



## Data Security of Cooperative Provable Data in Multi-Cloud

Asst. Prof. Ingale Vinod Bhimrao, Asst. Prof. Patil Pravin. Ramchandra

Department of Computer Science and Engineering,  
Adarsh Institute of Technology and Research Centre,  
Vita, Shivaji University, Kolhapur, Maharashtra, India

**Abstract**— *The use of cloud computing has increased rapidly in many organizations. Cloud computing provides many benefits in terms of the low cost and accessibility of data. Ensuring the security of cloud computing is a major factor in the cloud computing environment, as users often store sensitive information with cloud storage providers but these providers may be interested. Dealing with “single cloud” providers is becoming less popular with customers due to risks of service availability, cloud failure and the possibility of introducers insiders in the single cloud. An idea of “multi-clouds”, or in other words, “inter clouds” or “cloud-of-clouds” has emerged newly in cloud computing.*

*This paper surveys new research related to a multi-cloud security and addresses possible solutions. It is found that the research into the use of multi-cloud providers to maintain security has received less attention from the researchers than has the use of single clouds. This work is useful for multi-clouds due to its ability to reduce security risks that affect the cloud computing user.*

*In this paper, we give Cooperative Provable data possession (CPDP) is a technique for ensuring the integrity of data in storage outsourcing. In this paper, we address the construction of an efficient PDP scheme for distributed cloud storage to support the scalability of service and data migration, which we consider the existence of multiple cloud service providers to cooperative store and maintain the clients’ data. We present a cooperative PDP (CPDP) scheme based on homomorphic verifiable response and hash index hierarchy.*

**Keywords**— *Security risk, Aspects of data security in multi-cloud, Cooperative PDP, Data Integrity, Multi cloud storage, Trusted third party, Hash Index Hierarchy for CPDP.*

### I. INTRODUCTION

The use of cloud computing has increased rapidly in many organizations that small and medium companies use cloud computing services for various reasons, including because these services provides fast access to their applications and reduce their infrastructure costs. Cloud providers should address privacy and security issues as a matter of high and urgent priority. Dealing with “single cloud” providers is becoming less popular with customers due to potential problems such as service availability failure and the possibility that there are malicious insiders in the single cloud. In recent years, there has been a move towards “multi-clouds”, “Inter-cloud” or “cloud-of-clouds”.

This paper focuses on the issues related to the data security aspect of cloud computing. As data and information will be shared with a third party, cloud computing users want to avoid an untrusted cloud provider. Protecting private and important information, such as credit card details or a patient’s medical record from attackers or malicious insiders is of critical importance. In addition, the potential for migration from a single cloud in a multi-cloud environment is examined and research related to security issues in single and multi-clouds in cloud computing are surveyed. The paper analyses the new generation of cloud computing, that is, multi -clouds and recent solutions to address the security of cloud computing, as well as examining their limitations. In the commercial world, various computing needs are provided as a service. The service providers take care of the customer's needs by, for example, maintaining software or purchasing expensive hardware. For instance, the service EC2, created by Amazon, provides customers with scalable servers. As another example, under the CLuE program, NSF joined with Google and IBM to offer academic institutions access to a large-scale distributed infrastructure [4]. Cloud service providers should ensure the security of their customers’ data and should be responsible if any security risk affects their customers’ service infrastructure. A cloud provider offers many services that can benefit its customers, such as fast access to their data from any location, scalable, pay-for-use, data storage, data recovery, protection against hackers, on-demand security controls, and use of the network and infrastructure facilities. Reliability and availability are other benefits of the public cloud, in addition to low cost. However, there are also concerns issues of public cloud computing, most notably, issues surrounding data integrity and data confidentiality. Any customer will be worried about the security of sensitive information such as medical records or financial information

### II. SECURITY RISK

From different cloud service models, the security responsibility between cloud users and cloud service providers is different. In different cloud environment addresses security control in relation to physical, environmental, and virtualization security, whereas, the users remain responsible for addressing security control of the IT system including the operating systems, applications and data.

According to Tabakiet al. [9], the way the responsibility for privacy and security in a cloud computing environment is shared between cloud users and cloud service providers differs between delivery models. In SaaS, cloud service providers are more responsible for the security and privacy of application services than the cloud users. This responsibility is more relevant to the public than the private cloud environment because the clients need stricter security requirements in the public cloud. With PaaS, users are responsible for taking care of the applications that they build and run on the platform, while cloud service providers are responsible for protecting one user's applications from others. In IaaS, users are responsible for protecting operating systems and applications, whereas cloud service providers must provide protection for the users' data [9].

Ristenpart et al. [10] claims that the levels of security issues in IaaS are different. The impact of security issues in the public cloud is greater than the impact of the private cloud. For instance, any damage which occurs to the security of the physical infrastructure or any failure in relation to the management of the security of the infrastructure will cause many problems. In the cloud environment, the physical infrastructure that is responsible for data processing and data storage can be affected by a security risk.

### III. ASPECTS OF DATA SECURITY IN MULTI-CLOUD

**Confidentiality:** confidential is term in which cloud service provider also unknown to cloud users data which is uploaded on his own cloud, the cloud storage provider does not learn any information about customer data.

**Integrity:** any unauthorized or illegal modification and updating the contents of client data from the cloud storage provider can be detected by the customer while retaining the main benefits of a public storage service:

**Availability:** data of cloud user are available to the user at anytime, anywhere, anyplace from the cloud server. Customer data is accessible from any machine and at all-time reliability: customer data is reliably backed up

**Efficient retrieval:** data retrieval times are comparable to a public cloud storage service

**Data sharing:** customers can share their data with trusted parties. **Data sharing:** cloud users can share data securely with trusted parties.

### IV. STRUCTURE AND TECHNIQUES

This paper presents verification framework for multi-cloud storage and a formal definition of Cooperative Provable Data Possession. We introduce two fundamental techniques for constructing our CPDP scheme: hash index hierarchy (HIH) on which the responses of the clients' challenges computed from multiple Cloud Service Providers can be combined into a single response as the final result; and homomorphism verifiable response (HVR) which supports distributed cloud storage in a multi-cloud storage and implements an efficient construction of collision resistant hash function, which can be viewed as a random oracle model in the verification protocol.

Verification Framework for Multi-Cloud and availability of outsourced data with respect to public information stored in Trusted Third party 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 are 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. 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.

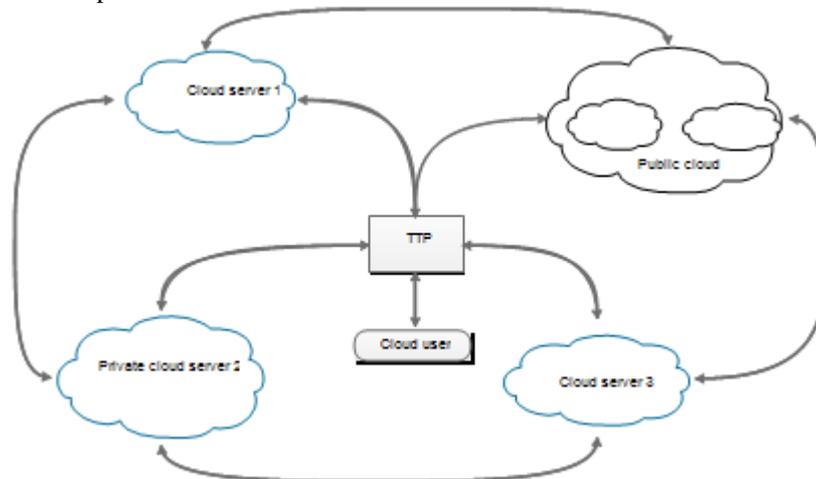


Fig. 1. Verification architecture for data integrity and data security.

In this architecture, we consider the existence of multiple Cloud Service Providers to cooperatively store and maintain the cloud user's data. Moreover, a cooperative PDP is used to verify the integrity and availability of clients stored data in all Cloud service Providers. The verification procedure is described as follows: Firstly, a

cloud user (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 Cloud Service Providers, 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.

## **V. COOPERATIVE PDP**

Cooperative Provable Data Possession (CPDP) schemes adopting zero-knowledge property and three-layered index hierarchy, that three layers are express layer, service layer and storage layer 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.

## **VI. 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. One of the most important issues related to cloud security risks is data integrity. The data stored in the cloud may suffer from damage during transition operations from different cloud users or due to introducers attack on cloud storage provider. when multiple clients use cloud storage or when multiple devices are synchronized by one user, it is difficult to address the data security and integrity. Data Integrity is very important in database operations in particular and Data warehousing and We neither assume that Cloud Service Provider 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 have been found. To achieve this goal, a Trusted Third Party 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 to set up and maintain the CPDP cryptosystem; to generate and store data owners public key and to store the public parameters used to execute the verification protocol in the CPDP scheme. TTP is not directly involved in the CPDP scheme in order to reduce the complexity of a cryptosystem.

## **VII. MULTI CLOUD STORAGE**

The term "multi-clouds" is similar to the "inter clouds" or "cloud-of-clouds". These terms suggest that cloud computing should not finish with only a single cloud. Using their illustration, a cloudy sky incorporates different colors and shapes of clouds which leads to different implementations and administrative domains. Distributed computing is used 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 service provider consist of data blocks .The cloud user 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 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.

## **VIII. TRUSTED THIRD PARTY**

Trusted Third Party (TTP) is the reliable and trusted one which is responsible for uploading cloud users' data on different clouds with secure verifying parameters .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 different distributed cloud. In distributed cloud environment each cloud has user data blocks. If any modification or updating tried by cloud owner or any introducer a alert is sent to the Trusted Third Party.

## **IX. 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 is 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.

## **X. HASH INDEX HIERARCHY FOR CPDP**

To support distributed cloud storage, we illustrate a representative architecture used in our cooperative PDP scheme as shown in Figure 2. Our architecture has a hierarchy structure which resembles a natural representation of file storage. This hierarchical structure consists of three layers to represent relationships among all blocks for storing resources. They are described as follows:

- 1) **Express Layer:** offers an abstract representation of the stored resources;
- 2) **Service Layer:** offers and manages cloud storage services; and
- 3) **Storage Layer:** realizes data storage on many physical devices.

We make use of this simple hierarchy to organize data blocks from multiple CSP services into a large size file by shading their differences among these cloud storage systems. For example, in Figure 2 the resources in Express Layer are split and stored into three CSPs, that are indicated by different layers, in Service

Layer. In turn, each CSP fragment and stores the assigned data into the storage servers in Storage Layer. We also make use of colors to distinguish different CSPs. Moreover, we follow the logical order of the data blocks to organize the Storage Layer.

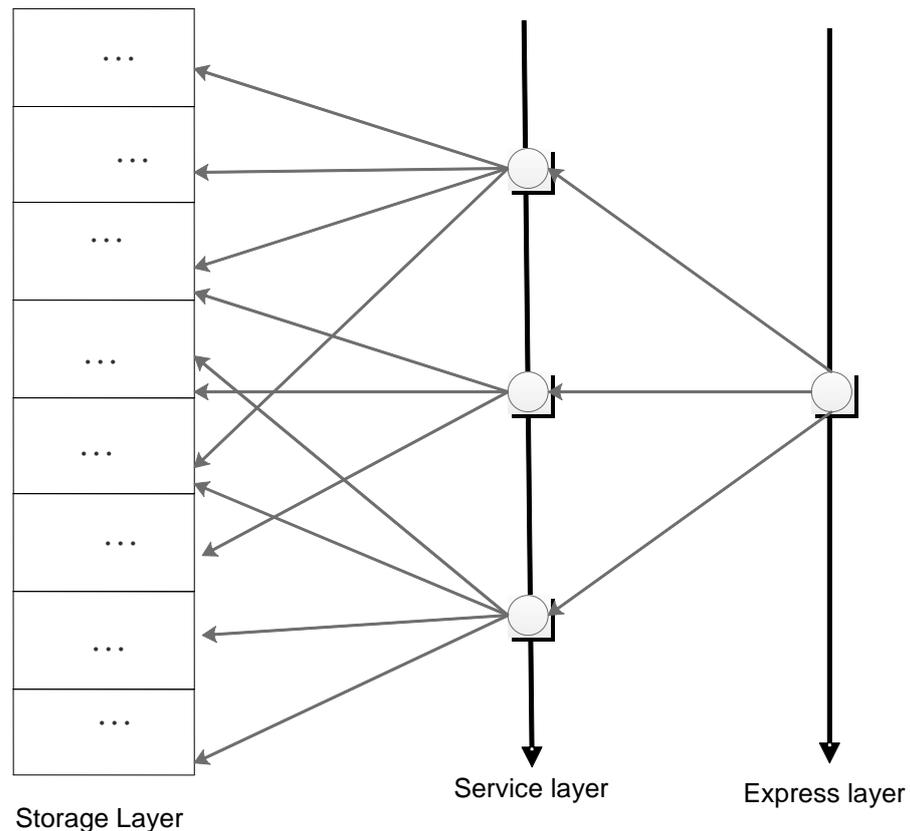


Fig. 2. CPDP applying three different layers

**XI. HADOOP DISTRIBUTED FILE SYSTEM (HDFS)**

This architecture also provides special functions for data storage and management there may exist overlaps among data blocks and discontinuous blocks, but these function may increase the complexity of storage management. Hadoop distributed file system is designed to run on commodity hardware with additional significant features compared to other DFS. HDFS works by using a master and slave relationship architecture. An HDFS cluster consists of a master server, Name Node, that manages the file system's namespace and controls the client's access to files. Each node in the cluster consists of slaves, Data nodes, which manages the storage attached to the nodes they run on. HDFS provides the file system namespace to allow user data to be stored in files. The HDFS architecture diagram is shown below for Name nodes and Data nodes. Having a Name Node in the cluster simplifies the architecture of the system by acting as the arbitrator and repository for all HDFS metadata. HDFS can reliably store large files across machines in a large cluster as a sequence of blocks.

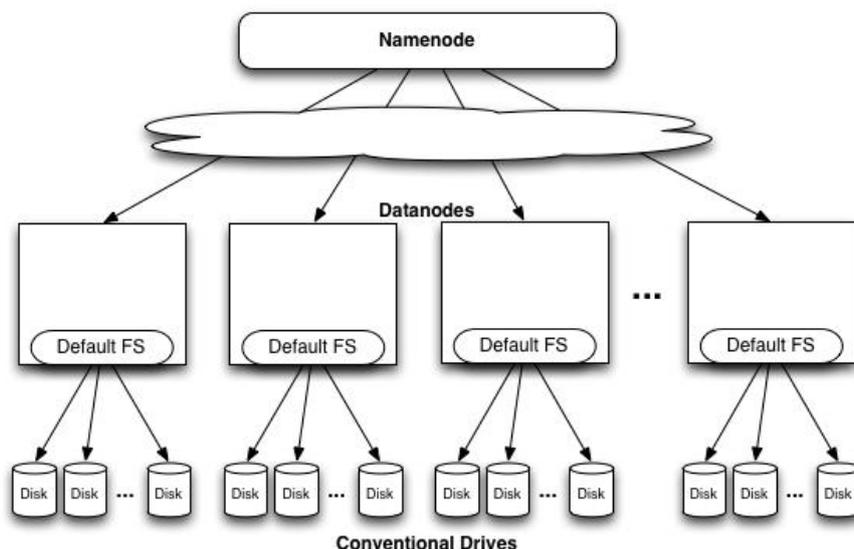


Fig. 3. Applying CPDP scheme in the Hadoop distributed file system (HDFS).

## XII. CONCLUSION

Firstly, we quantify the performance of our audit scheme under different parameters, such as file size, sampling ratio, sector number per block, and so on. Our analysis shows that the value should grow with the increase in  $z$  in order to reduce computation and communication costs. Thus, our experiments were carried out as follows: the stored files were chosen from 10KB to 10MB; the sector numbers were changed from 20 to 250 in terms of file sizes; and the sampling ratios were changed from 10% to 50%. The experimental results are shown in the left side of Figure 4. These results dictate that the computation and communication costs (including I/O costs) grow with the increase of file size and sampling ratio. Next, we compare the performance of each activity in our verification protocol. We have shown the theoretical results in Table 4: the overheads of “commitment” and “challenge” resemble one another, and the overheads of “response” and “verification” resemble one another as well.

In this paper, we presented the construction of an efficient PDP scheme for distributed cloud storage. Based on the homomorphic verifiable response and hash index hierarchy, we have proposed a cooperative PDP scheme to support dynamic scalability on multiple storage servers. We also showed that our scheme provided all security properties required by zero knowledge interactive proof system, so that it can resist various attacks, even if it is deployed as a public audit service in clouds. Furthermore, we optimized the probabilistic query and periodic verification to improve the audit performance. Our experiments clearly demonstrated that our approaches only introduce a small amount of computation and communication overheads. Therefore, our solution can be treated as a new candidate for data integrity and data security in outsourcing data storage system of CPDP scheme, especially 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 [2]. Next, from a practical point of view, we still need to address some issues about integrating our CPDP scheme smoothly with existing systems, for example, how to match the index hash hierarchy with HDFS’s two-layer namespace, how to match the index structure with cluster-network model, and how to dynamically update the CPDP parameters according to HDFS specific requirements. 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 an issue to provide the support of variable-length block verification.

## REFERENCES

- [1] A Trigger Identification Service for Defending Reactive Jammers in WSN IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 11, NO. 5, MAY 2012
- [2] M. Goodrich, M. Atallah, and R. Tamassia, “Indexing Information for Data Forensics,” Proc. Third Applied Cryptography and Network Security Conf. (ACNS), 2005.
- [3] Yan Zhu, Hongxin Hu, Gail-Joon Ahn, *Senior Member, IEEE*, Mengyang Yu” Cooperative Provable Data Possession for Integrity Verification in Multi-Cloud Storage”, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEM.
- [4] V. Guruswami and C.P. Rangan, “Algorithmic Aspects of Clique-Transversal and Clique-Independent Sets,” Discrete Applied Math., vol. 100, pp. 183-202, 2000.
- [5] W. Hang, W. Zanji, and G. Jingbo, “Performance of DSSS Against Repeater Jamming,” Proc. IEEE 13th Int’l Conf. Electronics, Circuits and Systems (ICECS), 2006.
- [6] XUAN ET AL.: A TRIGGER IDENTIFICATION SERVICE FOR DEFENDING REACTIVE JAMMERS IN WSN 805 Fig. 12. Solution robustness.
- [7] P. Tague, S. Nabar, J.A. Ritcey, and R. Poovendran, “Jamming-Aware Traffic Allocation for Multiple-Path Routing Using Portfolio Selection,” IEEE/ACM Trans. Networking, vol. 19, no. 1, pp. 184-194, Feb. 2011.
- [8] I. Shin, Y. Shen, Y. Xuan, M.T. Thai, and T. Znati, “Reactive Jamming Attacks in Multi-Radio Wireless Sensor Networks: An Efficient Mitigating Measure by Identifying Trigger Nodes,” Proc. Second ACM Int’l Workshop Foundations of Wireless Ad Hoc and Sensor Networking and Computing (FOWANC), in conjunction
- [9] R.C. Merkle, "Protocols for public key cryptosystems", IEEE Symposium on Security and Privacy, 1980, pp. 122-134.
- [10] H. Tabakiet, J.B.D. Joshi and G.-J. Ahn, "Security and Privacy Challenges in Cloud Computing Environments", IEEE Security & Privacy, 8(6), 2010, pp. 24-31.
- [11] T. Ristenpart, E. Tromer, H. Shacham and S. Savage, "Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds", CCS'09: Proc. 16th ACM Conf. on Computer and communications security, 2009, pp. 199-212.