# Summary of Data Storage Technology in Cloud Storage Service

Chunxia Tu[1, a], Zhonbing Yuan[2, b] and Xiaojun Liu[2, c*]

[1]School of Computer, Huanggang Normal University, Hubei Huanggang, China

[2]School of Electronic & Information, Huanggang Normal University, Hubei Huanggang, China

[a]18623582@qq.com, [b]1138002283@qq.com, [c]whutliuxiaojun@126.com

*The corresponding author

**Keywords:** Cloud storage; Storage security; Data holding verification; Data recovery

**Abstract.** This paper discusses the security and reliability of cloud storage service demand and the cloud storage service of provable data possession and recovery plan special requirements. Review can prove data holdings and recovery technology in the domestic and foreign research status, so as to identify the cloud storage environment provable data possession and recovery technology research direction.

## Introduction

In the cloud storage service, encryption technology can complete the data confidentiality, data possession (provable data possession, PDP) technology can realize data integrity, cloud service providers can prove to the user that no any unauthorized tampering or delete user data stored in the cloud, and the user's data is complete. The existing can prove that data hold scheme is divided into three categories: PDP scheme, POR scheme and the scheme based on the trusted third party.

## PDP Scheme.

Ateniese et al first defines the PDP scheme of [1], and puts forward 2 specific PDP program. The user to preprocess the documents, the document is divided into several data blocks, each block generates a homomorphic verification tag, the tag together with data stored in the server. Verify when users choose some random a challenge to the server, the server returns for holding these pieces of evidence, according to the user private key certificate calibration server returns whether the evidence is accurate. Because the label has a homomorphism, can be superimposed on each other, check both the amount of calculation is small, and allows for unlimited check. But their scheme uses a modular exponentiation in RSA in the generation of data, did not consider the data update, nor anti multiple server collusion attack, is not suitable for multi copy protocols. Then, they will be improved to support public homomorphic Tags Check the homomorphic tags, check that no private key, as long as you have your public key can prove the integrity of the document. The two schemes using RSA technology, file pre treatment process in computational overhead is significant. Later, they put forward the multi replica PDP Scheme [2], by increasing the copy of the document to eliminate PDP scheme file pre process, but check file integrity need multiple copies also in response to the challenges, files are large computational overhead is quite large.

Erway[3] et al Proposed 2 kinds of dynamic data holding dynamic DPDP (PDP), in order to achieve data update. Based on the RSA tree structure, a class of differential jump table is based on the tree structure. Their main contribution is to achieve the dynamic update of the data held. But because the whole project is still based on RSA modular index operation, the computation overhead is very big.

Curtmola et al [4] integrated forward error correction code to the PDP program, proposed a multi copy PDP protocol, the original file exchange location, from which to select a part of the RS encoding, so as to improve the coding efficiency. Since the attacker does not know what the redundant code is calculated from the block, so that the security of the program can be improved. But the disadvantage of this protocol is that only the data owners can verify the integrity of the data.

 

Check S-PDP [5] provides the basic idea is that the user will file is divided into fixed size blocks, the file block is divided into several groups, each group to calculate a hash value. Check file integrity checks a file block, requiring service providers to compute a proof, the proof and the original hash value to determine whether the file. Each file integrity checking to consume a hash value, so can only be limited. But it supports three and block as a unit of update operations (modify, delete, add, update the file block and the need to modify all the hash value calculation process is complicated, the communication overhead.

## POR Scheme

The basic idea of POR scheme is the user in the file random insertion and a plurality of file data do not distinguish the "sentinel", check file integrity requires service providers to return these random position of "sentinel", through the effectiveness of verification "sentinel" determine whether the documents are complete. In order to prevent replay attacks, each "sentinel" can only use a, each file integrity check to consume a group of "sentinel", so it can only make limited check. In processing file pre process have been identified on the position of the "sentinel" late does not allow files on any form of updates.

POR to achieve two goals: graders can use household timing, timely understanding of the cloud data of complete state; (2) when data is corrupted or lost less than a certain proportion, can guarantee to a higher probability of repair data integrity.POR generally the original data file first coding redundancy, although after in the original data file into a certain number of authentication data element, with the original files together into the cloud. Through "challenge response" mechanism of randomly selected part of the authentication data element, to some success rate detected data error, and use the file encoding provides redundancy recovery original data.

Juels et al [6] proposed a formal model and the definition of POR, and put forward based on the "sentinel" POR scheme. The basic idea is, first of all will use file encryption and error correcting code encoding, the file is divided into several data blocks, each block between the random insertion generated by the hash function with key "sentinel" when asked to prove every challenge; the verifier return random bits on a certain number of sentinels, to detect the integrity of the file to verify the integrity of the sentry; combined with the error correction encoding with a certain probability to ensure the file can be retrieved. They prove that if the probability is greater than a certain value to the server in response file is recovery. The advantages of the scheme is to have the extra storage overhead storage the small, small calculation overhead challenge and response, but due to the limited number of sentinel insertion and can only be challenged Once, therefore, only to support the integrity of the check. In order to ensure the privacy of the sentinel, the need for file encryption, resulting in a large file read overhead.

Shacham et al use homomorphic tags present 2 a POR Scheme [7], a is based on pseudo random function, and does not support the public verification; another based on BLS signatures to support public verification. The two schemes are using erasure code, support an unlimited number of challenges, but certified storage overhead is large, also did not consider dynamic updating of data.

Dodis et al [8] proposed a general POR framework for the first time, and a formal analysis was carried out, and the method of converting POR code into POR scheme was given.But their scheme does not consider the dynamic update of the data.

Bowers et al [9] put forward the theoretical framework of a design POR, based on damage degree when more easily detected this idea of, to one of the original cause POR scheme is improved, less storage cost and higher error rates. But their scheme does not support file update and publicly verifiable.

## Comparative Analysis of PDP and POR Schemes

The existing PDP and POR schemes have the following defects:

 

⑴The vast majority of the scheme is a technology based on public key cryptography, the large computational overhead;

⑵Most of the program cannot support the dynamic updating of data;

⑶The vast majority of programs only limited data have verified;

⑷Most of the schemes do not support public verification;

⑸The little scheme considering data recovery technology, detect the damage did not recover;

⑹The many schemes are not suitable for cloud storage service environment, scheme's computation complexity and the communication complexity is very high, the efficiency is relatively low.

⑺The test method of the trusted third party based on the coordination is not sufficient. In the actual cloud storage environment, how to achieve trusted third party inspection still need to be further studied.

Secondly, the performance of the commonly used PDP and POR schemes are compared as shown in Table 1 (C/S: User / service provider; Update operation, A: Additional, I: Insert, M: Modify, D: Delete ).

Table 1 Performance comparison of PDP and POR

| Programme | Check number | Public check | Storage overhead (C/S) | Computational overhead (C/S) | | Communication overhead | update operation |
|---|---|---|---|---|---|---|---|
| | | | | check | update | | |
| PDP | infinite | NO | O(1)/O(n) | O(1)/O(1) | | O(1) | |
| POS | infinite | YES | O(1)/O(n) | O(n)/O(n) | | O(1) | |
| DPDP | infinite | NO | O(1)/O(n) | O(logn)/O(logn) | O(logn)/O(logn) | O(logn) | A\I\M\D |
| S-PDP | finite | NO | O(1)/O(n) | O(1)/O(1) | O(1)/O(n) | O(1) | A\M\D |
| POR | finite | NO | O(1)/O(t) | O(1)/O(1) | | O(1) | |

## Scheme Based on Trusted Third Party

In order to reduce the client and service provider for the end of the computation complexity and communication complexity, 2011. Paper [10] in the previous POR scheme based on the increased a trusted third party is improved, the user can integrity check and dynamic update of its files stored in different trusted cloud server. At the same time, the scheme is extended to multi user shared situation, support multiple users access the same file on the cloud storage server, check and update. Since most of the user side storage, computing and communication overhead is transferred to the trusted third party, its computing power and storage space needs to reduce, more in line with the needs of the computing environment, lightweight customers. In 2012. Paper [11] proposed a applicable to archival storage of lightweight data can retrieve that algorithm (light-weight POR, L-POR and redundant data produced by the algorithm in coding with user authentication message, so as to avoid the other similar algorithms due to the insertion of additional recognition certificate metadata storage overhead, which greatly improves the authentication efficiency. In order to solve cloud storage in existing data hold checking scheme on the efficiency and function of defects is proposed based on implicit trusted third party data possession audit architecture (AITTP audit with implicit trusted third party).With trusted hardware as implicitly trusted third party, instead of the user performs data possession checking, and generate can display tamper resistant trusted audit logs, audit logs stored in the cloud server, through measures of cryptography and tripartite interaction to ensure that the daily log integrity.

## Conclusion

Currently, cloud storage service data can prove that there are three solutions to the storage technology, in which the PDP and POR programs have different degrees of defects.In recent years based on trusted third party program development rapidly, and trusted hardware and the cloud storage servers integrated in together, can effectively solve the problem of reality in the trusted third party is difficult to

 

achieve. Therefore, based on the trusted third party solutions will be the future development direction of the mainstream.

## Acknowledgment

The paper was supported by the project of excellent ICT engineer training program in Huanggang Normal University (Grant No. 2014zy04); and the project of zxfz2016A014 in Huanggang Normal University

## References

[1] Ateniese G, Burns R, Curtmola R, Herring J, et al. Provable data possession at untrusted stores. In P. Ning, S. De Capitani di Vimercati, and P. Syverson, editors, ACM Conference on Computer and Communication Security (CCS '07). ACM Press, 2007.

[2] Ateniese G, Pietro RD, Mancini RV, Tsudik G. Scalable and efficient provable data possession. In Proc. of SecureComm 2008.

[3] Erway C, Kupcu A, Papamanthou C, and Tamassia R. Dynamic provable data possession. In ACM conference on Computer and communications security (CCS '09), 2009. ACM.

[4] Curtmola R, Khan O, Burns R, and Ateniese G. MR-PDP: Multiple-replica provable data possession. In: Proc. of ICDCS '08, pp.411-420, 2008.

[5] Swaminathana A, Mao YN, Su GM, et al. Confidentiality preserving rank-ordered search[C]// Proc of the 2007 ACM Workshop on storage security and survivability(StorageSS'07). New York, NY, USA: ACM, 2007:7-12.

[6] Juels A,Kaliski BS. PORs: Proofs of retrievability for large files// Proc of CCS '07. New York: ACM, 2007:584-597.

[7] Shacham H, Waters B. Compact proofs of retrievability. In 14th ASIACRYPT, 2008:90-107.

[8] Dodis Y, Vadhan S, Wichs D. Proofs of retrievability via hardness amplification. In Theory of Cryptography Conference, volume 5444 of LNCS, pp. 109-127. Springer, 2009.

[9] Bowers K, Juels A, Opra A. Poofs of retrievability: Theory and implementation. Technical Report 2008/175, Cryptology ePrint Archive, 2008.

[10] Liu Feifei. Research on provable data possession in cloud computing environment[D], Shanghai Jiaotong University, 2011

[11] An Baoyu. Research on the key technologies of data integrity protection in cloud storage[D], Beijing University of Post and Telecommunication,2012

 