International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 1 Issue 1 –AUGUST 2013.

Multi-Cloud Storage based Integrity Verification System using Batch Auditing

Grandhi Prasuna[#1], Bandla Srinivasa Rao[*2]

[#]*PG Scholar, Dept. of CSE, VRS & YRN College of Engg. & Technology, Chirala, AP*

[1] *ch.prasuna@yahoo.com*

[*]*Associate Professor & HOD, Dept. of CSE, VRS & YRN College of Engg. & Technology, Chirala, AP*

***Abstract*— Cloud computing provides the increased level of scalability, availability and durability services to be easily consumed over the Internet on an as-needed basis for the customers. The more copies are asked to store by the service provider in cloud. For that more amounts has been charged from the customers. We introduce the new technique as Provable Data Possession (PDP) for ensuring the integrity of data in storage outsourcing. In this paper, we address the efficient scheme of PDP and its distributed cloud storage to support the qualities of the services provides. In which we consider and maintains the existence of multiple cloud service providers to the cooperatively stores and it's also maintain the clients data in safe and security. We present a cooperative PDP (CPDP) scheme based on homomorphic verifiable response and hash index hierarchy. In this paper, we prove the security of our scheme based on the multi-prover zero-knowledge proof performance optimization mechanisms for our scheme. In particularly, increasing the efficiency for our scheme and minimize the system, which can satisfy completeness, knowledge soundness, and zero-knowledge properties. In addition, we clear the cost of clients and also storage service providers. Our experiments show that our solution introduces lower computation and communication overheads in comparison with non-cooperative approaches.**

***Keywords*— Storage Security, Provable Data Possession, Interactive Protocol, Zero-knowledge, Multiple Cloud, Cooperative**

I. INTRODUCTION

In recent years, cloud storage services has become a faster growth and makes huge profits by providing the services to the customer for their needs on the internet and data sharing. Cloud provides the services at low cost, scalable position-independent platform for the storage of client's data. Though it provides good services, its environments are based on the open architecture and with interfaces. It has capabilities to incorporate with multiple internal or external services together to provide ability to exchange data's and use information so that we call such a distributed cloud environments as a multicloud or hybrid cloud.

There are many existence tools and technologies are used as platform for cloud services provides. For example, VM Orchestrator is a used as platform. This will help the cloud providers to construct a distributed cloud storage platform (DCSP) for maintaining the client's data.

However it has good platform and provides good services. The main problem is the platform is vulnerable to security attacks it will leads to irretrievable data losses to the clients. Therefore it is essential for the cloud service to provide a high security technique for managing their storage services. So that, we introduce Provable data possession (PDP) as a new technique for a cloud storage provider to prove the integrity and ownership of clients' data without downloading data. The proof checking is used to find the damage or any missing data in the large size files and folders without downloading the latest version data. It is able to replace traditional hash and signature functions in storage outsourcing. Various PDP schemes have been recently proposed, such as Scalable PDP and Dynamic PDP. However, these schemes mainly focus on PDP issues at untrusted servers in a single cloud storage provider and are not suitable for a multi-cloud environment.

We introduce this scheme for the motivation to provide a low-cost, scalable, location independent platform for managing clients' data, current cloud storage systems adopt several new distributed file systems. These file systems share some similar features: a single metadata server provides centralized management by a global namespace; files are split into blocks or chunks and stored on block servers; and the systems are comprised of interconnected clusters of block servers. Those features enable cloud service providers to store and process large amounts of data. However, it is crucial to offer an efficient verification on the integrity and availability of stored data for detecting faults and automatic recovery. Moreover, this verification is necessary to provide reliability by automatically maintaining multiple copies of data and automatically redeploying processing logic in the event of failures.

Many existing schemes can make their true or false decision for the data possession without downloading the data at any untrusted stores because they are suitable for the distributed cloud storage environment.

There are two schemes are taken for the examples one of the schemes based on Merkle Hash tree (MHT), such as DPDP-I, DPDP-II and SPDP and they did not provide any algorithms for constructing distributed Merkle trees that are necessary for efficient verification in a multi-cloud environment. The other scheme is such as PDP, CPOR-I and CPOR-II are constructed on homomorphic verification tags,

by which the server can generate tags for multiple file blocks in terms of a single response value.

Even though it as many existing schemes have addressed various security properties such as public verifiability, scalability and privacy preservation. But we still need a careful consideration in the security level. There are two types of attacks are happening frequently. One is Data Leakage Attack and other is Tag Forgery Attack. In the first attack, adversary can easily obtain the stored data through verification process after running or wiretapping sufficient verification communications and in the second attack, which a dishonest CSP can deceive the clients. These two attacks may cause potential risks for privacy leakage and ownership cheating. Although there are many security models are proposed in the existing PDP schemes, these cannot cover all security requirements. So that many threading occur in the client's data and cloud storage service providers.

To overcome the above issues, we proposed a new technique in this paper namely has Cooperative Provable Data Possession (CPDP). In this paper we addressed all problems in the PDP and to overcome by following aspects: high security, transparent verification, and high performance.

To achieve the above goals we introduce the two techniques for the verification framework for multicloud storage. The two techniques are: hash index hierarchy (HIH) and homomorphic verifiable response (HVR).

We then demonstrate that the possibility of constructing a cooperative PDP (CPDP) scheme without compromising data privacy based on modern cryptographic techniques, such as interactive proof system (IPS). We further introduce an effective construction of CPDP scheme using above-mentioned structure. Moreover, we give a security analysis of our CPDP scheme from the IPS model. We prove that this construction is a multi-prover zero-knowledge proof system (MP-ZKPS) [11], which has completeness, knowledge soundness, and zero-knowledge properties. These properties ensure that CPDP scheme can implement the security against data leakage attack and tag forgery attack. Our experiments show that our solution introduces very limited computation and communication overheads.

The rest of the paper is organized as follows. In Section II, we discuss about the related work of the PDP. In Section III formally introduces our proposed system of the paper. In Section IV we summarize about the algorithm used for simulation. In Section V, we present the full simulation study of the proposed scheme. Finally, we conclude the paper and discuss future work in Section VI.

II. Related Works

Ayad F. Barsoum and M. Anwar Hasan [21], proposes a pairing based provable multi-copy data possession (PB-PMDP) scheme, which provides evidence to the customers that all outsourced copies are actually stored and remain intact. Moreover, it allows authorized users (i.e., those who have the right to access the owner's file) to seamlessly access the file copies stored by the CSP, and supports public verifiability. The proposed scheme is proved to be secure against colluding servers. We illustrate the performance of the PB-PMDP scheme through theoretical analysis, which is validated by experimental results. The verification time of our scheme is practically independent of the number of file copies. Additionally, we discuss how to identify corrupted copies by slightly modifying the proposed PB-PMDP scheme.

N.Janardhan, Y.Rajasree, and R .Himaja [22], proposes a address the construction of an efficient PDP scheme for distributed cloud storage to support the scalability of service and data migration, in which we consider the existence of multiple cloud service providers to cooperatively store and maintain the clients' data .Offering strong data protection to cloud users while enabling rich applications is a challenging task. We explore a new cloud platform architecture called Data Protection as a Service, which dramatically reduces the per-application development effort required to offer data protection, while still allowing rapid development and maintenance. We prove the security of our scheme and we also provide distributed auditing mechanisms. We provide extensive experimental studies that demonstrate the efficiency and effectiveness of the proposed approaches.

Thanasis G. Papaioannou, Nicolas Bonvin and Karl Aberer [23], in this paper1, we introduce Scalia, a cloud storage brokerage solution that continuously adapts the placement of data based on its access pattern and subject to optimization objectives, such as storage costs. Scalia efficiently considers repositioning of only selected objects that may significantly lower the storage cost. By extensive simulation experiments, we prove the cost-effectiveness of Scalia against static placements and its proximity to the ideal data placement in various scenarios of data access patterns, of available cloud storage solutions and of failures.

Singh.Y. [24], In this work we observed that, from a customer's point of view, relying upon a solo SP for his outsourced data is not very promising. In addition, providing better privacy as well as ensuring data availability can be achieved by dividing the user's data block into data pieces and distributing them among the available SPs in such a way that no less than a threshold number of SPs can take part in successful retrieval of the whole data block. In this paper, we propose a secured cost-effective multi-cloud storage (SCMCS) model in cloud computing which holds an economical distribution of data among the available SPs in the market, to provide customers with data availability as well as secure storage. Our results show that, our proposed model provides a better decision for customers according to their available budgets.

To check the availability and integrity of outsourced data in cloud storages, researchers have proposed two basic approaches called Provable Data Possession (PDP) [2] and Proofs of Retrievability (POR) [3]. Ateniese et al. [2] first proposed the PDP model for ensuring possession of files on untrusted storages and provided an RSA-based scheme for a

static case that achieves the $O(1)$ communication cost. They also proposed a publicly verifiable version, which allows anyone, not just the owner, to challenge the server for data possession. This property greatly extended application areas of PDP protocol due to the separation of data owners and the users. However, separation of data owners and the users. However, dynamic scenarios because of the dependencies on the index of blocks. Moreover, they do not fit for multi-cloud storage due to the loss of homomorphism property in the verification process. In order to support dynamic data operations, Ateniese et al. developed a dynamic PDP solution called Scalable PDP [4]. They proposed a lightweight PDP scheme based on cryptographic hash function and symmetric key encryption, but the servers can deceive the owners by using previous metadata or responses due to the lack of randomness in the challenges. The numbers of updates and challenges are limited and fixed in advance and users cannot perform block insertions anywhere. Based on this work, Erway et al. [5] introduced two Dynamic PDP schemes with a hash function tree to realize $O(\log n)$ communication and computational costs for a n-block file. The basic scheme, called DPDP-I, retains the drawback of Scalable PDP, and in the 'blockless' scheme, called DPDPII, the data blocks $\{mij\}j\in[1,t]$ can be leaked by the response of a challenge, $M = \Sigma tj=1\ ajmij$, where aj is a random challenge value. Furthermore, these schemes are also not effective for a multi-cloud environment because the verification path of the challenge block cannot be stored completely in a cloud [8]. Juels and Kaliski [3] presented a POR scheme, which relies largely on preprocessing steps that the client conducts before sending a file to a CSP. Unfortunately, these operations prevent any efficient extension for updating data. Shacham and Waters [6] proposed an improved version of this protocol called Compact POR, which uses homomorphic property to aggregate a proof into $O(1)$ authenticator value and $O(t)$ computation cost for t challenge blocks, but their solution is also static and could not prevent the leakage of data blocks in the verification process. Wang et al. [7] presented a dynamic scheme with $O(\log n)$ cost by integrating the Compact POR scheme and Merkle Hash Tree (MHT) into the DPDP. Furthermore, several POR schemes and models have been recently proposed including [9], [10]. In [9] Bowers et al. introduced a distributed cryptographic system that allows a set of servers to solve the PDP problem. This system is based on an integrity-protected error correcting code (IP-ECC), which improves the security and efficiency of existing tools, like POR. However, a file must be transformed into l distinct segments with the same length, which are distributed across l servers. Hence, this system is more suitable for RAID rather than cloud storage.

III. PROPOSED WORK

Although there are many security models are proposed in the existing PDP schemes, these cannot cover all security requirements. So that many threading occur in the client's data and cloud storage service providers. To overcome the above issues, we proposed a new technique in this paper namely has Cooperative Provable Data Possession (CPDP). In this paper we addressed all problems in the PDP and to overcome by following aspects: high security, transparent verification, and high performance. To achieve the above goals we introduce the two techniques for the verification framework for multicloud storage. The two techniques are: hash index hierarchy (HIH) and homomorphic verifiable response (HVR).

We then demonstrate that the possibility of constructing a cooperative PDP (CPDP) scheme without compromising data privacy based on modern cryptographic techniques, such as interactive proof system (IPS). We further introduce an effective construction of CPDP scheme using above-mentioned structure. Moreover, we give a security analysis of our CPDP scheme from the IPS model. We prove that this construction is a multi-prover zero-knowledge proof system (MP-ZKPS), which has completeness, knowledge soundness, and zero-knowledge properties. These properties ensure that CPDP scheme can implement the security against data leakage attack and tag forgery attack. Our experiments show that our solution introduces very limited computation and communication overheads.

IV. EXPERIMENTAL RESULTS

We have simulated our system in Java. We implemented and tested with a system configuration on Intel Dual Core processor, Windows XP and using Netbeans 7.0. We have used the following modules in our implementation part. The details of each module for this system are as follows. We have implemented and tested with the 5 moudles.

Multi cloud storage

Distributed computing is used to refer to any large collaboration in which many individual personal computer owners allow some of their computer's processing time to be put at the service of a large problem. In our system the each cloud admin consist of data blocks. The cloud user uploads the data into multi cloud. cloud computing environment is constructed based on open architectures and interfaces, it has the capability to incorporate multiple internal and/or external cloud services together to provide high interoperability. We call such a distributed cloud environment as a *multi-Cloud* .A multi-cloud allows clients to easily access his/her resources remotely through interfaces.

Cooperative PDP

Cooperative PDP (CPDP) schemes adopting zero-knowledge property and three-layered index hierarchy, respectively. In particular efficient method for selecting the optimal number of sectors in each block to minimize the computation costs of clients and storage service providers. Cooperative PDP (CPDP) scheme without compromising data privacy based on modern cryptographic techniques

Data Integrity

Data Integrity is very important in database operations in particular and Data warehousing and Business intelligence in general. Because Data Integrity ensured that data is of high quality, correct, consistent and accessible.

Third Party Auditor

Trusted Third Party (TTP) who is trusted to store verification parameters and offer public query services for these parameters. In our system the Trusted Third Party, view the user data blocks and uploaded to the distributed cloud. In distributed cloud environment each cloud has user data blocks. If any Modification tried by cloud owner an alert is send to the Trusted Third Party.

Cloud User

The Cloud User who has a large amount of data to be stored in multiple clouds and have the permissions to access and manipulate stored data. The User's Data is converted into data blocks. The data blocks are uploaded to the cloud. The TPA views the data blocks and Uploaded in multi cloud. The user can update the uploaded data. If the user wants to download their files, the data's in multi cloud is integrated and downloaded.

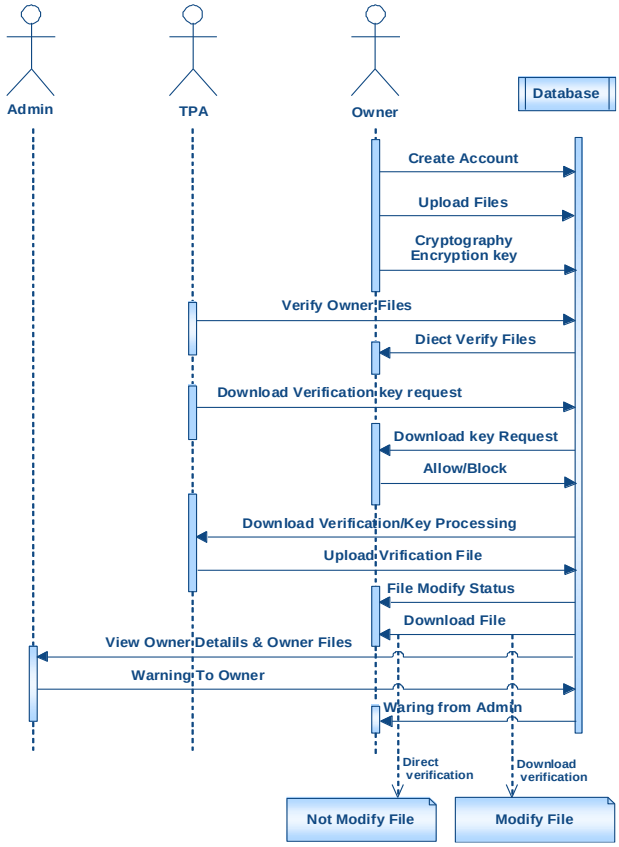

Figure 1: Sequence Diagram of our system

V. Conclusion and Future Works

In this paper, we proposed the scheme of PDP and its distributed cloud storage to support the qualities of the services provides. In which we consider and maintains the existence of multiple cloud service providers to the cooperatively stores and it's also maintain the clients data in safe and security. We presented a cooperative PDP (CPDP) scheme based on homomorphic verifiable response and hash index hierarchy. In this paper, we proved that the security of our scheme based on the multi-prover zero-knowledge proof performance optimization mechanisms for our scheme. In particularly, increasing the efficiency for our scheme and minimize the system, which can satisfy completeness, knowledge soundness, and zero-knowledge properties. In addition, we reduce the cost of clients according to the usage and also storage service providers. Our experiments show that our solution introduces lower computation and communication overheads in comparison with non-cooperative approaches.

For future work, we would like to extend our work with more effective CPDP constructions. First in our experiment for large files, is affected by the bilinear mapping operations due to its high complexity. To solve this problem, RSA based constructions may be a better choice, but this is still a challenging task because the existing RSA based schemes have too many restrictions on the performance and security. Secondly, from a practical point of view, we still need to address some issues about integrating our CPDP scheme smoothly with existing systems. And finally, it is still a challenging problem for the generation of tags with the length irrelevant to the size of data blocks. We would explore such a issue to provide the support of variable-length block verification.

VI. References

[1] B. Sotomayor, R. S. Montero, I. M. Llorente, and I. T. Foster, "Virtual infrastructure management in private and hybrid clouds," IEEE Internet Computing, vol. 13, no. 5, pp. 14–22, 2009.

[2] G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, L. Kissner, Z. N. J. Peterson, and D. X. Song, "Provable data possession at untrusted stores," in ACM Conference on Computer and Communications Security, P. Ning, S. D. C. di Vimercati, and P. F. Syverson, Eds. ACM, 2007, pp. 598–609.

[3] Juels and B. S. K. Jr., "Pors: proofs of retrievability for large files," in ACMConference on Computer and Communications Security, P. Ning, S. D. C. di Vimercati, and P. F. Syverson, Eds. ACM, 2007, pp. 584–597.

[4] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in Proceedings of the 4th international conference on Security and privacy in communication netowrks, SecureComm, 2008, pp. 1–10.

[5] C. C. Erway, A. K¨upc¸ ¨u, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in ACM Conference on Computer and Communications Security, E. Al-Shaer, S. Jha, and A. D. Keromytis, Eds. ACM, 2009, pp. 213–222.

[6] H. Shacham and B. Waters, "Compact proofs of retrievability," in ASIACRYPT, ser. Lecture Notes in Computer Science, J. Pieprzyk, Ed., vol. 5350. Springer, 2008, pp. 90–107.

[7] Q. Wang, C.Wang, J. Li, K. Ren, and W. Lou, "Enabling public verifiability and data dynamics for storage security in cloud computing," in ESORICS, ser. Lecture Notes in Computer Science, M. Backes and P. Ning, Eds., vol. 5789. Springer, 2009, pp. 355–370.

[8] Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu, and S. S. Yau, "Dynamic audit services for integrity verification of outsourced storages in clouds," in SAC, W. C. Chu, W. E. Wong, M. J. Palakal, and C.-C. Hung, Eds. ACM, 2011, pp. 1550–1557.

[9] K. D. Bowers, A. Juels, and A. Oprea, "Hail: a high-availability and integrity layer for cloud storage," in ACM Conference on Computer and Communications Security, E. Al-Shaer, S. Jha, and A. D. Keromytis, Eds. ACM, 2009, pp. 187–198.

[10] Y. Dodis, S. P. Vadhan, and D. Wichs, "Proofs of retrievability via hardness amplification," in TCC, ser. Lecture Notes in Computer Science, O. Reingold, Ed., vol. 5444. Springer, 2009, pp. 109–127.

[11] L. Fortnow, J. Rompel, and M. Sipser, "On the power of multiprover interactive protocols," in Theoretical Computer Science, 1988, pp. 156–161.

[12] Y. Zhu, H. Hu, G.-J. Ahn, Y. Han, and S. Chen, "Collaborative integrity verification in hybrid clouds," in IEEE Conference on the 7th International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom, Orlando, Florida, USA, October 15-18, 2011, pp. 197–206.

[13] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: A berkeley view of cloud computing," EECS Department, University of California, Berkeley, Tech. Rep., Feb 2009.

[14] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in Advances in Cryptology (CRYPTO'2001), vol. 2139 of LNCS, 2001, pp. 213–229.

[15] O. Goldreich, Foundations of Cryptography: Basic Tools. Cambridge University Press, 2001.

[16] P. S. L. M. Barreto, S. D. Galbraith, C. O'Eigeartaigh, and M. Scott, "Efficient pairing computation on supersingular abelian varieties," Des. Codes Cryptography, vol. 42, no. 3, pp. 239–271, 2007.

[17] J.-L. Beuchat, N. Brisebarre, J. Detrey, and E. Okamoto, "Arithmetic operators for pairing-based cryptography," in CHES, ser. Lecture Notes in Computer Science, P. Paillier and I. Verbauwhede, Eds., vol. 4727. Springer, 2007, pp. 239–255.

[18] H. Hu, L. Hu, and D. Feng, "On a class of pseudorandom sequences from elliptic curves over finite fields," IEEE Transactions on Information Theory, vol. 53, no. 7, pp. 2598–2605, 2007

[19] Bialecki, M. Cafarella, D. Cutting, and O. O'Malley, "Hadoop: A framework for running applications on large clusters built of commodity hardware," Tech. Rep., 2005. [Online]. Available: http://lucene.apache.org/hadoop/

[20] E. Al-Shaer, S. Jha, and A. D. Keromytis, Eds., Proceedings of the2009 ACM Conference on Computer and Communications Security, CCS 2009, Chicago, Illinois, USA, November 9-13, 2009. ACM, 2009.

[21] Ayad F. Barsoum and M. Anwar Hasan,-"Integrity Verification of Multiple Data Copies over Untrusted Cloud Servers"

[22] N.Janardhan, Y.Rajasree, and R .Himaja-" Data invulnerability and data integrity verification of multi cloud storage"- International Journal of Computer Trends and Technology (IJCTT) - volume4Issue4 –April 2013

[23] Thanasis G. Papaioannou, Nicolas Bonvin and Karl Aberer-" Scalia: An Adaptive Scheme for Efficient Multi-Cloud Storage"

[24] Singh.Y.-" A secured cost-effective multi-cloud storage in cloud computing"