ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 8, Issue, 11, pp.40961-40965, November, 2016

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

SECURITY ENSURED MULTICOPY DATA MANAGEMENT FRAMEWORK FOR CLOUD SERVICE PROVIDERS

***,1Hima Bindu, 2Vellaingiri, S. and 3Nagendra Babu, P.**

Sri Venkatesa Perumal College of Engineering and Technology, Puttur

ARTICLE INFO

Article History:
Received 15th August, 2016
Received in revised form
09th September, 2016
Accepted 23rd October, 2016
Published online 30th November, 2016

Key words:

Cloud computing,
Data replication,
Outsourcing data storage,
Dynamic environment.

ABSTRACT

The Cloud Service Providers (CSP) is deployed to support data storage and computational resources for the users. Data owners share the data values under the Cloud Service Providers. Authorized users access the data from the CSPs. Increasingly more and more organizations are opting for outsourcing data to remote cloud service providers (CSPs). Customers can rent the CSPs storage infrastructure to store and retrieve almost unlimited amount of data by paying fees metered in gigabyte/month. For an increased level of scalability, availability and durability, some customers may want their data to be replicated on multiple servers across multiple data centers. The more copies the CSP is asked to store, the more fees the customers are charged. Therefore, customers need to have a strong guarantee that the CSP is storing all data copies that are agreed upon in the service contract and all these copies are consistent with the most recent modifications issued by the customers. The cloud data sharing system is constructed with a map-based provable multi copy dynamic data possession (MB-PMDDP) scheme. The MB-PMDDP scheme achieves the following benefits. 1) it provides an evidence to the customers that the CSP is not cheating by storing fewer copies; 2) it supports outsourcing of dynamic data, i.e., it supports block-level operations, such as block modification, insertion, deletion and append and 3) it allows authorized users to seamlessly access the file copies stored by the CSP. The system verifies the data storage with multiple data copy model and data distribution operations.

Citation: ***Hima Bindu, Vellaingiri, S. and Nagendra Babu, P.*** **2016**. "Security ensured multicopy data management framework for cloud service providers", *International Journal of Current Research*, 8, (11), 40961-40965.

INTRODUCTION

One vision of 21st century computing is that users will access Internet services over lightweight portable devices rather than through some descendant of the traditional desktop PC. Because users won't have powerful machines, who will supply the computing power? The answer to this question lies with cloud computing. Cloud computing is a recent trending in IT that moves computing and data away from desktop and portable PCs into large data centers. It refers to applications delivered as services over the Internet as well as to the actual cloud infrastructure — namely, the hardware and systems software in data centers that provides these services. The key driving forces behind cloud computing is the ubiquity of broadband and wireless networking, falling storage costs and progressive improvements in Internet computing software. Cloud-service clients will be able to add more capacity at peak demand, reduce costs, experiment with new services and remove unneeded capacity, whereas service providers will increase utilization via multiplexing and allow for larger investments in software and hardware.

Corresponding author:* ***Hima Bindu*
Sri Venkatesa Perumal College of Engineering and Technology, Puttur

Currently, the main technical underpinnings of cloud computing infrastructures and services include virtualization, service-oriented software, grid computing technologies, management of large facilities and power efficiency. Consumers purchase such services in the form of infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), or software-as-a-service (SaaS) and sell value-added services to users. Within the cloud, the laws of probability give service providers great leverage through statistical multiplexing of varying workloads and easier management — a single software installation can cover many users' needs. Two different architectural models are considered for clouds. The first one is designed to scale out by providing additional computing instances on demand. Clouds can use these instances to supply services in the form of SaaS and PaaS. The second architectural model is designed to provide data and compute-intensive applications via scaling capacity (Ayad F. Barsoum and M. Anwar Hasan, 2015). In most cases, clouds provide on-demand computing instances or capacities with a "pay-as-you-go" economic model. The cloud infrastructure can support any computing model compatible with loosely coupled CPU clusters. Organizations can provide hardware for clouds internally, or a third party can provide it externally. A cloud might be restricted to a single organization or group, available to the general public over the Internet, or

shared by multiple groups or organizations. A cloud comprises processing, network and storage elements and cloud architecture consists of three abstract layers. Infrastructure is the lowest layer and is a means of delivering basic storage and compute capabilities as standardized services over the network. Servers, storage systems, switches, routers and other systems handle specific types of workloads, from batch processing to server or storage augmentation during peak loads. The middle platform layer provides higher abstractions and services to develop, test, deploy, host and maintain applications in the same integrated development environment. The application layer is the highest layer and features a complete application offered as a service. The shift of computer processing, storage and software delivery away from desktop and local servers, across the Internet and into next-generation data centers results, opportunities regarding data management. Data is replicated across large geographic distances, where its availability and durability are paramount for cloud service providers. It's also stored at untrusted hosts, which creates enormous risks for data privacy. Computing power in clouds must be elastic to face changing conditions. For instance, providers can allocate additional computational resources on the fly to handle increased demand. They should deploy novel data management approaches, such as analytical data management tasks, multitenant databases for SaaS, or hybrid designs among database management systems (DBMSs) and MapReduce-like systems so as to address data limitations and harness cloud computing platforms' capabilities.

Related Work

The problem of data integrity auditing in cloud have been extensively studied in past years by a number of Proof of Retrievability (POR) and Proof of Data Possession (PDP) schemes (Juels and Kaliski, 2007). Concepts of POR and PDP were first proposed separately using RSA-based homomorphic authentication tags. The efficiency of POR scheme was later enhanced by Shacham and Waters (Shacham and Waters, 2008), based on the BLS (Boneh-Lynn-Shacham) signature. To further enhance the efficiency of data integrity auditing, batch integrity auditing was introduced by Wang *et al*. Recently, Xu and Chang and Yuan and Yu proposed private and public POR schemes respectively with constant communication cost by using a nice algebraic property of polynomial. To support dynamic operations in verification, Ateniese *et al*. (2008) proposed another private PDP scheme with symmetric encryption. A Public integrity auditing with dynamic operations is introduced by Wang *et al*. (2009), based on the Merkle Hash Tree. Based on the rank information, Erway *et al*. also achieve the dynamic PDP. Zhu *et al*. (2011), later utilized the fragment structure to save storage overhead of authentication tags with the support of dynamic data. A private POR scheme with the support of dynamic data is recently proposed by Cash *et al*. (2013), by utilizing Oblivious RAM.

Although many efforts have been made to guarantee the integrity of data on remote server, most of them only consider single data owner who has the system secret key and is the only party allowed to modify the shared data on cloud. In order to improve the previous works to support multiple writers, Wang *et al*. (2012), first proposed a public integrity auditing scheme for shared data on cloud based on ring signature-based homomorphic authenticators. In their scheme, user revocation is not considered and the auditing cost grows with group size and data size. Recently, Wang *et al*. (2013), enhanced their previous public integrity verification scheme with the support of user revocation. However, if the cloud node responsible for tag update is compromised during user revocation process, attackers can discover the secret keys of all other valid users. What is more, verification cost of the TPA is significantly influenced by the error detection probability requirement and is also linear to the number of data modifier. Batch verification is not supported in their design. Therefore, this scheme is limited in its scalability.

Multicopy Data Management under Clouds

Outsourcing data to a remote cloud service provider (CSP) allows organizations to store more data on the CSP than on private computer systems. Such outsourcing of data storage enables organizations to concentrate on innovations and relieves the burden of constant server updates and other computing issues. Moreover, many authorized users can access the remotely stored data from different geographic locations making it more convenient for them. Once the data has been outsourced to a remote CSP which may not be trustworthy, the data owners lose the direct control over their sensitive data. This lack of control raises new formidable and challenging tasks related to data confidentiality and integrity protection in cloud computing. The confidentiality issue can be handled by encrypting sensitive data before outsourcing to remote servers. As such, it is a crucial demand of customers to have strong evidence that the cloud servers still possess their data and it is not being tampered with or partially deleted over time. Consequently, many researchers have focused on the problem of provable data possession (PDP) and proposed different schemes to audit the data stored on remote servers. PDP is a technique for validating data integrity over remote servers. In a typical PDP model, the data owner generates some metadata/information for a data file to be used later for verification purposes through a challenge-response protocol with the remote/cloud server.

The owner sends the file to be stored on a remote server which may be un-trusted and deletes the local copy of the file. As a proof that the server is still possessing the data file in its original form, it needs to correctly compute a response to a challenge vector sent from a verifier — who can be the original data owner or a trusted entity that shares some information with the owner. Researchers have proposed different variations of PDP schemes under different cryptographic assumptions. One of the core design principles of outsourcing data is to provide dynamic behavior of data for various applications. This means that the remotely stored data can be not only accessed by the authorized users, but also updated and scaled by the data owner. PDP schemes presented in (Sebé *et al*., 2008), focus on only static or warehoused data, where the outsourced data is kept unchanged over remote servers. Examples of PDP constructions that deal with dynamic data are (Erway *et al*., 2009). The latter are however for a single copy of the data file. Although PDP schemes have been presented for multiple copies of static data, see (Hao *et al*., 2010), to the best of our knowledge, this work is the first PDP scheme directly dealing with multiple copies of dynamic data. We provide a summary of related work. When verifying multiple data copies, the overall system integrity check fails if there are one or more corrupted copies. To address this issue and recognize which copies have been corrupted, we discuss a slight modification to be applied to the proposed scheme. Increasingly more and more organizations are opting for

outsourcing data to remote cloud service providers (CSPs). Customers can rent the CSPs storage infrastructure to store and retrieve almost unlimited amount of data by paying fees metered in gigabyte/month. For an increased level of scalability, availability and durability, some customers may want their data to be replicated on multiple servers across multiple data centers. The more copies the CSP is asked to store, the more fees the customers are charged.

Therefore, customers need to have a strong guarantee that the CSP is storing all data copies that are agreed upon in the service contract and all these copies are consistent with the most recent modifications issued by the customers (Xiaolong Xu *et al*., 2016). In this paper, we propose a map-based provable multi copy dynamic data possession (MB-PMDDP) scheme that has the following features: 1) it provides an evidence to the customers that the CSP is not cheating by storing fewer copies; 2) it supports outsourcing of dynamic data, i.e., it supports block-level operations, such as block modification, insertion, deletion and append; and 3) it allows authorized users to seamlessly access the file copies stored by the CSP. We give a comparative analysis of the proposed MB-PMDDP scheme with a reference model obtained by extending existing provable possession of dynamic single-copy schemes. The theoretical analysis is validated through experimental results on a commercial cloud platform. In addition, we show the security against colluding servers and discuss how to identify corrupted copies by slightly modifying the proposed scheme.

Problem Statement

Once the data has been outsourced to a remote CSP which may not be trustworthy, the data owners lose the direct control over their sensitive data. This lack of control raises new formidable and challenging tasks related to data confidentiality and integrity protection in cloud computing. The confidentiality issue can be handled by encrypting sensitive data before outsourcing to remote servers. As such, it is a crucial demand of customers to have strong evidence that the cloud servers still possess their data and it is not being tampered with or partially deleted over time. The following problems are identified from the current cloud security systems.

- There is no proof the client is using full utilized space allocated to him.
- Utilization is not effective and efficiency.
- When verifying multiple data copies, the overall system integrity check fails if there are one or more corrupted copies. To address this issue and recognize which copies have been corrupted.

Components of Cloud Service Provider

The cloud computing storage model considered in this work consists of three main components. (i) a data owner that can be an organization originally possessing sensitive data to be stored in the cloud; (ii) a CSP who manages cloud servers (CSs) and provides paid storage space on its infrastructure to store the owner's files; and (iii) authorized users — a set of owner's clients who have the right to access the remote data.
The storage model used in this work can be adopted by many practical applications. For example, e-Health applications can be envisioned by this model where the patients' database that contains large and sensitive information can be stored on the cloud servers. In these types of applications, the e-Health organization can be considered as the data owner and the physicians as the authorized users who have the right to access the patients' medical history. Many other practical applications like financial, scientific and educational applications can be viewed in similar settings.

Multicopy Data Access Model

The data owner has a file F consisting of m blocks and the CSP offers to store n copies $\{F_1, F_2, \ldots, F_n\}$ of the owner's file on different servers — to prevent simultaneous failure of all copies — in exchange of pre-specified fees metered in GB/month. The number of copies depends on the nature of data; more copies are needed for critical data that cannot easily be reproduced and to achieve a higher level of scalability. This critical data should be replicated on multiple servers across multiple data centers. On the other hand, non-critical, reproducible data are stored at reduced levels of redundancy. The CSP pricing model is related to the number of data copies. For data confidentiality, the owner encrypts his data before outsourcing to CSP. After outsourcing all n copies of the file, the owner may interact with the CSP to perform block-level operations on all copies. These operations includes modify, insert, append and delete specific blocks of the outsourced data copies. An authorized user of the outsourced data sends a data access request to the CSP and receives a file copy in an encrypted form that can be decrypted using a secret key shared with the owner. According to the load balancing mechanism used by the CSP to organize the work of the servers, the data-access request is directed to the server with the lowest congestion and thus the user is not aware of which copy has been received. We assume that the interaction between the owner and the authorized users to authenticate their identities and share the secret key has already been completed and it is not considered in this work.

Threat Model

The integrity of customers' data in the cloud may be at risk due to the following reasons. First, the CSP — whose goal is likely to make a profit and maintain a reputation — has an incentive to hide data loss or reclaim storage by discarding data that has not been or is rarely accessed. Second, a dishonest CSP may store fewer copies than what has been agreed upon in the service contact with the data owner and try to convince the owner that all copies are correctly stored intact. Third, to save the computational resources, the CSP may totally ignore the data-update requests issued by the owner, or not execute them on all copies leading to inconsistency between the file copies. The goal of the proposed scheme is to detect the CSP misbehavior by validating the number and integrity of file copies.

Map-Based Provable Multi Copy Dynamic Data Possession (MB-PMDDP) Scheme

We propose a MB-PMDDP scheme allowing the data owner to update and scale the blocks of files copies outsourced to cloud servers which may be un-trusted. Validating such copies of dynamic data requires the knowledge of the block versions ensure that the data blocks in all copies are consistent with the most recent modifications issued by the owner. Moreover, the verifier should be aware of the block indices to guarantee that the CSP has inserted or added the new blocks at the requested

positions in all copies. To this end, the proposed scheme is based on using a small data structure call a map-version table. The cloud data management model is divided into three major modules. They are Data owner, Cloud service provider and Authorized user. The data owner module is designed to manage the upload process. The Cloud Service Provider handles the data storage and distribution operations. The client application is designed to support the data download activities.

Data owner

A data owner that can be an organization originally possessing sensitive data to be stored in the cloud. The data owner has a file F consisting of m blocks and the CSP offers to store n copies of the owner's file on different servers. It is to prevent simultaneous failure of all copies, in exchange of pre-specified fees metered in GB/month. The number of copies depends on the nature of data; more copies are needed for critical data that cannot easily be reproduced and to achieve a higher level of scalability. This critical data should be replicated on multiple servers across multiple data centers.

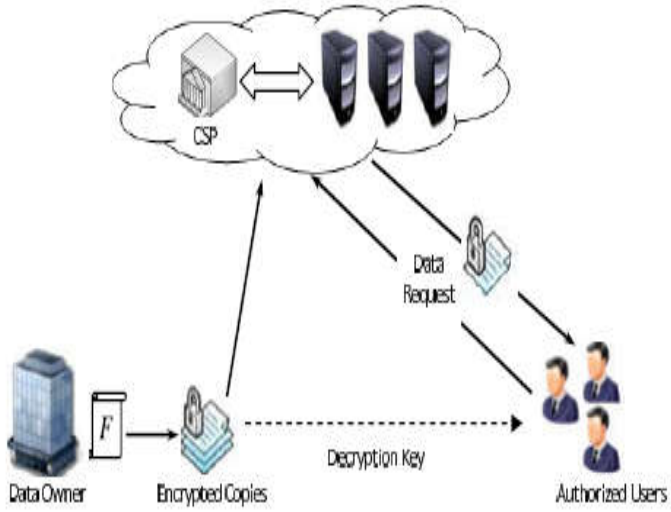

Fig. 1. Cloud Computing Data Storage System Model

Cloud service provider

A CSP who manages cloud servers (CSs) and provides paid storage space on its infrastructure to store the owner's files; According to the load balancing mechanism used by the CSP to organize the work of the servers, the data-access request is directed to the server with the lowest congestion and thus the user is not aware of which copy has been received.

Authorized user

A set of owner's clients who have the right to access the remote data. An authorized user of the outsourced data sends a data access request to the CSP and receives a file copy in an encrypted form that can be decrypted using a secret key shared with the owner.

Conclusion and Future Enhancement

We have proposed a new PDP scheme, which supports outsourcing of multi-copy dynamic data, where the data owner is capable of not only archiving and accessing the data copies stored by the CSP, but also updating and scaling these copies on the remote servers. To the best of our knowledge, the proposed scheme is the first to address multiple copies of dynamic data. The interaction between the authorized users and the CSP is considered in our scheme, where the authorized users can seamlessly access a data copy received from the CSP using a single secret key shared with the data owner. Moreover, the proposed scheme supports public verifiability, enables arbitrary number of auditing and allows possession-free verification where the verifier has the ability to verify the data integrity even though he neither possesses nor retrieves the file blocks from the server. In the future development the multi copy data management scheme can be enhanced to support intrusion detection and commercial cloud operations.

REFERENCES

Ateniese, G., Di Pietro, R., Mancini, L. V. and Tsudik, G. 2008. "Scalable and efficient provable data possession," in Proc. 4th Int. Conf. Secur. Privacy Commun. Netw. (SecureComm), Art. ID 9.

Ayad, F. 2015. Barsoum and M. Anwar Hasan, "Provable Multicopy Dynamic Data Possession in Cloud Computing Systems", IEEE Transactions On Information Forensics And Security, Vol. 10, No. 3, March.

Cash, D., Küpçü, A. and Wichs, D. 2013. "Dynamic proofs of retrievability via oblivious RAM," in Advances in Cryptology, vol. 7881, T. Johansson and P. Q. Nguyen, Eds. Berlin, Germany: Springer-Verlag, pp. 279–295.

Dražen Lučanin and Ivona Brandic, "Pervasive Cloud Controller for Geo temporal Inputs", IEEE Transactions On Cloud Computing, April.

Erway, C., Küpçü, A., Papamanthou, C. and R. Tamassia, 2009. "Dynamic provable data possession," in *Proc. 16th ACM Conf. Comput. Commun. Secur. (CCS)*, New York, NY, USA, pp. 213–222.

Hao, Z. and Yu, N. 2010. "A multiple-replica remote data possession checking protocol with public verifiability," in *Proc. 2nd Int. Symp. Data, Privacy, E-Commerce*, Sep. pp. 84–89.

Juels, A. and Kaliski, B. S. Jr., 2007. "PORs: Proofs of retrievability for large files," in Proc. 14th ACM Conf. Comput. Commun. Secur. (CCS), pp. 584–597.

Sebé, F., Domingo-Ferrer, J., Martinez-Balleste, A., Deswarte Y. and Quisquater, J.J. 2008. "Efficient remote data possession checking in critical information infrastructures," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 8, pp. 1034–1038, Aug.

Shacham, H. and Waters, B. 2008. "Compact proofs of retrievability," in Proc. 14th Int. Conf. Theory Appl. Cryptol. Inf. Secur. (ASIACRYPT), Melbourne, Vic., Australia, pp. 90–107.

Wang, B., Li, B. and Li, H. 2012. "Oruta: Privacy-preserving public auditing for shared data in the cloud," in Proc. IEEE 5th Int. Conf. Cloud Comput. (CLOUD), Washington, DC, USA, Jun., pp. 295–302.

Wang, B., Li, B. and Li, H. 2013. "Public auditing for shared data with efficient user revocation in the cloud," in Proc. *32nd IEEE Int. Conf. Comput. Commun.* (INFOCOM), Turin, Italy, Apr., pp. 2904–2912.

Wang, Q., Wang, C., Li, J., Ren, K. and Lou, W. 2009. "Enabling public verifiability and data dynamics for storage security in cloud computing," in Proc. 14th Eur. *Conf. Res. Comput. Secur.,* Saint-Malo, France, pp. 355–370.

Xiaolong Xu, Wanchun Dou, XuyunZhang and Jinjun Chen, 2016. "EnReal: An Energy-Aware Resource Allocation Method for Scientific Workflow Executions in Cloud Environment", IEEE Transactions On Cloud Computing, Manuscript Id, Jun.

Zhu, Y., Wang, H., Hu, Z., Ahn, G.-J., Hu, H. and Yau, S. S. 2011. "Dynamic audit services for integrity verification of outsourced storages in clouds," in Proc. ACM Symp. Appl. Comput. (SAC), 2011, pp. 1550–1557.

****