Australian Journal of Basic and Applied Sciences, 9(16) Special 2015, Pages: 462-466

ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Efficient Privacy-Preserving Protocol for Ensuring Remote Data Integrity

[1]L. Yamuna Devi and [2]Dr. K. Thilagavathy

[1]*Department of Computer Applications, Coimbatore Institute of Technology, Coimbatore, India*
[2]*Department of Physics, Coimbatore Institute of Technology, Coimbatore, India*

ARTICLE INFO

Article history:
Received 12 March 2015
Accepted 28 April 2015
Available online 14 June 2015

Keywords:
Cloud Computing, Remote Data Possession Checking, Data integrity, Privacy-Preserving, File Block grouping.

ABSTRACT

Cloud Computing lets applications and files to be hosted on a "Cloud" consisting of thousands of computers and servers, all linked together and accessible via the Internet. Besides the many benefits it offers, Cloud Computing poses many challenges to the cloud users as well as to the data owners, since the data is no longer stored in the local computers, instead they are stored and managed by the cloud providers. Anyone with permission can not only access the data stored in the cloud, but can also edit and collaborate on those data in real time. This brings many security challenges that are unique to this cloud paradigm. In this paper, we try to elaborate some of the major security challenges faced by the cloud users and the various techniques available to address those security challenges. This paper also proposes an efficient, privacy-preserving protocol for ensuring the integrity of user's data stored in Cloud Servers based on grouping of file blocks.

To Cite This Article: L. Yamuna Devi and Dr. K. Thilagavathy, Efficient Privacy-Preserving Protocol for Ensuring Remote Data Integrity. ***Aust. J. Basic & Appl. Sci.,*** *9(16): 462-466, 2015*

INTRODUCTION

Cloud Computing lets the computing resources, such as storage, database, application development, application services, and so on, that exists outside of the organizational premise to be leveraged by enterprise IT over the Internet. Cloud Computing allows the user to expand and contract their costs in direct proportion to their needs There are several service and deployment models exist for Cloud Computing. Each deployment model instance has one of two types: internal cloud or external cloud. Internal Clouds are deployed within an organization's network security perimeter, and external Clouds are deployed outside the perimeter.

In spite of the many advantages it provides, Cloud Computing poses new challenges which are unique to it. These challenges fall under two broad categories: performance and security. In this paper we analyze the potential security risks which are unique to the cloud paradigm and an efficient, privacy-preserving protocol for ensuring the integrity of remotely stored data is implemented.

Architecture of cloud computing:

Cloud Computing lets the users to avail the applications, platform, or infrastructure as a service through internet on demand basis. The benefits of Cloud Computing include minimized capital expenditure, location and device independence, scalability, multi-tenancy, and efficient use of computing power. Cloud offerings are available in multiple service and deployment models.

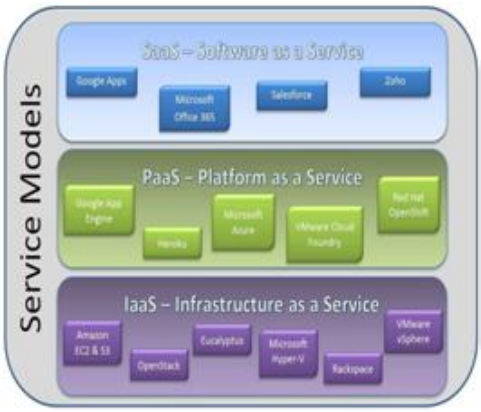

Fig. 1: Service Models of Cloud

Service and Deployment models of Cloud:

Service Models:

A service is defined as a fine-grained reusable resources (i.e., infrastructure or business processes) available from a service provider; this is now what is popularly called ―as a service. Cloud computing offers the following service models:

Corresponding Author: L. Yamuna Devi, Department of Computer Applications, Coimbatore Institute of Technology, Coimbatore, India
E-mail: yamunalogan@gmail.com

Australian Journal of Basic and Applied Sciences, 9(16) Special 2015, Pages: 462-466

Software as a Service (SaaS) is a kind of application that is available as a service to users; it delivers software as a service over the Internet, without installing and running the application on local computers in order to simplify the maintenance and support.

Platform as a Service (PaaS) model enables the deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers. The application can be deployed directly on the cloud infrastructure (without managing and controlling that infrastructure) using the programming languages and tools supported by a provider.

Infrastructure as a Service (IaaS) delivers a computer infrastructure that is a fundamental resource like processor cycles, storage capacity and network bandwidth to customers and provide automatic support for on demand scalability of computing and storage resources.

Deployment models:

Following Figure 2 shows the most commonly known deployment models of Cloud:

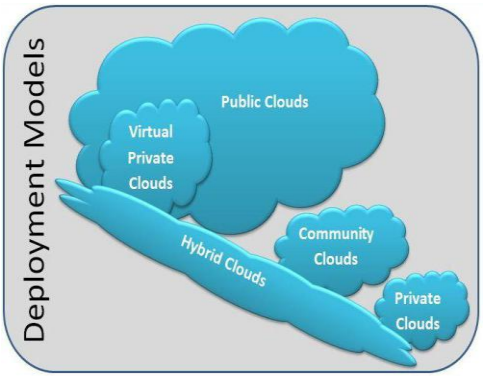

Fig. 2: Deployment Models of Cloud.

- *Private cloud:* The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.
- *Community cloud:* The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.
- *Public cloud:* The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.
- *Hybrid cloud:* The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

Challenges to cloud computing:

Even though cloud computing poses many features that makes it the new paradigm of choice for organizations, educational institutions, and individuals, it also poses some challenges that hinders its adoption in a wide manner (D. Sudha Devi et al., 2010). The main challenges in adopting cloud computing by many organizations are Reliability and Security (Yamuna Devi. L. et al., 2011). Among them the security issues are more prominent since cloud computing uses a distributed architecture, more data in transmitted over the net than traditional infrastructures, and because both customer data and programs are residing in cloud service provider premises users no longer physically posses the storage of their data.

According to ENISA, the European Network and Information Security Agency, the security risks identified in cloud computing are classified into three categories viz. Policy and organizational risks, Technical risks, and Legal risks. ENISA also states that some of the Policy and Organizational risks include Data Lock-In, Loss of Governance, Compliance Challenges, etc. Technical risks include Cloud provider malicious insider, Intercepting data in transit, Data leakage, Distributed Denial of Service (DDoS), compromise service engine, etc., and the Legal risks include data protection risks, licensing risks, etc.

In the above stated risks, Data Loss or Leakage is the deletion or alteration of records without a back-up of the original content (Cloud Security Alliance, March 2010). The threat of data compromise increases in the Cloud due to the number of and interactions between risks and challenges which are either unique to Cloud, or more dangerous because of the architectural or operational characteristics of the cloud environment (Cloud Security Alliance, March 2010). The data which are stored in the cloud data centers may be accidentally or maliciously deleted or modified by any unauthorized users or the Cloud Providers themselves. Hence it is of paramount important to introduce a secure auditing method to verify the integrity of cloud data.

Common Security Requirements:

When we refer security risks in cloud computing, it refers to both the security provided to the virtual machines in the cloud data center and the security given to the user data stored in the cloud. Since the VM security is beyond the scope of this paper, we will discuss only the data security. The common data security requirements include *confidentiality* which ensures that information is not disclosed to unauthorized persons, *integrity* which ensures that information held in a system is a proper representation of the information intended and that it

Australian Journal of Basic and Applied Sciences, 9(16) Special 2015, Pages: 462-466

has not been modified by an unauthorized person, *availability* which ensures that that information processing resources are not made unavailable by malicious action, and *non-repudiation* which ensures that agreements made electronically can be proven to have been made.

Apart from the security measures provided by the Cloud Service Provider(CSP)s, the cloud users may themselves introduce some additional measures to protect their data and verify the integrity of their outsourced data.

Cloud data integrity verification:

When data is moved across a network, the user might require a verification to ensure that data has not been modified after it has been sent. Users need to verify the integrity of their outsourced data against malicious users and misbehaving servers. How to efficiently verify the correctness of outsourced cloud data without the local copy of data files becomes a big challenge for data storage security in Cloud Computing.

The cloud data integrity verification problem can be addressed in two ways: either to empower the customers with tools to periodically conduct integrity checks of their data, or to introduce a Third Party Auditor (TPA) who will do the periodic task of checking cloud data integrity on behalf of the user (Cong Wang et al., 2011).

To perform this cloud data integrity verification, the TPA should meet the following requirements:

- TPA should audit the cloud data without demanding the local copy of data.
- TPA could not learn any knowledge about the data content stored in the Cloud Server during the auditing process.

There are several techniques available to perform this privacy-preserving public auditing for data stored in the cloud:

- Remote Data Possession Checking (RDPC) schemes
- Homomorphic Authentication

RDPC Schemes:

RDPC schemes allows data owners who have stored their data at a remote, untrusted server (such as Cloud Storage) to verify that the server posses the original data without retrieving the entire data. The RDPC schemes can fall under two groups :

(a) Provable Data Possession (PDP) schemes, and
(b) Proof of Retrievability (POR) schemes.

POR schemes check the possession of data and they can recover data in case of corruption. PDP schemes can be transformed to a POR scheme by adding erasure or error-correcting codes (Lanxiang Chen, 2012).

Homomorphic Authentication:

Homomorphic encryption is a form of encryption which allows specific types of computations to be carried out on ciphertext and obtain an encrypted result which decrypted matches the result of operations performed on the plaintext.

Homomorphic Authentication allows for arbitrary computing over encrypted data and allows for data processing without decryption. This technique encrypts the data in such a way that performing a mathematical operation on the encrypted information and then decrypting the result produces the same answer as performing an analogous operation on the unencrypted data. The correspondence between the operations on unencrypted data and the operations to be performed on encrypted data is known as homomorphism. For example, a user can store his data on an untrusted server in an encrypted form. Later he can send a query on the data to the server. The server can construct an encrypted response to the user's query without decrypting the user's data.

Proposed system:

System Model:

Instead of relying on a TPA to verify the correctness of data stored in remote Server, the Data Owner himself can perform the integrity verification checks on his data periodically. Remote Data Possession Checking (RDPC) scheme can be adopted for this purpose. This scheme follows the Challenge / Response Protocol in which the Verifier (Data Owner) possess a challenge for which the Prover (Storage Server) will respond. If this response meets the challenge possessed by the Verifier, then the integrity of the data is verified (Francesc Sebe et al., 2008).

Desirable Properties:

Efficiency of RDPC scheme can be assessed based on:

- Storage Overhead - Storage space used by the Verifier and the Prover for storing the file blocks and the tokens.
- Communication Overhead - Initial transfer of file blocks to the Server, transfer of challenge and response by the Verifier and the Prover.
- Computational Overhead - Initial computation of tokens by the Verifier and the response computation by the Prover.

Storage Overhead - The Verifier computes tokens on the encrypted file blocks and uploads the file blocks to the Storage Server. If storage space is of much concern to the Verifier, the Verifier can encrypt the tokens and uploads the encrypted tokens along with the encrypted file blocks.

Table 1: Resource Overheads at Prover's End and Verifier's End.

	Storage Overhead		Communication Overhead
Cloud Server storing encrypted file blocks + encrypted tags	Prover	High	High
	Verifier	Low	

Cloud Server storing only encrypted file blocks	Prover	Low	Low
	Verifier	High	

Design Goals:

- Privacy Preserving - The original file contents should not be revealed to the Prover during the tenure of the data storage as well as during computing tags for the challenged data blocks as response to the Verifier's challenge. This goal is not achieved in many works like (Cong Wang et al., 2012). But the proposed protocol achieves this goal by encrypting the file blocks using PBE encryption algorithm.
- After uploading the file blocks to the storage server, data will no longer be available to the data owner. Hence verification of the correctness of the data in remote Cloud storage must be conducted without explicit knowledge of the whole data. This requirement is satisfied in this scheme since the Data Owner / Verifier needs only the precomputed tags for integrity verification.
- Acceptable Storage Overhead to the Verifier since he / she need not possess the original file for verification of the response from the Prover, instead he only needs the precomputed tags for group of file blocks. Grouping of file blocks reduces the number of tags to be computed and stored at the Verifier's end.
- Minimal Communication Overhead since after the initial upload of the file blocks to the Storage Server, only the tags computed by the Server for the challenged blocks are transferred to the Verifier which again involves only minimal network bandwidth.

Notion and Preliminaries:

The Challenge / Response protocol for verifying the integrity of remotely stored data involves the following five steps: (i) Setup Phase, (ii) Tag Generation, (iii) Challenge, (iv) Proof Generation, (v) Proof Verification.

(i) Setup Phase:

During Setup Phase, the Data Owner split the original file F into fixed size blocks(f1, f2, … fn) and then encrypt the file blocks using PBE algorithm. Password-Based Encryption (PBE) derives an encryption key from a user-given password. In order to make the key sufficiently stronger to withstand against any brute force attack, a 64 bit random number, known as salt, is added with the password and (key + salt) is hashed multiple times to derive the encryption key.

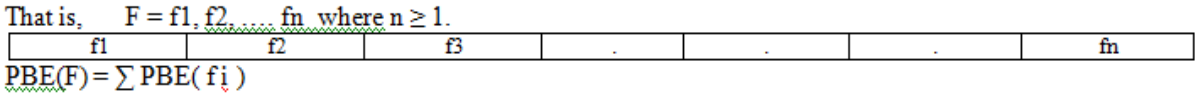

Fig. 3: Original File split into n blocks.

(ii) Tag Generation:

In this phase, instead of computing tags for each individual file block, the blocks are grouped based on their index, and tags are computed for the group of blocks. This grouping reduces the number of tags to be pre computed and stored at the Verifier's side. Each Tag i for the group 'i' is computed as follows:

$$Tag_i = \sum_{i=1}^{m} \sum_{j=0}^{k-1} hash(f_{(i+j*k)})$$

where, m = number of file block groups,
k = number of file blocks within a group

(iii) Challenge Phase:

The Data Owner / Verifier, either at periodical intervals or during any suspicion, challenges the Sever with necessary metadata.

(iv) Proof Generation:

The Storage Server / Prover, when challenged by the Data Owner, computes tags for the specified group of data blocks (Tag r) using the same algorithm and sends it to the Data Owner.

(v) Proof Verification:

Upon receiving the response from the Storage Server / Prover, the Verifier compares the tag (Tag r) returned by the Server with pre computed tag (Tag i) for that group of blocks. If Tag r = Tag i , then the Data Owner / Verifier is assured of the integrity of the data stored at the remote Cloud Server.

Conclusion:

A Challenge / Response Protocol for verifying the integrity of remotely stored data based on PDP scheme is implemented. The above protocol is benefitted by the fact that the verifiable tags are computed for a group of data blocks, instead of individual blocks. This reduces the storage as well as computation overhead at the Verifier's side since only less number of tags are computed and stored. This Protocol can also support dynamic data updates since the data block grouping operation can be extended to add additional file blocks by recomputing tags only for the affected file block groups.

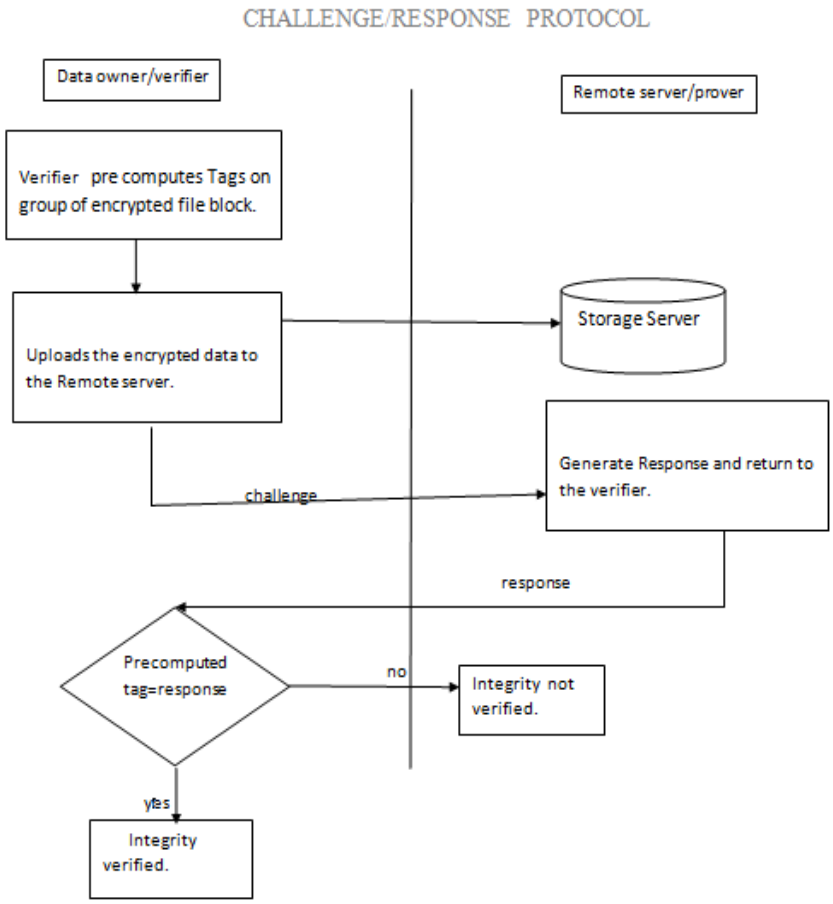

Fig. 4: Challenge/Response Protocol for Data Integrity Verification.

REFERENCES

Cloud Computing: Benefits, risks and recommendations for Information Security", European Network and Information Security Agency (ENISA).

Cong Wang, Qian Wang, Kui Ren, Ning Cao and Wenjing Lou, 2012. "Toward Secure and Dependable Storage Services in Cloud Computing", IEEE Transactions on Services Computing, 5(2): 220-232.

Cong Wang, Sherman S.M. Chow, Qian Wang, Kui Ren and Wenjing Lou, 2011. "Privacy-Preserving Public Auditing for Secure Cloud Storage", IEEE Transactions on Computers.

Sudha Devi, D., P. Aruna, L. Yamuna Devi, P. Raghu, 2010. "Managing Data in the Cloud: An Analysis", National Conference on Communication, Networking and Computing (NCCNC 2010) proceedings.

Francesc Sebe, Josep Domingo-Ferrer, Antoni Martinez-Balleste, Yves Deswarte and Jean-Jacques Quisquater, 2008. "Efficient Remote Data Possession Checking in Critical Information Infrastructures", IEEE Transactions on Knowledge and Data Engineering, VOL 20, NO. 8.

http://en.wikipedia.org/wiki/Homomorphic_encryption

Lanxiang Chen, 2012. "Using Algebraic Signatures to Check Data Possession in Cloud Storage", Future Generation Computer Systems.

Top Threats to Cloud Computing V1.0", Cloud Security Alliance (CSA), March 2010.

Yamuna Devi, L., P. Aruna, D. Sudha Devi and N. Priya, 2011. "Security in Virtual Machine Live Migration for KVM", International Conference on Process Automation, control and Computing (IEEE).