ZKG INTERNATIONAL ISSN: 2366-1313

A Block Chain-Based Cloud Data Reduplication Technique

[1] Pathakota Sindhu, [2] Dr. B. Gohin,

[1] MCA Student, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

pathakotasindhu01@gmail.com

[2.] Associate Professor, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

Abstract: *To deal with the trouble of un accept as true with well worth in seaming entities faced in the manner of information reduplication in cloud storage surroundings, this paper proposes a cloud records reduplication scheme primarily based on block chain. Firstly, to make certain the trustworthiness of facts ownership proof inside the fact's reduplication procedure, a statistics ownership verification set of rules primarily based on random locations ampullinids signed with the assist of block chain era and Merkle hash tree; Secondly, an oblivious pseudo-random protocol is used to gain convergent encryption key to facilitate reduplication by cloud servers gives. Finally, the scheme safety is proved by using accomplishing safety analysis, while simulation experiments are carried out to verify the effectiveness of the scheme.*

Keywords— cloud storage; data reduplication; convergent encryption; block chain; Merkle hash tree

I. INTRODUCTION

With the fast improvement of cloud computing and sizeable popularity of smart cell devices, extra and more humans shop their information within the cloud. On the only hand, cloud computing has large storage area, that can efficaciously relieve the storage space of local hardware. Alternatively, users should use their statistics every time and anywhere with mobile terminals. In current years, in keeping with an IDC document [1], the information stored inside the cloud could be anticipated to attain 44ZT in 2025. Storing this kind of top notch quantity of data is a intense mission for cloud service carriers. Moreover, there may be a good deal redundancy in such large data, which seriously reduces the utilization of cloud storage servers. To successfully keep storage area and improve storage server utilization, many cloud carrier companies, which include Google, Amazon, and Drop box use records reduplication generation.

ISSN: 2366-1313

The records reduplication technology [2] refers to the cloud garage server for redundant data. Only one replica of the source data is stored, then this source records is shared with the person who has uploaded the identical statistics. It can correctly keep storage space and improve the utilization of hardware. However, within the untrustworthy environment of cloud computing, information reduplication faces severe protection problems inclusive of records leakage. To protect information privacy, users encrypt it earlier than uploading to the cloud; however one-of-a-kind users use their different private keys to encrypt the statistics, then will get different cipher text, which brings inconvenience to the cloud server provider for reduplication. Meanwhile, in the method of reduplication, the conventional model is that the cloud service provider and the next up loader verify the data ownership through the task-reaction protocol. Never the less, in the practices network environment, the cloud service company is also untrustworthy, and there may be collusive attacks with the up loader, which eventually results in the leakage of customers' personal statistics. To this give up, this paper proposes a cloud data reduplication scheme based totally on block chain. Our contributions are as follows:

In summary, in the complex cloud computing

Environment, there is an urgent want to find a at ease and effective cloud information reduplication option to correctly shield the privateers of data, and enhance the distance utilization of cloud storage servers.

(1) A venture-reaction protocol is constructed, and a Merkle hash tree is used to affirm the possession of facts between the following up loader and the cloud service provider. To make certain the trustworthiness of the verification, whilst no 0.33-celebration depended on institutions are brought, smart contract of block chain is used for verification, which guarantees the general public traceability of the outcomes.

(2) In order to guard the statistics in undeniable textual content of records, users interact with the key server via oblivious pseudo- random protocol to achieve encryption keys.

(3) Through safety evaluation, the proposed scheme satisfies the set desires; meanwhile, the simulation test proves this scheme can effectively enhance the reduplication performance.

II RELATEED WORK

Now, information protection reduplication has been hot subject matter among

ISSN: 2366-1313

educational and enterprise. In order to defend the privacy-preserving facts, Decretal.[3] first of all proposed the convergent encryption(CE) to acquire information safety reduplication. Bellare et al.[4] proposed message-lock encryption primarily based CE. Bellare et al.[5] designed a scheme Dupes, which attain records encryption with the aided key server. Kelechi et al.[6] proposed a facts reduplication scheme based on key server encryption. Next, Li et al.[7] proposed a statistics reduplication scheme based on reliable convergent key management for correctly dealing with the important thing. Young hood et al.[8] proposed a dispensed mystery key server help encryption-primarily based statistics reduplication scheme, which replaces a unmarried mystery key server with more than one ones to efficiently solve the single point of failure hassle so that knowing intra-consumer reduplication and inter- person reduplication.

Next, Tang et al.[9] proposed a re-encryption move-consumer reduplication scheme to make sure comfortable records reduplication by way of interacting with the consumer and the cloud provider without introducing 1/3-celebration entities. Yuan et al.[10] proposed a random tag-based information reduplication scheme that permits dynamic person management even as reduplicating statistics. After that, Guo et al. [11] proposed a random move-user reduplication scheme and designed a key alternate encryption algorithm based on ElGamal (Diffie-Hellman) key exchange encryption algorithm for data encryption, reduces the creation of the 1/3-birthday party servers. As referred to above, while making sure that the dictionary attack does no longer blast the plaintext, the papers introduce a third-birthday celebration entity such as a key server or a user. Moreover, the scheme thinks the 0.33-celebration entity is depended on. However, it is not possible in an herbal environment. So with block chain and facet computing development, Zhang et al.[12] proposed a block chain-primarily based auditing scheme, which calls for TPA to submit audit results on the block chain. Further, xu et al. [13] proposed a block chain-primarily based scheme that employs a no longer absolutely trusted TPA to implement information auditing underneath the supervision of the block chain. Next, Yuanetal. [14] proposed an OTP A public auditing scheme, which uses smart contracts to execute audit duties and truthful arbitration. In quick, almost all of these schemes treat the block chain as an easy black container.

Considering the integrity verification of

ISSN: 2366-1313

information, we must reap the accept as true with of information ownership proof. Halevi et at. [15] discovered there existed security hassle among the sub-add and CSP while they have interaction with every other within the process of statistics ownership evidence. The papers [16-18] each proposed some solutions to deal with the facts possession proof, but, present solutions all require the creation of 0.33-party servers, which might be susceptible to performance bottlenecks and different troubles, affecting the practicality of the answer in actual scenarios. So this paper proposed a statistics reduplication scheme based totally on block chain.

III PRELIMINARIES

A. Bilinear map

Let G1and G2be linear multiplicative cyclic groups of equal massive high order p, and g be a producing detail ofG1, and e be a bilinear mapping e: G1x G1->G2.

(1) Bilinearity: for any a, b belonging to Zp*and u, v

belongingtoG1, there is e (ua, vb) = e (u, v) ab.

(2) Non-degeneracy: e (u, v) not equal1.

(Three) Computability: there exist valid algorithms for any u,v

Belonging to G1 that can compute the value of e (u, v).

B. Smart contract

Block chain is an emerging technology. It Characteristics are decentralized, traceable, and tamper- obvious. The block chain ought to successfully remedy the UN trust worthy problem between all entities in the distributed surroundings. With the rapid development of block chain, an automated execution script has started to show up. It called a clever contract is a string of code deployed on block chain nodes that robotically executes when its necessities are met in line with its very own designed situations. With the benefits of block chain, smart contracts can successfully prevent the problem of miss trust between entities. Meanwhile, to achieve an open and transparent verification, it facts the authentication procedure in the block.

IV SYSTEM DESCRPTION

System model

As shown in Figure 1, the scheme includes four entities: key server (KS), data owner (DO), cloud server provider (CSP), and block chain (BC).

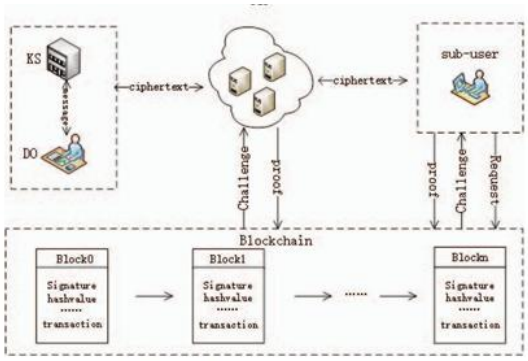

Figure1.Systemmodel

(1) The key server in particular generates public device parameters based

ISSN: 2366-1313

on safety parameters and then aids the DO to attain convergent encryption key by way of oblivious pseudo-random characteristic protocol.

(2) The statistics owner is split into the primary consumer and subsequent person. The one is a person who first uploads the statistics. Moreover, the other one is someone who uploads the facts to CSP, however the records already exists in CSP.

(3) The cloud service issuer particularly stores the records uploaded by way of the facts owner and authenticates the ownership of the facts for the following up loaders preserving reproduction data.

(Four) The most important mission of block chain is to verify the correctness of statistics possession evidence outcomes

B. Security model

This state of affairs is especially conducted in a semi-depended on surroundings, assuming that the records owner and the cloud service issuer are semi-relied on entities, while amongst them the important thing server is a fully relied on entity. Then there are three major adversaries gift in the paper as follows:

(1) Cloud service provider, wherein the CSP can also have snooping and reasoning about the user's plaintext data statistics in the course of the sincere execution of information de- duplication.

(2) Subsequent up loaders: In the process of appearing facts reduplication, to infer greater facts privateers records that does not belong to them.

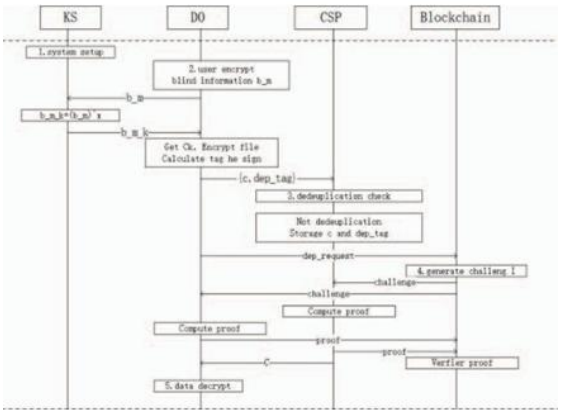

Figure2.The work flow of our scheme

When the cloud service provider receives the reduplication tag deep from the user, it checks its own stored reduplication list, and if there is no duplicate, it prompts the user to continue data uploading while the cloud service provider adds the reduplication tag to the reduplication list; if there is duplicate; it prompts the user to authenticate data ownership. First the subsequent up loader sends a data ownership authentication request to the block chain, which includes file name and file length $\{f_{name}, f_{length}\}$.

When the block chain receives the request sent by the user, then it starts to execute the challenge information production, which is mainly a random selection of a random set I of the number of leaf nodes of the complete binary tree, constituting the set challenge, and then sends it to the

ISSN: 2366-1313

cloud service provider and subsequent up loaders respectively.

V PERFORMANCE

Theoretical Analysis

This section analyses the computational cost and storage cost of this scheme and other schemes. The computational cost mainly includes four aspects of setup, prove and verify, as shown in table 1.

TABLE I COMPARISONSOFCOMPUTATIONCOSTS

Schemes	*setup*	*prove*	*verify*
Scheme[18]	$N(H+2EXP)$	*NO*	(2n+1)EXP+nH+2Pair
Our scheme	H+2EXP+SE	$n_c(H)$	*EQ*

TABLE II. COMPARISONS OF COMMUNICATION AND STORAGE

Schemes	*Upload.ini*	*Upload.sub*	*Block chain*
Scheme[18]	$N\|G\|+\|F\|$	$4\|Z_p\|$	$(2n_c+4n_s)\|ZP\|+n_c(y+1)\|G\|$
Our scheme	$N\|G\|+\|F\|$	$3\|Z_p\|$	$2n_f\|F_c\|$

Implementation

In this subsection, our experiment is run at a window system, where the processor is Intel Corei5, and the memory is 16G. Cloud storage server uses Ten cent cloud server, where the memory is 4G and storage is 80 GB, bandwidth is 4Mbps. Smart contract in the block chain is written using solidity language. The programming uses the Python3 language and the PYPBC (Python Pairing-Based Crypto Library) function library. The element's size in Zpisset as $|p|$=160 bits. The hash function is SHA-256, the symmetric encryption algorithm is AES-CBC, and the key is 128 bits.

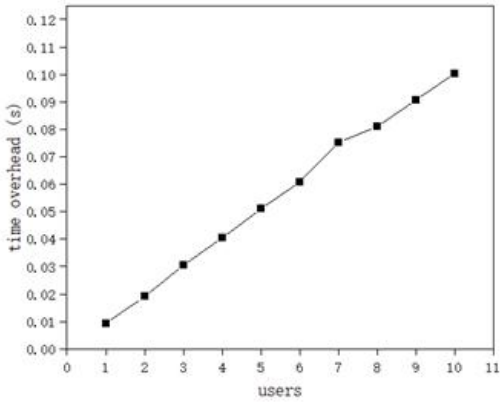

Figure3.Systemsetup

We first measure the performance of system initialization with a different number of users. It can be known from the fig.3 that the time overhead is increase with the user numbers.

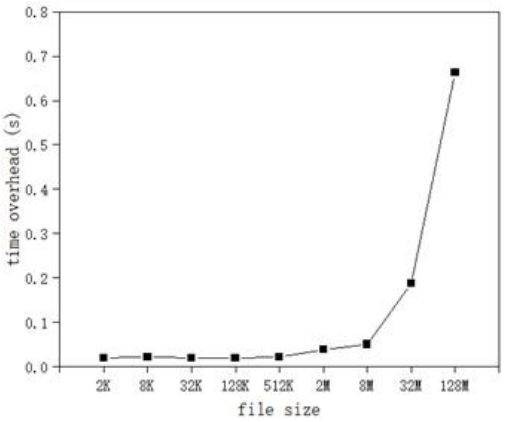

Figure4.Convergent key generation

Fig.4 tests the convergence key generation time. From the fig.4, we know that the time cost of the convergence key generation increases when the encrypted file size is from 2Kto 128M.

ISSN: 2366-1313

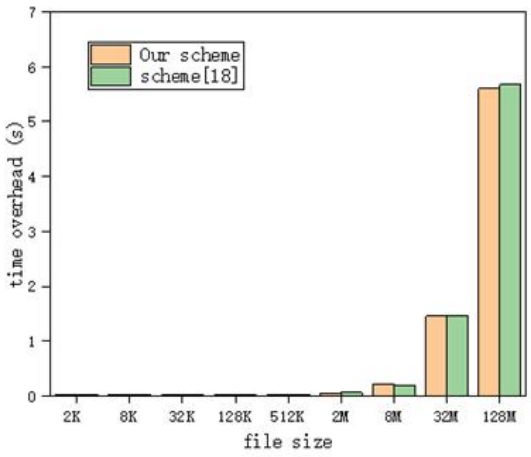

Figure5.Data encryption

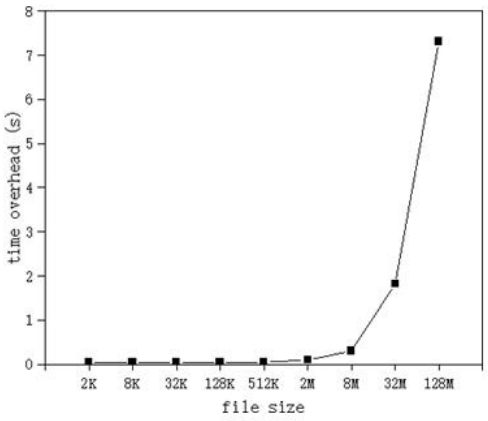

Figure7.Datadeduplication

Finally, we measure the data reduplication time of our schema. This experiment sets the block chain node number as ten, and data sizes are from 2K to 128M. As shown inFig.7, the time cost of the data reduplication process gradually increases as the file size increases, and the primary consumption is the computation of the value of the merkle hash tree.

V CONCLUSION

To address the trouble of UN consider worthiness within the system of statistics possession proof for the duration of facts reduplication within the complicated cloud garage surroundings. This paper proposes a block chain-based cloud records reduplication scheme through employing a Merkle hash tree and smart agreement generation in the block chain. To remedy the trustworthiness hassle without introducing different 0.33-birthday celebration companies, next user, and cloud carrier carriers conduct mutual authentication of statistics possession based on the concept of task-evidence protocol. Further extra, to confirm the results via clever contracts, which makes the verification consequences public and accountable. Next, to make certain that the cipher text with low data entropy will now not burst, the encryption secret's acquired the usage of the oblivious pseudo-random protocol. Finally, the security evaluation and overall performance prove that this scheme correctly solves the untrustworthiness hassle in records possession authentication.

REFERENCES

[1]Cisco global cloud index: Forecast and methodology, 2016-2021white https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html.

[2] J. Gantz and D. Reinsel, (2020) 'The digital universe decade-are youready,'https://hk.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf.

[3] Douceur, J., Adya, A., Bolosky, W., Simon, D., Theimer, M..Reclaiming space from duplicate files in a server less distributed file system. In: Proc. of IEEE TDCS, pp. 617–624, 2002.

[4] Prasadu Peddi (2019), "Data Pull out and facts unearthing in biological Databases", International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.

[5] Bellare, M., Keelveedhi, S., 'Interactive message-locked encryption and secure reduplication. In: Proc. of Springer PKC, pp. 516–538, 2015.

[6] Keelveedhi, S., Bellare, M., Ristenpart, T., 'Dup LESS: server-aided encryption for reduplicated storage. In: Proc. of USENIX Security Symposium, Washington, D. C., USA, pp. 179–194,2013.J.

[7] Li, X. Chen, M. Li, J. Li, P.P. Lee, W. Lou, 'Secure reduplication with efficient and reliable convergent key management, IEEE Transactions on Parallel and Distributed Systems, Vol.25, No 6, pp.1615–1625,2014.

[8] Shin, Y., Koo, D., Yun, J., & Hur, J. 'Decentralized Server-aided Encryption for Secure Deduplicationin Cloud Storage. Vol.13, No.6, pp.1021-1033, 2020.

[9] Tang, X., Zhou, L., Huang, Y., Chang, C., 'Efficient cross-user reduplication of encrypted data through re-encryption'. In: IEEE Trust Com/Big Data SE, New York, NY, pp. 897–904, 2018.

[10] Yuan, H., Chen, X., Jiang, T., Zhang, X., Yan, Z., Xiang, Y., 'Dedup DUM Secure and scalable data reduplication with dynamic user management. Inf. Sci. Vol.456, pp.159–173, 2018.

[11] Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp:23-37.