Healthcare Application Cloud - Ensuring Data Storage Security and Error Localization

Mukesh Kumar Patel[1] Upen Nathwani[2]
[1] PG Student [2] Assistant Professor
[1, 2] Computer Engineering Department Noble Group of Institutions, Junagadh, Gujarat

Abstract---Cloud computing is a recent trend 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 provide these services. Cloud computing faces many of the same challenges as other information and network technologies: performance, security, resiliency, interoperability, data migration, and transition from legacy systems. To ensure the correctness of users' Healthcare data in the Healthcare cloud, it is propose a scheme which is flexible and effective with two features, opposing to its predecessors. By making use of homomorphic token with distributed verification of erasure-coded data, the scheme achieves the integration of storage correctness insurance and data error localization, which means the identification of misbehaving server.

Keywords: Cloud Computing, Personal Health Record, Electronic Health Record, Electronic Medical Record, Token, Data Integrity.

I. INTRODUCTION

The new concept of *Cloud Computing* offers dynamically scalable resources provisioned as a service over the Internet and therefore promises a lot of economic benefits to be distributed among its adopters. Depending on the type of resources provided by the Cloud, distinct layers are present. The bottom-most layer provides basic infrastructure components such as CPUs, memory, and storage, and is henceforth often denoted as Infrastructure-as a- Service (IaaS). Amazon's Elastic Compute Cloud (EC2) is a prominent example for an IaaS offer. On top of IaaS, more platform-oriented services allow the usage of hosting environments tailored to a specific need. Google App Engine is an example for a Web platform as service (PaaS) which enables to deploy and dynamically scale Python and Java based Web applications. Finally, the top-most layer provides it users with ready to use applications also known as Software as-a-Service (SaaS). To access these Cloud services, two main technologies can be currently identified. Web Services are commonly used to provide access to IaaS services and Web browsers are used to access SaaS applications. In PaaS environments both approaches can be found. [1]

With the development of technology, the Internet has entered into a new phase. Rather than running software on a desktop computer or server, Internet users are now able to use the "Cloud" - a networked collection of servers, storage systems, and devices - to combine software, data, and computing power scattered in multiple locations across the network. Cloud computing is a topic on software and distributed computing based on Internet, which means users can access storages and applications from remote servers by web browsers or other fixed or mobile terminals. Because the constrained resources of fixed or mobile terminals, cloud computing will provide terminals with powerful complementation resources to acquire complicated services. Cloud computing is a complex infrastructure of software, hardware, processing, and storage that is available as a service. Cloud computing offers immediate access to large numbers of the world's most sophisticated supercomputers and their corresponding processing power, interconnected at various locations around the world, proffering speed in the tens of trillions of computations per second. [2]

II. WHAT IS CLOUD COMPUTING

A. *Background of cloud computing*

In recent years, Internet has been developing very quickly. The cost of storage, the power consumed by computer and hardware is increasing. The storage space in data center can't meet our needs and the system and service of original internet can't solve above questions, so we need new solutions. At the same time, large enterprises have to study data source fully to support its business. The collection and analysis must be built on a new platform. Why we need cloud computing? It is to utilize the vacant resources of computer, increase the economic efficiency through improving utilization rate, and decrease the equipment energy consumption.

B. *Cloud computing principle*

It is difficult to define the cloud computing. Computing is a virtual pool of computing resources. It provides computing resources in the pool for users through internet. Integrated cloud computing is a whole dynamic computing system. It provides a mandatory application program environment. It can deploy, allocate or reallocate computing resource dynamically and monitor the usage of resources at all times. Generally speaking cloud computing has a distributed foundation establishment, and monitor the distributed system, to achieve the purpose of efficient use of the system. Cloud computing collects all the computing resources and manages them automatically through software.

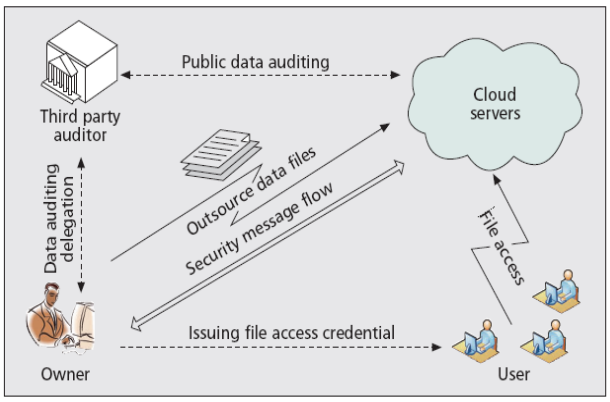

Fig.1: The architecture of the cloud data storage service [9]

In the process of data analysis, it integrates the history data and present data to make the collected information more accurate and provide more intelligent service for users and enterprises. The users need not care how to buy servers, software, solutions and so on. Users can buy the computing resource through internet according to their own needs. [3] The architecture of the cloud computing is shown in the figure 1.

Following network entities are described below:

1) User: An entity, which is used to retrieve the data from cloud server.

2) Cloud Server (CS): An entity, which is managed by cloud service provider (CSP) to provide data storage service and has significant storage space and computation resources (we will not differentiate CS and CSP hereafter).

3) Third Party Auditor (TPA): An optional TPA, who has expertise and capabilities that users may not have, is trusted to assess and expose risk of cloud storage services on behalf of the users upon request.

4) Owner: An entity, who has data to be stored in the cloud and relies on the cloud for data storage and computation, can be either enterprise or individual customers. In cloud data storage, a user stores his data through a CSP into a set of cloud servers, which are running in a simultaneous, cooperated and distributed manner. Data redundancy can be employed with technique of erasure correcting code to further tolerate faults or server crash as user's data grows in size and importance. Thereafter, for application purposes, the user interacts with the 3 cloud servers via CSP to access or retrieve his data. [9]

C. Cloud Service Models

1) *Infrastructure as a Service (IaaS):* This model allows users to rent processing, storage, networks, and other resources. The user can deploy and run the guest OS and applications. The user does not manage or control the underlying cloud infrastructure but has control over OS, storage, deployed applications, and possibly select networking components.

2) *Platform as a Service (PaaS):* This model provides the user to deploy user-built applications onto the cloud infrastructure that are built using programming languages and software tools supported by the provider (e.g., Java, python, .Net). The user does not manage the underlying cloud infrastructure.

3) Software as a Service (SaaS): This refers to browser initiated application software over thousands of cloud customers. On the customer side, there is no upfront investment in servers or software licensing. On the provider side, costs are rather low, compared with conventional hosting of user applications. Cloud offers four service deployment modes: private, public, managed, and hybrid. These modes demand different levels of security implications. The different service level agreements and service deployment modalities imply the security to be a shared responsibility of all the cloud providers, the cloud resource consumers and the third party cloud enabled software providers. [4]

III. Overview Of Healthcare Cloud

Below defines the concept of Personal Health Record (PHR), Electronic Health Record (EHR), and Electronic Medical Record (EMR) and then why cloud computing is attracted to healthcare IT.

A. The Definitions of PHR, EMR and EHR

The terms of EHRs and EMRs are used interchangeably by many in both healthcare industry and the press or health science literature. Strictly speaking, these two terms describe completely different concepts according to HIMSS (Health Information and Management System Society) Analytics. Both EMRs and EHRs are critical to the grand vision of healthcare digitization for improving safety, quality and efficiency of patient care and reducing healthcare delivery costs. EMRs are owned by individual healthcare providers, whereas EHRs are typically composed of some subsets of EMRs. The interoperability of EHRs is a fundamental enabling technology for EMRs to reach its full potential in revolutionizing the healthcare delivery with high quality and affordable cost.

1) *Personal Health Record* is typically a health record that is initiated and maintained by an individual. An ideal PHR would provide a complete and accurate summary of the health and medical history of an individual by gathering data from many sources, including EMRs and EHRs, and making this information accessible to those who have the necessary electronic credentials to view the information.

2) *Electronic Medical Record* is the legal record of what happened to the patient during their encounter at a Care Delivery Organization (CDO) across inpatient and outpatient environments and is owned by the CDO. EMR is created, used and maintained by healthcare practitioners to document, monitor, and manage health care delivery within a CDO. A core EMR system is composed of the clinical data repository (CDR), clinical decision support system (CDSS), controlled medical vocabulary (CMV), computerized provider order entry (CPOE), pharmacy management system, and the electronic medication administration record (eMAR), a functionality in the electronic clinical documentation systems of most vendors.

3) *Electronic Health Record* is a subset of EMR record maintained by each CDO and is created and owned by the patient. An HER typically has patient input and access that spans episodes of care across multiple CDOs within a community, region, or state, the primary purpose of the EHR is to provide a documented record of care which supports both present and future care received by the patient from the same or other clinicians or care providers. This documentation provides a mean of communication among clinicians controlling to the patient's care.

B. Why is cloud computing attractive to healthcare IT?

Many healthcare providers and insurance companies today have adopted some form of electronic medical record systems, though most of them store medical records in centralized databases in the form of electronic records. Typically, a patient may have many healthcare providers, including primary care physicians, specialists, therapists, and other medical practitioners. In addition, a patient may use multiple healthcare insurance companies for different types of insurances, such as medical, dental, vision, and so forth. Currently, each provider typically has its own database for electronic medical records (EMRs). Sharing information between healthcare practitioners across administrative boundaries is translated to sharing

information between EMR systems. The electronic records sharing between different EMR systems are called electronic health records (EHRs). The interoperation and sharing among different EMRs has been extremely slow. Cost and poor usability have been cited as the biggest obstacles to adoption of health IT, especially Electronic Health Records (EHR) systems. Cloud computing provides an attractive IT platform to cut down the cost of EHR systems in terms of both ownership and IT maintenance burdens for many medical practices. [5]

IV. DESIGN GOALS

To ensure the security and fidelity for cloud data storage under the aforesaid antagonist model, we aim to design efficient mechanisms for dynamic data verification and operation and achieve the following goals: (1) Storage correctness: to ensure users that their data are indeed stored appropriately and kept unharmed all the time in the cloud. (2) Fast localization of data error: to effectively locate the faulty server when data corruption has been erected. (3) Dynamic data support: to maintain the same level of storage correctness assurance even if users modify, delete or append their data files in the cloud. (4) Dependability: to enhance data availability against Intricate failures, malevolent data modification and server colluding attacks, i.e. minimizing the effect brought by data errors or server failures. (5) Lightweight: to enable users to perform storage correctness checks with minimum overhead.

Ensuring Cloud Data Storage In cloud data storage system, users store their data in the cloud and no longer possess the data locally. Thus, the correctness and availability of the data files being stored on the distributed cloud servers must be guaranteed. One of the key issues is to effectively detect any unauthorized data modification and corruption, possibly due to server compromise and/or random Byzantine failures. Besides, in the distributed case when such inconsistencies are successfully detected, to find which server the data error lies in is also of great significance, since it can be the first step to fast recover the storage errors. To address these problems, our main scheme for ensuring cloud data storage is presented in this section. The first part of the section then, the homomorphic token is introduced. IT defines the common theme is that a homomorphism is a function between two algebraic objects that respects the algebraic structure. The token computation function we are considering belongs to a family of universal hash function, chosen to preserve the homomorphic properties, which can be perfectly integrated with the verification of erasure- coded data. Subsequently, it is also shown how to derive a challenge response protocol for verifying the storage accuracy as well as identifying misbehaving servers. Finally, the method for file reclamation and error recuperation based on erasure-correcting code is outlined. [6]

F – Healthcare File. It is denoted as matrix of m equal- sized data vectors each consisting of l blocks.
R – Dispersal matrix used for Reed-Solomon coding
C – Encoded Healthcare file matrix
PRF – Pseudo random Function
PRP – Pseudo Random Permutation
Ver – Version number.

A. Algorithm: TOKEN PRE-COMPUTATION

Step. 1 : Start
Step. 2 : Choose Healthcare file F to upload & encrypt the file using AES
Step. 3 : Generate n*m Vector Matrix D on file F.
Step. 4 : Create Reed Solomon Matrix P over Galois Field GF (2w) where w=4.
Step. 5 : Generate Matrix C=D*P. It is Checksum Matrix created for fault tolerance.
Step. 6 : Compute Token over Matrix C i.e., Compute Token(c, l, t, r) where l-block size, t- no. of tokens, r-indices per verification. Compute the tokens by pseudorandom function & pseudorandom permutation function.
Step. 7 : Precomputed tokens to be stored on the main cloud server.
Step. 8 : Disperse the file over the Cloud. i.e. Disperse File Matrix D.
Step. 9 : End.

B. Verifying Data distribution

To eliminate the errors in storage systems key prerequisite is to locate the errors. The newly computed tokens from servers for each challenge are compared with pre-computed tokens to determine the correctness of the distributed storage. This also gives information to locate potential data errors.

C. Algorithm: CORRECTNESS VERIFICATION

Step. 1 : Start Challenge for n- total number of cloud servers.
Step. 2 : Fetching precomputed tokens from main cloud server.
Step. 3 : Reading file blocks from all cloud servers for calculating new tokens.
Step. 4 : Generate Vector Matrix D on all file blocks that are read in step 3.
Step. 5 : Create Reed Solomon Matrix P
Step. 6 : Generate Matrix C=D*P. On this matrix, new tokens will be computed.
Step. 7 : Compute token on Matrix. Compute Token(c,l,t,r)
Step. 8 : If (PrecomputedToken = Token which is newly computed) then, Data is complete .Else Data is Corrupt. For that, initiate the recovery.
Step. 9 : End [7]

V. CONCLUSION

In this paper, I investigated the problem of data security in Healthcare application cloud data storage, which is essentially a distributed storage system. The patient's data is stored in the Healthcare application cloud. To check whether the data is stored in the server is intact or not, I proposed an effective and flexible distributed scheme with explicit dynamic data support, including block update, delete, and append**.** By making use of homomorphic token with distributed verification of erasure coded data, it achieves the integration of storage correctness insurance and data error localization. Whenever data corruption has been detected during the storage correctness verification across the servers, it can identify the misbehaving server(s). Any unauthorized user accessed the healthcare files, automatically blocks the computer.

REFERENCES

[1] Meiko Jensen, J¨org Schwenk, Nils Gruschka, Luigi Lo Iacono "On Technical Security Issues in Cloud Computing", ISBN: 978-0-7695-4130-3/10© 2010 IEEE.

[2] Jian Wang Yan Zhao Shuo Jiang Jiajin Le "Providing Privacy Preserving in cloud computing", ISBN: 978-0-7695-4130-3/10© 2010 IEEE.

[3] Shuai Zhang, Shufen Zhang, Xuebin Chen, Xiuzhen Huo "C loud Computing Research and Development Trend", ISBN: 978-0-7695-4130-3/10© 2010 IEEE.

[4] Kai Hwang and Sameer Kulkarni, Yue Hu "Cloud Security with Virtualized Defense and Reputation-based Trust Management", ISBN: 978-0-7695-4130-3/10© 2010 IEEE.

[5] Rui Zhang and Ling Li "Security Models and Requirements for Healthcare Application Clouds", ISBN: 978-0-7695-4130-3/10© 2010 IEEE.

[6] Bhupesh Kumar Dewangan and Sanjay Kumar Baghel "Data Storage in Cloud Server by Token Pre-Computation" ISSN: 2277 128X, April 2013 International Journal of Advanced Research in Computer Science and Software Engineering.

[7] P.Dhanalakshmi, V.Ramesh "Remote Data Integrity In Cloud Security Services" ,January 2013 ISSN 2250-2459 (Online) ICISC-2013), INDIA

[8] Cong Wang, Qian Wang, and Kui Ren "Ensuring Data Storage Security in Cloud Computing", 2009 ISSN: 978-1-4244-3876-1 ©2009 IEEE

[9] Nandeesh.B.B, Ganesh Kumar R, Jitendranath Mungara "Secure and Dependable Cloud Services for TPA in Cloud Computing", August 2012 ISSN: 2278-3075, IJITEE

