, leading to more efficient utilization of storage resources. For instance, when 50% replicas of a file are present, the cloud service provider (CSP) is only obligated to store the remaining 50% of the data blocks. This exemplifies the capacity of deduplication to significantly optimize storage efficiency in cloud environments.

6.3 Implementation and evaluation of the BMAS-DID model

The BMAS-DID architecture comprises a group of agents that communicate through messages using the ACL language. Each agent includes a set of predefined rules to be executed when the architecture is started to perform the deduplication process. Figure 6 highlights a plot of computation time versus input file size for the two proposed models (B-DID and BMAS-DID), specifically for duplicate and non-duplicate files.

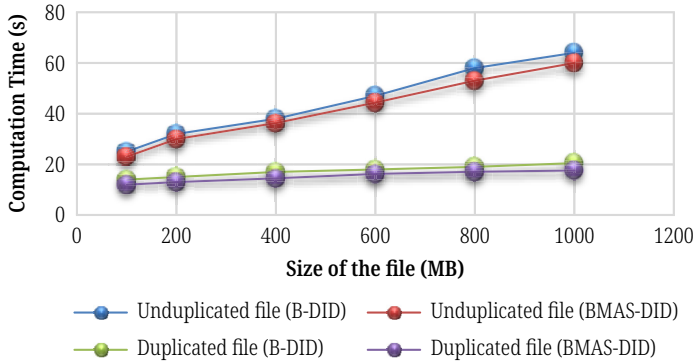

Fig. 6. A plot of computation time, versus input file size for B-DID and BMAS-DID models

The observed data show a significant reduction in processing time for storing duplicated and non-duplicated files when using the BMAS-DID model compared to scenarios using the B-DID model. This improvement is attributed to the deployment of MAS and their ability to distribute tasks among a set of intelligent agents. The efficiency gains are particularly evident in the context of the de-duplication process, where multi-agent systems effectively streamline and optimize task execution. Consequently, the results underline the effectiveness of MAS in managing the deduplication process in cloud environments, highlighting its suitability for improving both overall performance and time efficiency in file storage operations.

6.4 Implementation and evaluation of the audit process

As discussed in subsection 5.3, and considering that the cloud is a multi-client environment, the audit phase is a crucial step to verify the integrity of the data that has been outsourced.

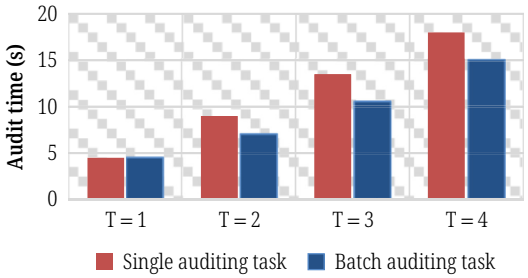

Fig. 7. Comparison of computation time between single and batch audit tasks in the audit phase

The effectiveness of the audit process is graphically represented in Figure 7, where both single and batch auditing methodologies have undergone testing. Notably, the results reveal a discernible advantage in the computation time of batch audit tasks when the parameter T (representing the number of tasks) exceeds 1, compared to the computation time associated with single audit tasks. This compelling observation underscores the efficiency of batch auditing, where the TPA simultaneously performs multiple audit tasks. The simultaneous execution of tasks in batch auditing significantly reduces computational overhead and substantially decreases the overall auditing time. This underscores the practical utility of employing batch auditing strategies to optimize the TPA's resources and enhance the expediency of auditing processes in a concurrent and parallelized manner.

7 CONCLUSION

This paper introduces a system based on blockchain technology that ensures data integrity and deduplication through a model called B-DID. After that, we compared this model with the one integrating multi-agent systems (BMAS-DID), and we emphasized the significance of incorporating MAS in the deduplication process. By indicating that these two models also ensure data integrity auditing.

The implementation and results obtained demonstrate that our models are realistic and valuable for ensuring deduplication and data integrity in a complex and dynamic environment like cloud computing.

The added value of our models over the others is that our proposals allow for data deduplication, ensuring efficient storage, with data integrity auditing, which validates the accuracy of the offshored data. The application of blockchain in our solution is particularly suited to situations where historical data is crucial, such as in the fields of justice, real estate, medical record storage, or tax collection.

8 REFERENCES

[1] "Statista – The Statistics Portal for Market Data, Market Research and Market Studies." https://www.statista.com/topics/1695/cloud-computing/#topicOverview

[2] "101 Shocking Cloud Computing Statistics." https://www.cloudzero.com/blog/cloud-computing-statistics/

[3] D. Tapscott and A. Tapscott, "Blockchain revolution: How the technology behind bitcoin is changing money, business, and the world," *Penguin*, 2016.

[4] L. A. Palmer, R. Fowler, N. Shi, H. Dastani, S. Abouzaid, and E. Kim, "PND61 impact of patient cost-sharing arrangements for disease modifying therapies on treatment compliance among patients with multiple sclerosis in the United States," *Value in Heal*, vol. 14, no. 7, pp. A328–A329, 2011. https://doi.org/10.1016/j.jval.2011.08.527

[5] I. C. Lin and T. C. Liao, "A survey of blockchain security issues and challenges," *Int. J. Netw. Secur.*, vol. 19, no. 5, pp. 653–659, 2018. https://doi.org/10.6633/IJNS.201709.19(5).01

[6] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: A survey," *Int. J. Web Grid Serv.*, vol. 14, no. 4, pp. 352–375, 2017. https://doi.org/10.1504/IJWGS.2018.095647

[7] C. Fromknecht and D. Velicanu, "A decentralized public key infrastructure with identity retention," *Cryptol. ePrint Arch.*, pp. 1–16, 2014.

[8] M. Ali, J. Nelson, R. Shea, and M. J. Freedman, "Block stack: Design and implementation of a global naming system with blockchains," 2016.

[9] D. Wilson and G. Ateniese, "From pretty good to great: Enhancing PGP using bitcoin and the blockchain," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9408, pp. 368–375, 2015. https://doi.org/10.1007/978-3-319-25645-0_25

[10] A. Ouaddah, A. Abou Elkalam, and A. Ait Ouahman, "FairAccess: A new blockchain-based access control framework for the Internet of Things," *Secur. Commun. Networks*, vol. 9, no. 18, pp. 5943–5964, 2016. https://doi.org/10.1002/sec.1748

[11] T. Wilkinson, S. Lowry, and J. Boshevski, "Metadisk: Blockchain-based decentralized file storage application," *Storj Labs Inc., Tech. Report, Hal*, vol. 1, no. 11, 2014.

[12] Y. Guo and C. Liang, "Blockchain application and outlook in the banking industry," *Financ. Innov.*, vol. 2, 2016. https://doi.org/10.1186/s40854-016-0034-9

[13] M. S. and J. Domingue, "The blockchain and Kudos: A distributed system for educational record, reputation and reward," *DM Rev.*, vol. 2, no. September 2016, p. 700, 2002. https://doi.org/10.1007/978-3-319-45153-4

[14] L. P. Johnson, A. Islam, N. Gogerty, and J. Zitoli, "Connecting the blockchain to the sun to save the planet," *SSRN Electron. J.*, pp. 1–16, 2015. https://doi.org/10.2139/ssrn.2702639

[15] J. Herbert and A. Litchfield, "A novel method for decentralised Peer-to-Peer software license validation using cryptocurrency blockchain technology," in *Conf. Res. Pract. Inf. Technol. Ser.*, 2015, vol. 159, pp. 27–35.

[16] C. Noyes, "BitAV: Fast anti-malware by distributed blockchain consensus and feedforward scanning," *arXiv*, no. 1601.01405, pp. 1–10, 2016. https://doi.org/10.48550/arXiv.1601.01405

[17] M. Shah, M. Shaikh, V. Mishra, and G. Tuscano, "Decentralized cloud storage using blockchain," in *2020 4th Int. Conf. Trends Electron. Informatics (ICOEI),* 2020, pp. 384–389. https://doi.org/10.1109/ICOEI48184.2020.9143004

[18] M. El Ghazouani, A. My, E. Kiram, L. Er-rajy, and Y. El Khanboubi, "Efficient method based on blockchain ensuring data integrity auditing with deduplication in Cloud," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 3, pp. 1–7, 2020. https://doi.org/10.9781/ijimai.2020.08.001

[19] K. Gai, J. Guo, L. Zhu, and S. Yu, "Blockchain meets Cloud computing: A survey," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 3, pp. 2009–2030, 2020. https://doi.org/10.1109/COMST.2020.2989392

[20] H. Tian *et al.*, "Dynamic-Hash-Table based public auditing for secure Cloud storage," *IEEE Trans. Serv. Comput.*, vol. 10, no. 5, pp. 701–714, 2017. https://doi.org/10.1109/TSC.2015.2512589

[21] K. M. Lee, K. M. Lee, and S. H. Lee, "Remote data integrity check for remotely acquired and stored stream data," *J. Supercomput.*, vol. 74, pp. 1182–1201, 2018. https://doi.org/10.1007/s11227-017-2117-4

[22] M. Shaik Saleem and M. Murali, "Privacy-preserving public auditing for data integrity in cloud," *J. Phys. Conf. Ser.*, vol. 1000, no. 1, p. 012164, 2018. https://doi.org/10.1088/1742-6596/1000/1/012164

[23] Y. Li, Y. Yu, G. Min, W. Susilo, J. Ni, and K. K. R. Choo, "Fuzzy identity-based data integrity auditing for reliable cloud storage systems," *IEEE Trans. on Dependable and Secur. Comput.*, vol. 16, no. 1, pp. 72–83, 2019. https://doi.org/10.1109/TDSC.2017.2662216

[24] H. Zhu *et al.*, "A secure and efficient data integrity verification scheme for Cloud-IoT based on short signature," *IEEE Access*, vol. 7, pp. 90036–90044, 2019. https://doi.org/10.1109/ACCESS.2019.2924486

[25] M. Zhao, Y. Ding, Y. Wang, H. Wang, B. Wang, and L. Liu, "A privacy-preserving TPA-aided remote data integrity auditing scheme in clouds," *Commun. Comput. Inf. Sci.*, vol. 1058, pp. 334–345, 2019. https://doi.org/10.1007/978-981-15-0118-0_26

[26] Y. Yu, Y. Li, B. Yang, W. Susilo, G. Yang, and J. Bai, "Attribute-based cloud data integrity auditing for secure outsourced storage," *IEEE Trans. on Emerg. Top. in Comput.*, vol. 8, no. 2, pp. 377–390, 2020. https://doi.org/10.1109/TETC.2017.2759329

[27] W. Tong, W. Chen, B. Jiang, F. Xu, Q. Li, and S. Zhong, "Privacy-preserving data integrity verification for secure mobile edge storage," *IEEE Trans. on Mob. Comput.*, vol. 22, no. 9, pp. 5463–5478, 2023. https://doi.org/10.1109/TMC.2022.3174867

[28] Y. Ji, B. Shao, J. Chang, and G. Bian, "Flexible identity-based remote data integrity checking for cloud storage with privacy preserving property," *Cluster Comput.*, vol. 25, pp. 337–349, 2022. https://doi.org/10.1007/s10586-021-03408-y

[29] C. Yang, B. Song, Y. Ding, J. Ou, and C. Fan, "Efficient data integrity auditing supporting provable data update for secure cloud storage," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022. https://doi.org/10.1155/2022/5721917

[30] N. Garg, A. Nehra, M. Baza, and N. Kumar, "Secure and efficient data integrity verification scheme for cloud data storage," in – *IEEE 20th Consum. Commun. Netw. and Conf. (CCNC)*, 2023, vol. 2023, pp. 1–6. https://doi.org/10.1109/CCNC51644.2023.10059690

[31] J. Li, X. Chen, F. Xhafa, and L. Barolli, "Secure deduplication storage systems supporting keyword search," *J. Comput. Syst. and Sci.*, vol. 81, no. 8, pp. 1532–1541, 2015. https://doi.org/10.1016/j.jcss.2014.12.026

[32] M. Miao, J. Wang, H. Li, and X. Chen, "Secure multi-server-aided data deduplication in cloud computing," *Pervasive and Mob. Comput.*, vol. 24, pp. 129–137, 2015. https://doi.org/10.1016/j.pmcj.2015.03.002

[33] Y. Zhang *et al.*, "A fast asymmetric extremum content defined chunking algorithm for data deduplication in backup storage systems," *IEEE Trans. on Comput.*, vol. 66, no. 2, pp. 199–211, 2017. https://doi.org/10.1109/TC.2016.2595565

[34] Y. Fan, X. Lin, W. Liang, G. Tan, and P. Nanda, "A secure privacy preserving deduplication scheme for cloud computing," *Futur. Gener. Comput. Syst.*, vol. 101, pp. 127–135, 2019. https://doi.org/10.1016/j.future.2019.04.046

[35] C. Wang, Z. G. Qin, J. Peng, and J. Wang, "A novel encryption scheme for data deduplication system," in *2010 Int. Conf. Commun. Circuits Syst. (ICCCAS)*, 2010, pp. 265–269. https://doi.org/10.1109/ICCCAS.2010.5581996

[36] Y. Won, R. Kim, J. Ban, J. Hur, S. Oh, and J. Lee, "PRUN: Eliminating information redundancy for large scale data backup system," in *2008 Int. Conf. Comput. Sci. its Appl. (ICCSA),* 2008, pp. 139–144. https://doi.org/10.1109/ICCSA.2008.46

[37] M. Wooldridge and N. R. Jennings, "Intelligent agents: Theory and practice," *The Knowledge Engineering Review,* vol. 10, no. 2, pp. 115–152, 1995. https://doi.org/10.1017/S0269888900008122

[38] G. Tsochev, R. Trifonov, and G. Naydenov, "Agent communication languages comparison: FIPA-ACL and KQML," 2015.

[39] J. Bajo, F. De la Prieta, J. M. Corchado, and S. Rodríguez, "A low-level resource allocation in an agent-based cloud computing platform," *Appl. Soft Comput.*, vol. 48, pp. 716–728, 2016. https://doi.org/10.1016/j.asoc.2016.05.056

9 AUTHORS

Mohamed El Ghazouani is a Professor at the Polydisciplinary Faculty of Sidi Bennour, Chouaib Doukkali University, Morocco. He has completed his doctoral degree in the area of Cloud Computing Security at the Department of Computer Science, Faculty of Science Semlalia, Cadi Ayyad University, Marrakesh, Morocco. He has over 14 years of rich experience in academics and administration. He is the author of numerous publications related to areas of interest, including Cloud Computing Security, Machine Learning, Big Data, and Blockchain. He is a member of reviewing committees for various national and international journals. He can be contacted via email at mohamed.elghazouani63@gmail.com.

Abdelouafi Ikidid is a professor at EST Salé, Mohammed V University in Morocco. He holds a PhD from the Computer Science Department of Cadi Ayyad University in Marrakesh, Morocco. His research interests are in software engineering, focusing on multi-agent systems and artificial intelligence. He is the author of numerous publications related to artificial intelligence, multi-agent systems, and IoT. He can be contacted via email at a.ikidid@gmail.com.

Charafeddine Ait Zaouiat is a Professor at the Polydisciplinary Faculty of Sidi Bennour, Chouaib Doukkali University in Morocco. He is an accomplished professional in the field of networks and telecommunications, specializing in IoT and AI

applied to health. His doctoral research focused on the application of IoT and AI techniques to enhance healthcare systems and improve patient outcomes. He has authored numerous papers and conducted research in collaboration with other experts in the field. He can be contacted via email at charafeddineaitzaouiat@gmail.com.

Layla Aziz is a Professor at the Polydisciplinary Faculty of Sidi Bennour, Chouaib Doukkali University in Morocco. She has completed her doctoral degree in the area of Computer Network and Wireless Sensor Network at the Department of Mathematics and Computer Science, Faculty of Sciences and Techniques, Cadi Ayyad University, Marrakesh, Morocco. She is the author of numerous publications related to areas of interest that include WSNs, Heterogeneous Networks, IoT networks, multi-criteria analysis, and machine learning. She can be contacted via email at laylaa.az@gmail.com.

Moulay Youssef Ichahane is a PhD candidate at the Information Technology Laboratory, National School of Applied Sciences, Chouaîb Doukkali University, Morocco. His research interests include software modeling and design, metamodel design, healthcare systems, and artificial intelligence. He can be contacted via email at y.ichahane@gmail.com.

Latifa Er-Rajy is a Professor at the Computer Science Department at Cadi Ayyad University in Marrakesh, Morocco. Her research interests include Android applications, mobile security and computer networks security. She is the author of several publications related to her research interests. She can be contacted via email at errajy.latifa@gmail.com.