International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 3, Issue 6, June 2016

Single and Multi Clouds Techniques for Provable Data Possession- A Comparative Survey

[1] Pradeep B Gaonkar, [2] Dr. J. A. Laxminarayana
Computer Engineering Department,
Goa College of Engineering,
Farmagudi Ponda Goa

Abstract - Due to increase in the growth of computational, maintenance and storage cost at user environment, an organisation prefer to host huge user data on the cloud server where users can access their data without any storage constraints. Data storage on a cloud introduces a security issue concerned with the safety of user's data. To check the data integrity of cloud data, a technique called Provable Data Possession (PDP) is introduced. This paper presents a survey of techniques used to establish the cloud data integrity.

***Keywords*— Storage Security, Provable Data Possession, Multiple Cloud**

I. INTRODUCTION

Storing a user data in cloud environment makes an organisation more profitable as the cost of computation and maintenance of storage are low. Further, cloud storage data is scalable and offers location-independent platform for managing clients' data. The cloud infrastructures are much more reliable and powerful compared to personal computing devices. However, security is critical in clouds environment due to various issues. In a cloud environment, there may be internal and external threats. The critical issue is the data integrity when a user updates his/her data on cloud server. In such scenario, it is essential to ensure the integrity of data. To ensure the integrity, a client may need to access whole data set from the cloud, which is usually consuming huge amount of space, bandwidth and time. Considering the huge data size and user's constraint to the resources the provable data possession technique is proposed

The Provable Data Possession (PDP) is a technique used to ensure data intactness of the user's data hosted on cloud server. In this technique, at client side a metadata is computed from the actual hosted data on cloud server to ensure integrity. A client can use the metadata to check data integrity. The cloud server stores actual data along with the appropriate metadata generated by the client. Whenever the user asks for verification, server returns the metadata as a response which is then verified by the user by comparing with backup copy of metadata stored at client side.

As PDP techniques are widely used in data integrity verification in clouds storage, various system and security models are proposed. It is difficult ascertain whether the data integrity is maintained when user accesses data at remote server. There may be a scenario that the Cloud Server Provider may delete a part of the user data and falsely claim that data is still intact in the cloud. The critical issue with the outsourced data is data integrity. Many PDP schemes are proposed under various system and security model to overcome data integrity issues.

II. CURRENT SCENARIO

At the outset, we can categorise the work done by the researchers as single cloud based or multi cloud based techniques. Some of the techniques use static PDP schemes whereas some employs dynamics PDP schemes. However, there are techniques using hybrid PDP schemes also.

The characteristics of the proposed techniques for PDP are critically analysed and empirically verified. The result of this study is presented in the following sections.

A. Single Cloud Based Techniques

We have considered five different works proposed by the researchers. All the methods are considering the existence of a single cloud.

i. Provable Data Possession (PDP)

The PDP is a technique used to ensure data intactness of the user's data hosted on cloud server [1]. In this technique, at client side a metadata is computed from the actual data in order to ensure integrity of hosted data on

cloud server. Client can use the metadata to check data integrity. The cloud server store actual data along with the appropriate metadata generated by client. Whenever the user asks for verification, server returns the metadata as a response which is then verified by the user by comparing with backup copy of metadata stored at his/her side.

ii. Dynamic Provable Data Possession (DPDP)

The Data dynamics plays an important role in data integrity checking techniques. The PDP scheme suits to only static outsourced data files and is not used for dynamic data files. To resolve this type of issue, Chris Erway, [2] et al. provided Dynamic Provable Data Possession (DPDP) schemes to allow data dynamics in outsourced data. DPDP is an extended version of PDP model to support dynamics update of store data based on rank information.

Bo Chen et al [3] proposed two different approaches of DPDP and a construction of rank-based authenticated dictionary using a skip list. The maintenance of log and communication in the DPDP schemes are same as the original PDP scheme. The next important reported work is an alternative construction of a rank-based authenticated dictionary using an RSA tree [2]. This construction results in a DPDP scheme with improved detection probability but increases server computation

iii. Designated-Verifier Provable Data Possession in Public Cloud Storage

In public cloud, data integrity is a crucial issue when the client cannot perform the outsource data possession checking. In the normal PDP approach it increases clients overhead to calculate tags and hash values for the data to check data integrity. Yongjun Ren, et al. [4], proposed the designated data verification model for the clients with less resources and computational power. The authors have proposed to use Elliptic Curve Cryptography (ECC)-based homomorphism authenticator to design PDP scheme, which consume small amounts of calculation and fewer communications. This type of scheme is best suited for mobile clouds.

In terms of complexities, an elliptic curves cryptography (ECC)[9] provides shorter key length compared to RSA based on the same level of security. It has been shown that 160-bit ECC provides comparable security to 1024-bit RSA. The communication overhead caused mostly comes from the DV-PDP response.

iv. Identity based Remote Data Possession Checking

The existing PDP protocols have been designed in the PKI (public key Infrastructure) setting. In PDP approach, the cloud server has to authenticate the users' certificates before storing the data uploaded by the users to the cloud server in order to prevent spam. This incurs considerable costs as many users may frequently upload data to the cloud server. Huaqun Wang {5}, addressed this problem with a new model of identity-based RDPC (ID-RDPC) protocols. The author provided first ID based PDP protocol which is secure by assuming the hardness of the standard computational Diffie-Hellman (CDH) problem. Further, in addition to the structural advantage of elimination of certificate management [8] and verification, the ID-RDPC protocol also outperforms existing PDP protocols in the PKI setting in terms of computation and communication. Firstly, the PKG (Private Key Generator) generates the system public key and the master secret key along with the private keys for the clients of an organization [5]. The main challenge to design of the ID-RDPC protocol is that it requires to generate aggregately ID-based signatures for the client like tags for blocks without applying the hash and sign paradigm to the original data. The authors addressed this issue with a variation of the well-known Schnorr signature [8].

v. Analysis

The method outlined above are compared with respect to their advantages and disadvantages as per table 1.1

Table 1.1 Comparison of single cloud methods

Method 1: **Provable Data Possession (PDP)**	
Advantages	1. Protection against small corruptions. 2. Reduced update block communication 3. RSA scheme for security. 4. Allows public verifiability
Disadvantages	1. Searching the block is poor (with brute force) 2. It is more efficient scheme but can applicable only for static files. 3. It is insecure against dynamic block of data.
Method 2: **Dynamic Provable Data Possession (DPDP)**	
Advantages	1. Block modification and updating of block is allowed. 2. Efficient integrity verification is made by querying and updating DPDP scenario
Disadvantages	1. Client needs to perform extra computations. 2. It provides efficient verification but construction of rank based scheme is complex
Method 3: **Designated-Verifier Provable Data Possession in Public Cloud Storage**	
Advantages	1. No client expertise is required.

	2. Elliptic curves cryptography (ECC) has shorter key length based on the same level of security. 3. The total communication overhead is more efficient.
Disadvantages	1. Extra setup is needed for designated verifier. 2. Pairing based approach increases complexity
Method 4: **Identity based Remote Data Possession Checking (ID-RDPC)**	
Advantages	1. Reduced Communication Overhead 2. It allows Data dynamics.
Disadvantages	1. The approach increases. 2. Can't adopt for multi-cloud storage.

B. Multi Cloud Based Techniques

We have considered three different works proposed by the researchers. All the method are considering the existence of multi cloud

i. Cooperative Provable Data Possession

The parallel computing can be implemented in several ways of computing like instruction, task and data level parallelism. In case of large volume of data PDP technique is not suitable for verification has it perform slower. In such situations, the data integrity verification can be done in parallel and data block storages can be on multiple clouds. The YanZhu, et al. [6] proposed Cooperative PDP (CPDP) model, which is based on zero knowledge proof mechanism and interactive proof system to prove the integrity of data stored in a multi cloud. A Cooperative PDP is a collection of two main algorithms (Key Gen, Tag Gen) and interactive proof system Proof [6].

The Key Generation Algorithm:- Input has been taken as security parameter and secret key has been return.

Tag Generation Algorithm: Secret key, set of cloud storage and file as input and returns triplet. A protocol called **GenProof** is used to generate a proof of data possession among the CSP's and data verifier.

The CPDP approach allows parallel computing. It also enhances performance and also provides support for large file storage on cloud.

ii. Identity Based Distributed PDP

In some scenarios, the clients have to store their data on multi-cloud servers to allow parallelism and huge data storage. So, the integrity checking protocol must be efficient to save the verifier's cost. Wang, H [7], proposed a novel PDP model as ID-DPDP (identity-based distributed provable data possession) in multi-cloud storage. Based on the bilinear pairing concept, the complete IDDPDP protocol is designed. The proposed ID-DPDP protocol is provably secure under the hardness assumption of the standard computational Diffie- Hellman (CDH) problem as tested by author. In addition to eliminate the managing certificate; the IDDPDP approach is efficient and flexible. Depending on the client's authorization, the proposed ID-DPDP protocol can identify private verification and public verification.

iii. Robust DPDP

A robust DPDP scheme implements mechanisms to mitigate arbitrary amounts of data corruption. The protection against small corruptions to ensures that attacks that modify a few bits do not destroy an encrypted file or invalidate authentication information. As updating a small portion of the file may require retrieving the entire file, the PDP scheme must be robust enough to perform dynamic updates. Bo Chen, et al. [3] proposed two approaches towards Robust DPDP. The first approach provides efficient encoding, but causes high communication cost for updates whereas the second approach overcomes this drawback through a combination of techniques that consists of RS codes based on Cauchy matrices, separating the encoding for robustness from the symbol position in the file, and reducing add/remove operations to append/modify operations when updating the RS-encoded parity data. Robustness is a vital property for all PDP schemes that rely on spot checking, which includes the both types of static and dynamic PDP protocols.

iv. Analysis

The method outlined above are compared based on their advantages and disadvantages as per table 1.2

Table 1.2 Comparison of Mutli cloud methods

Method1: **Cooperative Provable Data Possession**	
Advantages	1. Allows multi cloud storage. 2. Hash index hierarchy reduces search complexity
Disadvantages	1. Due to multi cloud storage, Combiner model needs to be added which may increase complexity
Method 2: **Identity Based Distributed PDP**	
Advantages	1. Allows multi-cloud storage. 2. Provides more flexibility as a block is divided in multiple parts.

Disadvantages	1. Incurred overhead due to combiner and PKG modules.
Method 3: **Robust DPDP**	
Advantages	1. Lightweight as it reduces overheads and communication 2. Spot checking allows clients to randomly check data integrity
Disadvantages	1. Provides high communication overhead in first model of RDPDP

III. FURTHER ENHANCEMENT

PDP-based schemes of Ateniese et al [1] consider public auditability in the model and then utilizing RSA-based homomorphic tags to achieve public auditability However, this scheme does not support dynamic updates. Further, Ateniese et al [1], proposed a scalable PDP scheme (SPDP) based on symmetric-key cryptography by extending PDP to partially support dynamic data operations. Erway et al [2], proposed two fully dynamic PDP schemes, DPDP I and DPDP II, by using rank-based authentication skip list structure and rank based RSA trees, respectively [2]. Esiner et al. proposed a flexible dynamic PDP scheme (F-DPDP) [8] by using FlexList (Flex Length-based Authenticated Skip List method) in order to overcome disadvantages of DPDP I and DPDP II. Wang et al [7] proposed a public PDP scheme (PPDP) by using Mekle Hash Tree (MHT) to authenticate block tags, which supports both dynamic updates and public auditability. Zhu et al [7] also proposed dynamic PDP scheme (i.e D-PPDP) which depends on random sampling, index-hash table and fragment structure. Huaqun Wang addressed this problem with a new model of identity-based RDPC (ID-RDPC) protocols [7]. They provided first ID based PDP protocol to be secure assuming that the hardness of the standard computational Diffie-Hellman (CDH) problem

IV. CONCLUSION

In Public cloud computing environment, the data integrity verification is crucial part. There are many PDP techniques which are available and further improved to achieve efficient integrity verification. We have identified latest PDP variations and compared those PDP schemes based on their approaches, techniques, advantages and disadvantages. As a result, we have proposed the enhanced Identity based PDP scheme for data integrity verification which will make client free from the data intactness checking and also will provide a scheme to perform administrative task

REFERENCES

[1]. G. Ateniese, R.D. Pietro, L.V. Mancini, Scalable and Efficient Provable Data Possession, Proc. Fourth Int'l Conf.Security and Privacy in Comm. Networks (SecureComm '08), pp. 1-10.

[2] C. Erway, A. Kupclu, C. Papamanthou, and R. Tamassia. Dynamic provable data possession. In Proc. of the 16th ACM Conference on Computer and Communications Security (CCS'09), Chicago, Illinois, USA, pages 213–222. ACM, 2009.

[3]. Bo Chen Reza Curtmola, Robust Dynamic Provable Data Possession. In IEEE Transactions on Distributed Computing Systems Workshop, 2012 32nd International Conference. Pp. 515- 525.

[4]. Yongjun Ren, Jiang Xu, Jin Wang, and Jeong-Uk Kim Designated Verifier Provable Data Possession in Public Cloud Storage. In International Journal of Security and Its Applications Vol.7, No.6 (2013), pp.11-20.

[5]. Huaqun Wang, Qianhong Wu, Bo Qin, and Josep Domingo-Ferrer, Identity-Based Remote Data Possession Checking in Public Clouds. In Information Security, IET (Volume:8 , Issue: 2), p.p.: 114-121.

[6]. Yan Zhu, Hongxin Hu,Gail-Joon Ahn and Mengyang Yu., Cooperative Provable Data Possession for Integrity Verification in Multicloud Storage. IEEE Transactions on Parallel and Distributed Systems, 23, 12(2012)

[7] Wang, H., Identity-Based Distributed Provable Data Possession in Multi-Cloud Storage. At Services Computing, IEEE Transactions on (Vol:PP , Issue: 99).

[8] E. Esiner, A. Kachkeev, and O. Ozkasap. FlexList: Optimized skip list for secure cloud storage. Technical report, Ko University, 2013.

[9] A. Miyaji, M. Nakabayashi, S. Takano. New Explicit Conditions of Elliptic Curve Traces for FR-reduction. IEICE Transactions Fundamentals, 5:1234-1243, 2001.