An overview of cloud storage data auditing schemes

Dongchen Cui

University of Electronic Science and Technology of China, Chengdu 610054, China

lucy2000cui@outlook.com

Abstract. With the rapid growth of the internet today, cloud storage services have had an impact on numerous internet users worldwide. Internet users, in order to avoid potential risk of data loss, can outsource their local data to remote cloud servers instead of using local media. However, this kind of cloud storage services is sometimes unreliable enough and the data security of users cannot be guaranteed. In this case, data auditing mechanisms for cloud storage are studied to avoid the destruction of user data. The traditional audit schemes are based on provable data possession mechanisms, while the third-party organization relies on the recent blockchain-based methods, conventional single-cloud storage system and recently emerging multi-cloud storage models. Some audit schemes not only provide auditing capability, but also have some other ancillary functions to ensure applicability for different potential malevolent situations. Gathered correlated studies about data auditing schemes of cloud service are analyzed and classified in this paper, based on research objects and research methods in order to seek probable innovation points. We also analyze experiments in their papers. Toward the finish of this review, we sum up all the examination, list the bearings not investigated, and give a few creative spots.

Keywords: data auditing, blockchain, cloud storage.

1. Introduction

A new technology that has gained prominence during the rapid development of Internet services is cloud computing. Cloud computing has attractive features, such as large-scale and on-demand services. Cloud storage, one of the key technologies in cloud computing. It is possible to store data in the cloud through cloud storage, which allows users to access it from anywhere, leaving them with virtually endless storage space. As a result, more and more individuals and businesses are taking advantage of third-party cloud platforms to store their data. Some well-known cloud storage services, such as Dropbox, Microsoft's OneDrive, Google Drive, Baidu's Baidu Disk and so on, have hundreds of millions of users worldwide. Clearly, cloud storage services have played an important role in people's daily lives. Nevertheless, using network computing services to outsource data, which means transferring the ownership of user data to cloud service providers, can result in security risks. Even though some famous big companies, such as Tencent and Alibaba, have had disputes over data loss from cloud storage. It can be seen that cloud service providers are not always reliable, which leads to a variety of data leakage or data damage caused by human or non-human reasons, including but not limited to hacking servers, internal staff leaking data, etc.

In addressing the security of outsourced data, the academic community has already achieved many methods to audit data in cloud storage. Early remote data auditing models were mostly based on the provable data possession protocol. In recent years, most auditing schemes have relied on a third-party auditor (TPA) such as a bank. In order to avoid the risk associated with the outsourcing of all data to a single cloud storage provider, the concept of multi-cloud storage developed. In addition, blockchain technology has been heralded by researchers recently for taking advantage of its distributed nature and traceability.

This paper classifies the existing research literature on cloud storage data integrity auditing. First, we analyze different research objects and find that topics about multi-cloud and TPA-independent auditing schemes are emerging. Then we analyze research methods which use blockchain technology and find that using blockchain as substitute of TPA and finding new verification parameter may be a chance. Finally, we analyze the measurement methods of various literature, and find that most of them studied the cost of the proposed model, while some of them analyzed the efficiency of the proposed model or algorithm.

Here are the remaining sections of this survey paper. In Section II, we provide a classification of research objects of cloud storage data auditing scheme. A classification of research techniques is given in Section III. We compare the experimental analysis in given papers about cloud storage auditing scheme with relevant measures in section IV. Future research opportunities about cloud storage auditing scheme are discussed in Section V. At last, section VI makes a conclusion of this survey paper.

2. Classification of research objects

Table 1. Different Research Objects about data auditing scheme of cloud storage.

Number of cloud storage services provider	Existence of Third-Party Auditor	
	TPA-independent	TPA-dependent
Multi-cloud	I. [1]	II. [2][3][4][5]
Single-cloud	III. [6][7]	IV. [8][9][10][11][12][13]

2.1. *Criteria*

In recent years, research on cloud storage security has increased. To categorize the research objects regarding data auditing scheme of cloud storage in this section, two mutually exclusive and distinct criteria are employed, which divide research objects in to four different types.

The number of data storage services. There are two kinds of outsourcing here: multi-cloud or single-cloud. Single-cloud means the users of cloud storage services outsource their data to a single cloud storage service provider to store their data. Multi-cloud is a more complex way that users outsource their data to multiple cloud service providers, for possible safety reasons. In practice, users may prefer using multi-cloud outsourcing for security reasons.

Audit scheme. There are two types of audit scheme here: TPA-dependent and TPA-independent. TPA-dependent schemes rely on trusted TPAs. This centralized character is trusted by all entities in the audit scheme. TPA-independent schemes use other methods to replace TPAs. Due to its decentralized nature, blockchain is a powerful and popular substitute for to TPA.

2.2. *The classification*

We provide the categorization findings in Table 1 based on the aforementioned two classification criteria, number of data storage services and audit scheme. The findings of the categorization divide the present papers into four groups.

Type I: This type focuses on a scheme that is not relying on trusted TPAs and was employed in a cloud storage system that let customers to distribute their data among several cloud storage companies.

Type II: This type focuses on a scheme that is relying on trusted TPAs and was employed in a cloud storage system that let customers to distribute their data into one cloud storage companies.

Type III: This type focuses on a scheme that is not relying on trusted TPAs and was employed in a cloud storage system that let customers to distribute their data among several cloud storage companies.

Type IV: This type focuses on a scheme that is relying on trusted TPAs and was employed in a cloud storage system that let customers to distribute their data into one cloud storage companies.

2.3. Explanation of different types

2.3.1. TPA-independent & multi-cloud. References [1] belong to Type I. As far as we are aware, just one study of this kind exists. In this paper, Zhang et al. focus on multiple cloud storage providers (CSPs). They build a public batch data integrity auditing scheme in multi-cloud situations without the requirement for a centralized TPA with blockchain since they think it is difficult to locate a TPA who is trusted by several CSPs. This paper also provides an algorithm to check whether entities, including users and CSPs in its scheme, have malicious behavior.

2.3.2. TPA-dependent & multi-cloud. References [2] [3] [4] [5] belong to Type II. This sort of study supports batch auditing to increase the effectiveness of data audits across several cloud storage services, with reliance to TPA. In reference [3], Multiple CSPs working together to produce integrity evidence for verification is an essential goal of the researchers. Additionally, they provide parameter optimization techniques to reduce the computation costs of storage service providers and clients. In reference [2], He et al. provide a multi-cloud storage protocol that enables TPA to complete batch verification. In this protocol, a TPA can verify multiple auditing requests from various users on different outsourced data on different cloud storage servers. In reference [4], Wang et al. provide an identity-based Provable Data Possesion (PDP) in multi-cloud. In their scheme, they use an entity that they call it combiner to receive storage request and distribute identity message. The combiner also deals with received challenges, distribute them to different CSPs, and finally send their response to a verifier, whose duties are the same as TPA. In reference [5], they at first focus on deduplication techniques in order to reduce burden on the system in the uploading process for CSPs. In their scheme, the data stored in different server, and their integrity can be audited simultaneously by TPA, who has significant computing resources.

2.3.3. TPA-independent & single-cloud. References [6] [7] belong to Type III. This kind of study provides data integrity verification for users without relying on TPA. They are designed for single-cloud scenarios. If this sort of scheme is employed in a multi-cloud setting, each CSP must be examined independently. In reference [6], Liu et al. provide a scheme that is applicable to the internet of things (IoT), in which the data is inherently dynamic. They provide a data integrity solution that is carried out through smart contracts. In reference [7], Xu et al. provide an arbitrable data auditing scheme with commutative hash and a decentralized arbitration process based on the use of smart contracts to settle data ownership issues without the need of TPAs.

2.3.4. TPA-dependent & Single-cloud. References [8] [9] [10] [11] [12] [13] belong to Type IV. Some of these researches use relatively traditional way to solve remote data audit and consider in simpler situations. Others, like reference [8] [9] [12] use blockchain in their design. In reference [8], Xu et al. focus on deduplication technique within their audit scheme. They set their deduplication step before the audit step. In reference [9], Zhang et al. focus on monitoring TPA's efficiency on detecting data corruption. In reference [10], Shen et al. avoid using hardware token to storage user's private key, instead they use biometric data to generate fuzzy signature. TPA's verification is based on the generated fuzzy signature. Reference [11] protects user' privacy from the public verifier during the auditing process based on ring signature technology. Reference [12] provide an identity-based auditing scheme. The scheme constructs random challenge message with the help of blockchain to constrain TPA.

Reference [13] is also identity-based. It also focuses on protecting users' privacy, and reducing computation cost.

3. Classification of research methods

Table 2. Different Research Methods of data auditing schemes of cloud storage.

Verification	Purpose of using blockchain	
	Constraint of TPA	Substitute of TPA
KGC	I. [9][5][12]	II.
Verification metadata	III.	IV. [1][7]
Others	V. [8]	VI.

3.1. Criteria

Recent advancements in blockchain technology have made it feasible to address certain centralized issues with cloud storage data audits. In terms of research methods, we focus on those researches that provide blockchain-based audit schemes. This section will categorize the study techniques of the data auditing scheme of cloud storage into six different categories using two mutually exclusive and distinct criteria.

1) Purpose of using blockchain. Nowadays, objectivity is becoming more important in data integrity auditing, especially when disputes happen. In the past, a trusted TPA was responsible for auditing. However, TPA may also be malicious. Some researches use blockchain as a way to constrain TPA, such as storing TPA's auditing logs in a trusted way. Other studies use smart contracts on the blockchain to replace TPA entirely in order to construct trusted auditing.

2) Verification parameters. There are three kinds of verification, key generation center (KGC), verification metadata and others. KGC is subject to an independent authority. In audition models, it usually generates a partial private key for the user by using the user's identity, and sends it to the user. Usually, it takes part in setup phases. Other researches do not depend on KGC to setup their models. Instead, entities in these schemes jointly generate verification metadata, as a distributed substitute for a centralized KGC.

3.2. The classification

Based on the two classification standards, purpose of using blockchain and verification parameters appealed, we can provide the classification of research methods results about data auditing scheme in cloud storage in Table 2. The classification results of data auditing scheme in cloud storage divide the current papers into six categories.

Type I: This type research proposes to use KGC as part of their scheme, and the purpose of using blockchain is to constrain TPA.

Type II: This type research proposes to use KGC as part of their scheme, and the purpose of using blockchain is to replace TPA.

Type III: This type research proposes to use verification metadata, and the purpose of using blockchain is to constrain TPA.

Type IV: This type research proposes to use verification metadata, and the purpose of using blockchain is to replace TPA.

Type V: This type research proposes to use other methods to generate verification parameter, and the purpose of using blockchain is to constrain TPA.

Type VI: This type research proposes to use other methods to generate verification parameter, and the purpose of using blockchain is to replace TPA.

3.3. Explanation of different types

3.3.1. KGC & Constraint of TPA. References [9] [5] [12] belong to Type I. Many public verification techniques rely on public key infrastructure (PKI), which maintains the user's certificates. PKI often faces problems such as verification, certificate storage and so on, which lead to the increase of cost. In addition, identity-based crypto-system is introduced to solve the problem, and its security depends on the trust of private key generator (PKG). In reference [9], researchers try to design a scheme that can resist procrastinating TPA, and make TPA able to verify data integrity without certificate management. The TPA's verification time is determined by users. Users can audit TPA's verification behavior by obtaining blocks according to verification time. Based on the identity of the user, the KGC generates a portion of the private key for the user. In reference [5], Yang et al. provide a multi-replica public auditing scheme. They introduce blockchain to generate random, unpredictable challenge messages to constrain TPAs. Users' private key is partly generated by KGC and partly by themselves, so that KGC cannot obtain the complete private key of users, so as to avoid the problem of key escrow. In reference [12], the nonce in a blockchain constructs unpredictable challenge messages to TPA, to avoid malicious TPA forgoing their audit results. And TPA's audit results are stored in blockchain, too. A fully trusted authorized agency manages the KGC as PKG. Each user is provided with a full private key.

3.3.2. Verification metadata & Substitute of TPA. References [1] [7] belong to Type IV. Entities in these schemes jointly generate verification metadata, as a distributed substitute for a centralized KGC. In reference [1], TPA is fully replaced by smart contract on blockchain, and the data storage in blocks is used for locating faults while arbitrating the malicious entity in the scheme. The scheme uses RSA-based hash function to construct Homomorphic Verification Tag. In reference [7], Xu et al. design an arbitration node to be the only full node to keep all transaction blocks and others only containing limited information. The system completes auditing by certifying the integrity of a group of blocks at random. Integrity metadata is created by researchers using hash value.

3.3.3. Others & Constraint of TPA. References [8] belong to Type III. In reference [8], users entrust TPA to verify their data on CSPs. A blockchain audit result will be published by TPA after CSP validates the proof generated by CSP. Users will check these results periodically. The scheme uses a system manager to generate the user's secrete key. In fact, the system manager in their scheme has similar duties to those of KGCs.

4. Review of experimental analysis

Table 3. Experiments with Different Metrics and Parameters about data auditing schemes of cloud storage.

Metric	Parameters			
	Each phase	Number of data blocks	Number of events	Consensus mechanism
Time	[1][8][14][5]	[10][4][12][13]	[2][11][6]	[7]
Gas	[1][8][7][14]			
Bloating rate		[7]	[8][9]	
Transaction throughput			[7][14]	[7][14]
Transaction latency				[7]

The system parameters and evaluation metrics from relevant studies will be categorized in this section, as shown in Table 3. The metrics and parameters of experimental analysis are also categorized in Table 3. Table 3 shows that the majority of the references concerning data auditing schemes in cloud storage focus on time and gas measurements.

4.1. Metric evaluation

Most studies choose time as their basic measurement of efficiency of their auditing scheme. Those that use blockchain-based methods additionally use gas as an elementary consideration of their system cost. For those considering space cost, bloating rate is an adjunctive metric they are concerned about. Transaction throughput is a pressure gauge for research that concerns the compressive strength of their system. For research that concerns strictly response time, transaction latency is a crucial metric.

Time: Time cost is a basic metric for measuring system performance.

Gas: Gas is the fee that must be paid in order to successfully complete a contract or conduct a transaction on the Ethereum blockchain platform. This fee reflects the amount of work required to carry out operations on the Ethereum virtual machine (EVM). Researchers usually measure the gas unit amount of gas consumed, and the exact price of the gas is not discussed.

Bloating rate: Blockchain bloat has been identified as a chronic side effect of blockchain ecosystems, and the rate of bloating is inversely related to system scalability and storage cost of network entities. Hence, bloating rate is an important metric to measure whether the system is sustainable in practical scenarios.

$$\text{Bloating\_rate} = \text{storage\_increased} / \text{time\_unit} \tag{1}$$

Transaction throughput: Throughput reflects a system's ability to handle several concurrent operations. Some studies, like reference [7], examine the link between the concurrency value of transactions and the speed of the method for dealing with these transactions to determine transaction throughput.

Transaction latency: Transaction latency is a negative performance measure that represents an auditing system's responsiveness. It is measured by some studies that take latency of system very seriously, like reference [7]. Different from time cost, transaction latency is more real-time and can have a great impact on users’ experience.

4.2. System parameters

Each phase: Audit schemes are divided into different phases, including setup phase, audit phase, verification phase and so on.

Consensus mechanism: Different consensus mechanisms used by blockchain are compared by several references, such as PoW, PoA and others.

4.3. Experimental comparison

In reference [1], the author compares time cost in each phase of the off-chain operation. Furthermore, the author compares the time costs of their scheme to earlier research on blockchain-based multi-cloud storage data auditing schemes with varying accuracy rates. For the on-chain computational overhead, the researcher tests the cost of gas of the blockchain system.

The author of reference [8] evaluates the time delay of operations in the off-chain experiment part. The author evaluates the amount of gas spent by executing various operations in the scheme to determine on-chain computing overhead. In terms of storage costs, the author examines the blockchain bloating rate under various scenarios. The author also examines the system's storage load on various entities.

In reference [7], the author compares the average gas consumption in different phases of the scheme. The author also evaluated the transaction latency of their system by launching transactions every day and recording the latency adopting the PoA and the PoW consensus mechanism respectively. In addition, the author evaluates the bloating rate under diverse conditions.

In reference [6], the author compares batch audit efficiency by measuring time cost under different number of audit tasks (events) with other schemes. The author also compares the time cost under different number of broken blocks.

In reference [14], the author compares the average gas consumption under different operations of the scheme. The author deploys systems under PoW and PoA consensus mechanisms to compare the transaction throughput. For bloating rate, the author tested the blockchain bloating rate for different transaction frequencies.

In reference [5], The author compares the calculation time cost of their technique to previous studies in various auditing phases.

In reference [12], the author compares the computation time cost of their scheme with others under different amount of data blocks.

5. Discussion and suggestions

This survey paper discusses the research methods and research objects of data auditing schemes of cloud storage from various references, and the conclusion can be drawn as follows.

1) In the past, integrity auditing usually relied on TPA. In recent years, the research on TPA-independent auditing scheme has become a trend. In addition, most of the implementation methods are through blockchain.

2) In the process of making audit schemes, whether to use KGC is different in different studies. In the study of replacing TPA with blockchain, KGC has not been used. In addition, non-KGC schemes still have the opportunity to use blockchain as an alternative to TPA.

At last, this study proposes the following directions for cloud data auditing schemes research, which may give some future prospects for cloud data auditing schemes research.

1) Improve the existing audit model to make it more secure and reliable. Some research focuses can be shifted from single-cloud to multi-cloud storage.

2) Classified research methods II, III, VI may be considered as new methods for further research.

3) A new assistant function for audit scheme can be developed. In addition to locating faults, there can also be insurance mechanisms and withdrawal mechanisms. Audit schemes with back-up checking and recovery may also be a direction.

6. Conclusions

Through the previous analysis, we can see that although there are many researches on data integrity audit of cloud storage, studying different objects and using different methods, some classified research methods have not yet been used by any previous research institute. This paper puts forward some suggestions for future research. New research methods and more new functions can be added into the schemes, which requires systematic research.

The limitation of our work is that we haven't deeply studied the specific algorithms used in these papers, so the classification method we provide are the external view of these models.

References

[1] Cheng, Z., Yang, X., Yupeng, H., Jiajing, W., Ju, R., Yaoxue, Z.: A Blockchain-Based Multi-Cloud Storage Data Auditing Scheme to Locate Faults. IEEE Transactions on Cloud Computing (Early Access), pp. 1-1. (2021).

[2] Kai, H., Chuanhe, H., Jinhai, W., Hao, Z., Xi, C., Yilong, Lu., Lian, Z., Bin, W.: An efficient public batch auditing protocol for data security in multi-cloud storage. In: 8th ChinaGrid Annual Conference, pp. 51–56. IEEE, Los Alamitos, CA, USA (2013).

[3] Yan, Z., Hongxin, H., Gail-Joon, A., Mengyang, Y.: Cooperative provable data possession for integrity verification in multicloud storage. IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 12, pp. 2231–2244. (2012).

[4] Huaqun, W.: Identity-based distributed provable data possession in multicloud storage. IEEE Transactions on Services Computing, vol. 8, no. 2, pp. 328–340. (2014).

[5] Xiaodong, Y., Xizhen, P., Meiding, W., Ting, L., Caifen, W.: Multi-replica and multi-cloud data public audit scheme based on blockchain. IEEE Access, vol. 8, pp. 144809-144822. (2020).

[6] Bin, L., Xiaoliang, Y., Shiping, C., Xiwei, X., Liming, Z.: Blockchain based data integrity service framework for IoT data. In: 24th International Conference on Web Services (ICWS), pp. 468-475. IEEE, Honolulu, HI, USA (2017).

[7] Yang, X., Ju, R., Yan, Z., Cheng, Z., Bo, S., Yaoxue, Z.: Blockchain empowered arbitrable data auditing scheme for network storage as a service. IEEE Transactions on Services Computing, vol. 13, no. 2, pp. 289–300. (2019).

[8] Yang, X., Cheng, Z., Guojun, W., Zheng, Q., Quanrun, Z.: A blockchain-enabled deduplicatable data auditing mechanism for network storage services. IEEE Transactions on Emerging Topics in Computing, vol. 9, no. 3, pp. 1421-1432. (2020)

[9] Yuan, Z., Chunxiang, X., Xiaodong, L., Xuemin, S.: Blockchain-based public integrity verification for cloud storage against procrastinating auditors. IEEE Transactions on Cloud Computing, vol. 9, no. 3, pp. 923-937. (2019).

[10] Wenting, S., Jing, Q., Jia, Y., Rong, H., Jian, H., Jinxin, M.: Data integrity auditing without private key storage for secure cloud storage. IEEE Transactions on Cloud Computing, vol. 9, no. 4, pp. 1408-1421. (2019).

[11] Boyang, W., Baochun, L., and Hui, L.: Oruta: Privacy-preserving public auditing for shared data in the cloud. IEEE Transactions on Cloud Computing, vol. 2, no. 1, pp. 43–56. (2014).

[12] Jingting, X., Chunxiang, X., Jining, Z., Jianfeng, M.: Identity-based public auditing for cloud storage systems against malicious auditors via blockchain. Science China Information Sciences, vol. 62, no. 3, p. 32104. (2019).

[13] Yong, Y., Liang, X., Man, A., Willy, S., Jianbing, Ni., Yafang, Z., Athanasios, V., Jian, S.: Cloud data integrity checking with an identity-based auditing mechanism from rsa. Future Generation Computer Systems, vol. 62, pp. 85–91. (2016).

[14] Yang, X., Cheng, Z., Quanrun, Z., Guojun, W., Ju, R., Yaoxue, Z.: Blockchain-enabled accountability mechanism against information leakage in vertical industry services. IEEE Transactions on Network Science and Engineering, vol. 8, no. 2, pp. 1202-1213. (2020).