Computation of Data Integrity at Sector Level Using PDP Scheme in Multi-Cloud Storage

R.Ramanjulu
Department of Computer Science and Engineering,
Kuppam Engineering College, Kuppam.

S.Santha Kumari
Associate Professor,
Department of Computer Science and Engineering,
Kuppam Engineering College, Kuppam.

S.Rajan
Associate Professor,
Department of Computer Science and Engineering,
Kuppam Engineering College, Kuppam.

Abstract - 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. By using Provable Data Possession (PDP) technique, we can provide the integrity of data in Multi cloud storage. The design and implementation of an efficient PDP scheme for Multi cloud storage at sector level to support the scalability of service and distributed data, in which the multiple cloud service providers to store and manage the client's data. Also the security is used based on multi-prover zero-knowledge proof system to fulfill the soundness, completeness and zero-knowledge properties.

Keywords — Cloud Storage, Computational, Data Integration, Provable Data Possession, Sector level.

I. INTRODUCTION

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. By using Provable Data Possession (PDP) technique, we can provide the integrity of data in Multi cloud storage. The design and implementation of an efficient PDP scheme for Multi cloud storage at sector level to support the scalability of service and distributed data, in which the multiple cloud service providers to store and manage the client's data. Also the security is used based on multi-prover zero-knowledge proof system to fulfill the soundness, completeness and zero-knowledge properties.

In this paper, we use PDP scheme based on response for verification of the same type of data and hash index hierarchy. The security is provided for the cloud system using the multi-prover zero-knowledge proof system which can also satisfy completeness, knowledge soundness, and zero-knowledge properties, by providing lightweight PDP scheme based on cryptographic hash function and symmetric key encryption. The numbers of updates and challenges are limited and fixed in advance. Users cannot perform block insertions anywhere.

Recently, the cloud storage service became prominent area for research in market by providing a comparably low-cost, scalable, position-independent platform for clients' data. Both open architectures and interfaces are used for building cloud computing system; its ability to include many cloud services together for high interoperability. The virtual infrastructure management (VIM), a multi-cloud allows clients to easily access his/her resources remotely through interfaces such as Web services provided by Amazon EC2. There exist various tools and technologies for multicloud, such as Platform VM Orchestrator, VMware vSphere, and Ovirt. These tools help cloud providers construct a distributed cloud storage platform (DCSP) for managing clients' data. Provable data possession (PDP) [2] (or proofs of retrievability (POR)) is such a probabilistic proof technique for a storage provider to prove the integrity and ownership of clients' data without downloading data. To check the availability and integrity of outsourced data in cloud storages, researches have proposed two basic approaches called provable data possession and proofs for retrieve. 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 lack of randomness in the challenges.

II. ARCHITECTURE OF CLOUD STORAGE

Although existing PDP schemes offer a publicly accessible remote interface for checking and managing the tremendous amount of data, the majority of existing PDP schemes is incapable to satisfy the inherent requirements from multiple clouds in terms of communication and computation costs. To address this problem, we consider a multi-cloud storage service as illustrated in Figure 1.

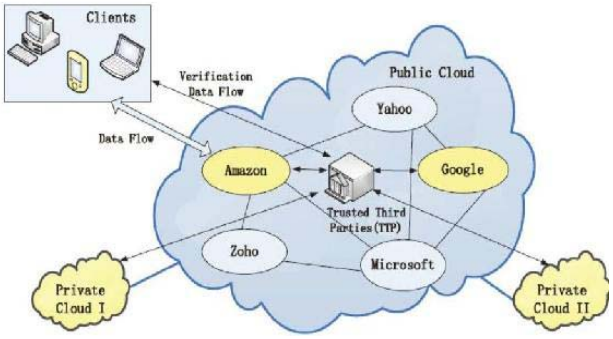

Figure 1: clouds storage system architecture.

In this architecture, a data storage service involves three different entities: Clients who have a large amount of data to be stored in multiple clouds and have the permissions to access and manipulate stored data; Cloud Service Providers (CSPs) who work together to provide data storage services and have enough storages and computation resources; and Trusted Third Party (TTP) who is trusted to store verification parameters and offer public query services for these parameters.
In this architecture, we consider the existence of multiple CSPs to cooperatively store and maintain the clients' data. Moreover, a cooperative PDP is used to verify the integrity and availability of their stored data in all CSPs. The verification procedure is described as follows: Firstly, a client (data owner) uses the secret key to pre-process a file which consists of a collection of blocks, generates a set of public verification information that is stored in TTP, transmits the file and some verification tags to CSPs, and may delete its local copy; Then, by using a verification protocol, the clients can issue a challenge for one CSP to check the integrity and availability of outsourced data with respect to public information stored in TTP. We neither assume that CSP is trust to guarantee the security of the stored data, nor assume that data owner has the ability to collect the evidence of the CSP's fault after errors has been found. To achieve this goal, a TTP server is constructed as a core trust base on the cloud for the sake of security. We assume the TTP is reliable and independent through the following functions [12]: to setup and maintain the PDP cryptosystem; to generate and store data owner's public key; and to store the public parameters used to execute the verification protocol in the PDP scheme. Note that the TTP is not directly involved in the PDP scheme in order to reduce the complexity of cryptosystem.

This application consists of the following modules.
Multi-cloud Storage
PDP scheme
Data Integrity
Third Party Auditor
Cloud User.

A. 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 users upload 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.

B. PDP scheme: PDP 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, compromising data privacy based on modern cryptographic techniques.

C. 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.

D. 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 a alert is send to the Trusted Third Party.

E. 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.

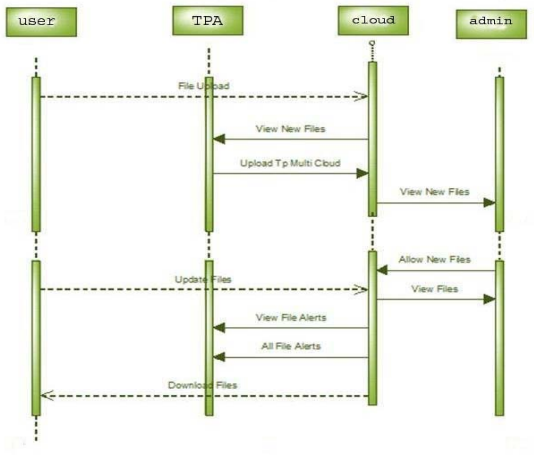

Figure 2: Working of the cloud system for computation of time with Third Party Authority.

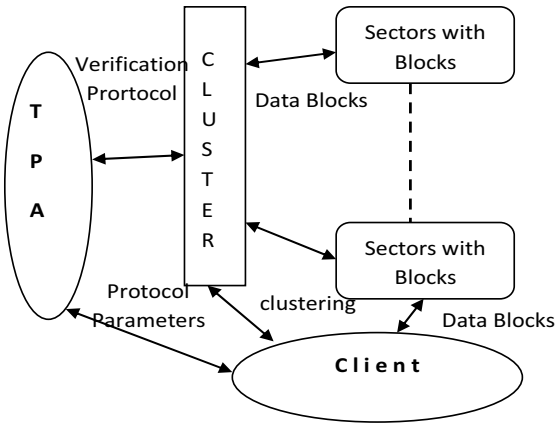

Figure 3 Structure of Sectors for the computation among cloud elements.

Based on PDP scheme, the system structure is with outsourced data in multiple clouds by replacing the TTP with a third party auditor (TPA) in Fig 3. Here the structure can be constructed into a visualization infrastructure of cloud-based storage service [1]. In Figure 3, a distributed, scalable, and portable file system [19].

The structure is composed of sectors and Blocks, where Blocks maps a file name to a set of indexes of blocks and clusters indeed stores data blocks. To support the PDP scheme, the index-hash hierarchy and the metadata of Sectors should be integrated together to provide an enquiry service for the hash value or index-hash record.

III. IMPLEMENTATION & EXPERIMENTAL RESULTS

We have implemented the computation of at sector using PDP scheme and validated the effect of dispersed secret data on clouds storage. For the sake of comparison, our experiments were executed in the following

scenario: a fixed-size file is used to generate the tags and prove data possession under the different number of sectors s. For a 256 K-Bytes file, the computational overheads of the verification protocol are shown in Fig 4(a) when the value of s is ranged from 1 to 56 and the size of sector is 32-Bytes. Moreover, there exists an optimal value of s from 16 to 32. The computational overheads of the tag generation are also shown in Fig 4(b). The results indicate that the overheads are reduced when the values of s are increased. Hence, it is necessary to select the optimal number of sectors in each block to minimize the computation costs of clients and storage service providers.

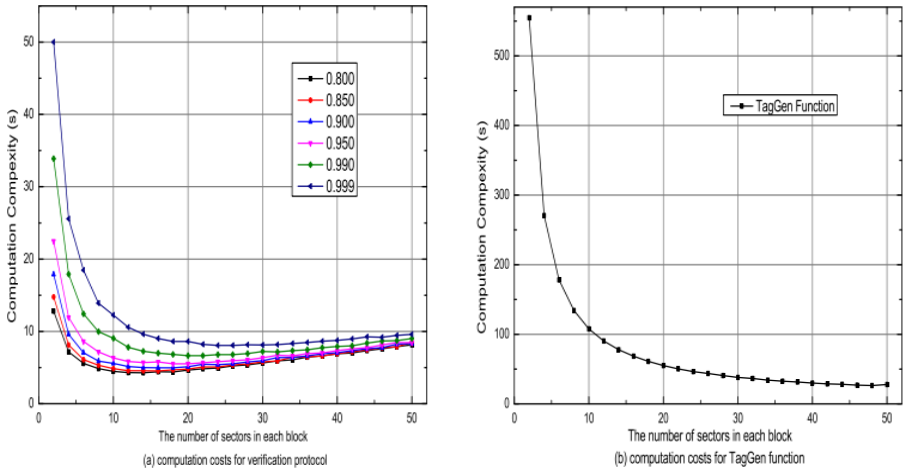

Figure 4: Experimental results of sector level computation of data process in clouds.

More accurately, we show the influence of parameters, $sz \cdot w$, s, and t, under different detection probabilities in Table 1. It is easy to see that computational cost raises with the increase of P. Moreover, we can make sure the sampling number of challenge with following conclusion: Given the detection probability P, the probability of sector corruption ρ, and the number of sectors in each block s, the sampling number of verification protocol are a constant $t = n \bullet w \geq \log(1-p) / (s \bullet \sum p_{k \in} r_k \bullet \log(1-p_k))$ for different files

TABLE I
Sectors and Time computation for the data in cloud storage

P	0.8	0.85	0.9	0.95	0.99	0.999
$sz \cdot w$	142.60	168.09	204.02	265.43	408.04	612.06
s	7	8	10	11	13	16
t	20	21	20	29	31	38

Finally, we observe the change of t under different s and P. It is obvious that the optimal value of s raises with increase of P. We choose the optimal value of s on the basis of practical settings and system requisition.

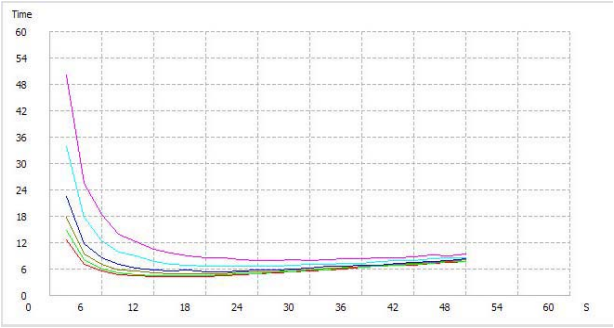

Figure 5. The computational values of s and t for the cloud stoage.

IV. CONCLUSION

We presented by using PDP scheme for sharing many cloud storage with sector size. The experimental results show computation of time with respect to the sectors size in the cloud. Security can also be provided based on response of same type and hash index hierarchy, provided all security properties required by zero knowledge interactive proof system Our experiments clearly demonstrated that our approaches only introduce a small amount of computation and communication overheads. Finally, it is still a challenging problem for the generation of tags with the length irrelevant to the size of data as sectors and blocks.

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 un trusted 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] A. 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 ComputerScience, 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] A. 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 the 2009 ACM Conference on Computer and Communications Security, CCS 2009, Chicago, Illinois, USA, November 9-13, 2009. ACM, 2009.