Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 03, March-2016

A Survey on Provable Data Possession in Cloud Computing Systems

Poornashree B. R.
Dept of ISE,
BNM Institute of Technology,
Bangalore, India.

S.Srividhya
Assistant Professor, Dept of ISE,
BNM Institute of Technology,
Bangalore, India.

***Abstract*— In Provable data possession scheme the customer outsources the data to the remote cloud service provider which is responsible for storing and preserving the data. Customers can rent the storage infrastructure from the cloud service providers to store their data by paying fees. Therefore the customers need to verify whether the server possesses the original data and should have strong guarantee that the service provider is storing all the data copies issued as per the agreement. In this process the issues such as data security, data dynamics, integrity protection and multi cloud storage have remained the most important task. To achieve this various PDP techniques and its extensions are discussed in this paper. This paper surveyed different types of PDP techniques and the focus is done on comparing the best method for achieving the efficient and secure PDP technique.**

Keywords—Data dynamics; Data Security; Confidentiality; Integrity Protection.

I. INTRODUCTION

Organizations outsource their data to remote cloud service providers to store their data since the cloud service providers provide large amount of storage infrastructure which relieves the burden on the organizations to maintain storage infrastructure, constantly updating the server and other computing issues [1]. Such outsourcing to cloud also provides security to the data stored in the cloud rather than storing it in the private computer systems [2]. Many authorized users from the organizations can remotely access the data stored in the cloud across different geographic locations. Customers can lose direct control over their data by outsourcing their sensitive data to the CSP which may not be trustworthy. Data owners lose control over their sensitive data which raises confidentiality and integrity issues [3]. Before storing data to the cloud the data copies can be encrypted and then stored to the remote CSP which provides security against attacks [4]. The remote CSP guarantees about the authenticity of the data copies which is stored but it is insufficient to trust the CSP because apart from intentional dishonest like tampering and deleting partial data the server might be exposed to data loss because of administration errors such as backup and restore, migration of data to new systems or it may be vulnerable to latent faults, correlated faults and recovery faults [5]. Thus to solve integrity issue enough evidence has to be provided to the customer that all their data copies are stored across all the servers with the most recent modifications which is given by the customer. To verify that the server possesses the original data copies the entire file cannot be accessed because of expensive I/O costs and transferring the files incur high network costs hence the verification is carried out without retrieving the file. The client is able to verify that the server has retained original file data without retrieving the data from the server and without having the server to access the entire file [6]. In the PDP scheme the data owner generates a metadata for all the files stored on the CSP which is used later for verification purposes using a challenge-response protocol [7].

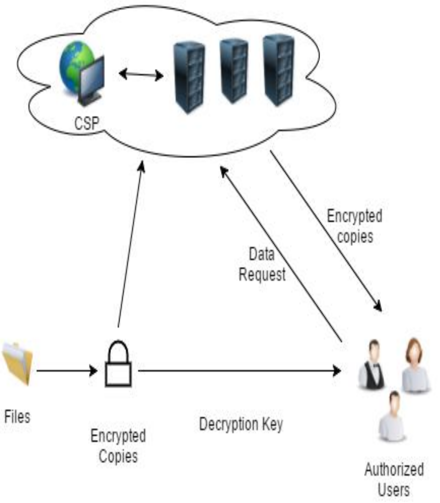

Fig. 1. Cloud Computing Data Storage System Model

II. PROVABLE DATA POSSESSION SCHEMES

In this paper various provable data possession techniques are discussed based on their efficiency.

A. Provable Data Possession (PDP)

Ateniese [5] has discussed a provable data possession technique where a PDP protocol checks whether the data outsourced to the CSP is retained as per the service agreement. The client pre-processes the file, generating a metadata which is stored locally and transmits the file to the CSP and he may delete his local copy of file. The server stores the file and responds to the challenge issued by the client. The client can alter the data in the file which is to be stored in the server. The client can execute data possession challenge to make sure that the server has retained the file before deleting his local copy of file. Before outsourcing the data to the CSP the client can encrypt the file for the security purpose but metadata does not contain any encryption keys. Whenever the client needs to verify the integrity of the file data possession challenge is issued for which the server has to compute response, using the metadata which is stored locally the client can verify whether the server has successfully retained the file. The server has to respond to the challenges issued by the client failing to do so indicate that there may be a data loss and the server could not be trusted. Even though the file is partially or totally missing the server might try to convince the client that it possess the original data. The intention of this scheme is to detect the misbehaviour of the server.

The drawback of this scheme is it only applies to the static files. For improving this disadvantage Dynamic provable data possession was proposed however this is the first technique to propose provable data possession.

B. Dynamic Provable Data Possession(DPDP)

C. Chris Erway and Alptekin Kupcu [8] have proposed an efficient way of proving the integrity of data stored in the CSP. In the PDP model the client pre-processes the data and then stores it in the server by keeping the metadata and the server responds to the challenge issued by the client. However this model applies only to static files [9]. Hence in the DPDP model the PDP model is extended to support dynamic updates to the stored data. In the PDP model the file that is outsourced can never be changed whereas in the DPDP model dynamism is supported where the client can insert, modify or delete the stored blocks. Such scheme is essential in practical scenario [10].

In this DPDP scheme an efficient construction for dynamic provable data possession is proposed which extends the PDP model to support provable updates on the stored data. Given a file F consisting of n blocks, update is defined as either insertion of a new block, or modification of an existing block, or deletion of any block. Therefore update operation is the most general form of modifications a client may wish to perform on a file. In this scheme the rank information is used to organize dictionary entries. Thus it is able to support efficient authenticated operations.

This scheme provides an efficient fully dynamic PDP solution. But the scheme does not guarantee that multiple copies of the data file are actually maintained [11].

C. Multiple Replica Provable Data Possession(MR-PDP)

Reza Curtmola [12] has proposed multiple-replica provable data possession (MR-PDP) system. In order to improve the data availability and reliability of a single replication PDP system the data copies are replicated and stored across multiple servers. By storing the data files on multiple servers across different locations, even though if some of the copies are destroyed, the data can still be recovered from the remaining copies. The replication systems can tolerate failures only if the failure modes of the replicas are independent. Suppose if the failure mode of replicas is dependent then all the replicas may fail simultaneously this is because all the replicas are stored in the same geographical location or because data dependencies exist among replicas. The main aim of the replication systems is to tolerate independent, accidental and non-malicious failures such as hardware failures. When the storage servers are non-malicious, storing data in different geographic locations can ensure failure independence. The situation is different when the servers are untrusted, i.e., servers are malicious and can collude. The failure independence cannot be assumed in the replication systems which rely on untrusted servers, such servers cannot offer the same level of assurance as a system relying on trusted servers. Initially the replicas might be stored on servers in different geographic locations, but later the servers can move all the replicas to one location and access them from that location when client demands. Another important open problem is establishing physical location of data. The generic limitation faced by the replication systems is to prove the data availability; upon client’s challenge, the servers can produce replicas however this does not prove that the actual replicas are stored all the times. The malicious servers may introduce dependencies among replicas stored across different geographic locations, by encrypting them before storing. Replicas can be decrypted and served whenever they are requested by clients. The malicious servers can effectively decrease the reliability improvements achieved by storing the replicas at different locations by storing the encryption key in a single location. Loss of the encryption key means loss of all the replicas.

The efficient multiple-replica provable data possession (MR-PDP) scheme is discussed that guarantees that the storage servers are storing multiple unique replicas. However the drawback of the scheme is authorized users face problem in accessing the file copies from the CSP.

D. Efficient Multicopy Provable Data Possession(EMC_PDP)

Ayad F.Barsoum and M.Anwar Hasan [13] have proposed secure and efficient protocol to provide strong evidence to the customers that CSP is storing all the data copies as per the service agreement. The Efficient Multi-Copy Provable Data Possession (EMC-PDP) scheme is proposed which utilizes BLS Homomorphic Linear Authenticators (HLAs) [14]. The HLAs finger prints every block of file in such a way that it satisfies any challenge vector issued by the customer, by authenticating value the server can homomorphically construct the tag. The main task in designing a multi-copy provable data possession model is to generate unique distinguishable copies of data file, a simple and efficient method is used to generate these copies. The EMC-PDP model adopts to the diffusion property of any secure encryption scheme. Diffusion means that the output bits of the ciphertext should depend on the input bits of the plaintext in a very complex way. In an encryption scheme with strong diffusion property, if there is a change in one single bit of the plaintext, then there will be drastic change in the cipher text in an unpredictable way [15]. This methodology of generating distinct copies is efficient, and also successful in solving the authorized users problem of the MRPDP scheme to access the file copy received from the CSP. In this scheme, the data owner or the authorized users need only to keep a single secret shared key to decrypt the file copy. This is a secure, complete, and efficient protocol that addresses the storage integrity of multiple data copies over cloud computing.

E. Map Based Provable Multicopy Dynamic Data Possession(MB-PMDDP)

Ayad F.Barsoum and M.Anwar Hasan [16] have proposed a map-based provable multicopy dynamic data possession (MB-PMDDP) technique which provides evidence to the customer that CSP is not cheating by storing only a fewer copies. This scheme also supports dynamic behaviour of data [17]. When large numbers of verifiers are connected to the CSP the computation overhead increases on servers, the MB-PMDDP scheme significantly reduces the computation time in the challenge-response phase which makes it more practical for applications. Besides, it also reduces the storage overhead on the CSP, and thus reduces the fees paid by the cloud customers. The communication cost incurred for dynamic block operations of the map-based approach is less. The map-based PDP scheme validates the integrity and consistency of all file copies outsourced to the CSP by using a map-version table (MVT), it is a small dynamic data structure which is stored on the verifier side. The MVT consists of three columns: serial number (SN), blocks number (BN), and block version (BV). The file blocks are indexed using the serial number. The serial number indicates the physical position of a block in a data file. The block number is a counter which is used to number the file blocks. Thus, the relation between block number and serial number can be viewed as a mapping between the logical number block number and the physical position serial number. The block version indicates the current version of file blocks. When a data file is initially created the block version of each block is 1. If a specific block is being updated, its block version is incremented by 1. The verifier keeps only one table for unlimited number of file copies, i.e., the storage requirement on the verifier side does not depend on the number of file copies on cloud servers. To simplify the insertion and deletion of entries to the table the MVT is implemented as a linked list. For actual implementation, the serial number is not needed to be stored in the table; serial number is considered to be the entry/table index, i.e., each table entry contains just two integers block number and block version.

CONCLUSION

The demand for outsourcing the data to the cloud is tremendously increasing. So the need for efficient and secure PDP technique is also abundantly increasing. To overcome those aspects the desired efficiency and security goals must be achieved. In this paper, the survey of different PDP scheme is mentioned with their advantage and disadvantage. The different variation of this scheme are compared and discussed according to the rise in the efficiency and security issues in provable data possession. The comparisons and study of those PDP schemes are done according to the problems arises and the solutions on those problems are mentioned.

REFERENCES

[1] Bing Rao, Zhigang Zhou, Hongli Zhang, Shuofei Tang and Renfu Yao "Outsourcing Cloud Data Privacy-Preserving Based on Over-Encryption," Communications in Computer and Information Science pp 109-116.

[2] Swapna Lia Anil and Roshni Thanka "A Survey on Security of Data outsourcing in Cloud," International Journal of Scientific and Research Publications, Volume 3, Issue 2, February 2013 .

[3] Y. Deswarte, J.-J. Quisquater, and A. Saïdane "Remote integrity checking," in Proc. 6th Working Conf. Integr. Internal Control Inf. Syst. (IICIS), 2003, pp. 1–11.

[4] Yongjun Ren, Zhenqi Yang, Jin Wang and Liming Fang "Attribute based Provable Data Possession in Public Cloud Storage," Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014.

[5] G. Ateniese "Provable data possession at untrusted stores," in Proc. 14th ACM Conf. Comput. Commun. Secur. (CCS), New York, NY, USA, 2007, pp. 598–609.

[6] Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Osama Khan, Lea Kissner, Zachary Peterson and Dawn Song "Remote Data Checking Using Provable Data Possession," ACM Transactions on Information and System Security, Vol. 14, No. 1, Article 12, Publication date: May 2011.

[7] F. Seb´e, J. Domingo-Ferrer, A. Martinez-Balleste, Y. Deswarte, and J.-J. Quisquater, "Efficient remote data possession checking in critical information infrastructures," IEEE Trans. on Knowl. and Data Eng., vol. 20, no. 8, 2008.

[8] C. C. Erway, A. Küpçü, C. Papamanthou, and R. Tamassia "Dynamic provable data possession. Cryptology," ePrint 2008/432.

[9] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik "Scalable and efficient provable data possession," In SecureComm, pp. 1–10, 2008.

[10] M. Kallahalla, E. Riedel, R. Swaminathan, Q. Wang, and K. Fu. Plutus "Scalable secure file sharing on untrusted storage," In FAST, pp. 29–42, 2003.

[11] B.-G. Chun, F. Dabek, A. Haeberlen, E. Sit, H. Weatherspoon, M. F. Kaashoek, J. Kubiatowicz, and R. Morris "Efficient replica maintenance for distributed storage systems," in NSDI'06: Proceedings of the 3rd Conference on Networked Systems Design & Implementation, Berkeley, CA, USA, 2006.

[12] R. Curtmola, O. Khan, R. Burns, and G. Ateniese, "MR-PDP: Multiple-Replica Provable Data Possession," in 28th IEEE ICDCS, 2008, pp. 411–420.

[13] Ayad F.Barsoum and M.Anwar Hasan "Provable possession and replication of data over cloud servers".

[14] H. Shacham and B. Waters "Compact proofs of retrievability," In ASIACRYPT, pp. 90–107, 2008.

[15] C. E. Shannon "Communication theory of secrecy systems," Bell Syst.Tech. J., vol. 28, no. 4, 1949.

[16] Ayad F.Barsoum and M.Anwar Hasan "Provable multicopy dynamic data possession," Information Forensics and Security, IEEE Transactions on (Volume:10 , Issue: 3).

[17] Cong Wang, Qian Wang, Kui Ren and Wenjing Lou "Ensuring data storage security in cloud computing," in IEEE Quality of Service, 2009. IWQoS.