Volume: 9 Issue: VII Month of publication: July 2021

DOI: https://doi.org/10.22214/ijraset.2021.37109

Sharing Healthcare Records in the Cloud Using Attribute-Based Encryption and De-Duplication

Rushikesh Umak
Dept. of Computer Science and Engineering, PRM institute of technology & reseach Amravati, India

***Abstract:** Cloud based healthcare computing have changed the face of healthcare in many ways. The main advantages of cloud computing in healthcare are scalability of the required service and the provision to upscale or downsize the data storge, collaborating Artificial Intelligence (AI) and machine learning. The current paper examined various research studies to explore the utilization of intelligent techniques in health systems and mainly focused into the security and privacy issues in the current technologies. E-Healthcare is an emerging field of medical informatics, referring to the delivery of health services and information using the Internet and related technologies. Rendering efficient storage and security for all data is very important for cloud computing. Securing and privacy preserving of data is of high priority when it comes to cloud storage. E-Healthcare is the most important source in the healthcare society. E-healthcare system is now being popularized globally. Implementing the E-healthcare system will have more advantages such as online services for teleconsultation (second medical opinion), e-prescription, e-referral, telemonitoring, telecare etc. E-healthcare system provides high level of security and cost-effective use of patients records, information and communication in support of healthcare and health related issues.*

***Keywords:** E-Healthcare, PHI (Personal Health Information), cloud computing, deduplication, ABE (Attribute Based Encryption).*

I. INTRODUCTION

Cloud computing is one of the emerging technology, which helped several organizations to save money and time adding convenience to the end users. Thus the scope of cloud storage is vast because the organizations can virtually store their data's without bothering the entire mechanism. Cloud Computing provides key advantage to the end users like cost savings, Able to access the data irrespective of location, performance and security. Innovative changes have permitted progressive answers for be actualized to upgrade the nature of human existence. Analysts considering the development of innovation have identified and assessed wellbeing data from these sources to acquire information and take care of wellbeing-related issues. In this manner, the advancement of incorporated medical care innovation has the likelihood to enhance efficiency and improve understanding of the results at each level of the medical care framework.. In current healthcare systems, there is a high demand on establishing a framework that minimizes time-consuming work and expensive procedures to retrieve a patient's medical record and integrating this varying set of medical data consistently to deliver it to the healthcare industry. Electronic health records (EHRs) have been widely accepted to allow patients, insurance companies, and healthcare providers to initiate, control and process patients' healthcare information from any place, and at any time. Thus, healthcare providers accept moving their data and operations to the clouds that can perform their operations more efficiently and eliminate the physical distance concern between patients and providers. Cloud service enables different doctors to obtain an access to a patient's health record even if they are kilometres apart. There is no need for the doctors to make a phone call to ask for a move of the health records; they will just access them in the clouds. Despite all the benefits cloud computing provides for healthcare systems, data privacy and security are among the major concerns, which make healthcare move slowly towards the acceptance of these new technologies. Cloud computing benefits come at a price of the emergence of different risks related to information security that must be cautiously addressed. Risks differ according to the criticality of the data to be processed or stored, and how the specific cloud provider has developed their specific cloud services. In order to be appealing to healthcare community, cloud computing should maintain required guarding to address HIPAA (Health Information Portability and Accountability Act) of U.S. Department of Health and Human Services (2013) and other security and privacy requirements. Although Electronic Health Records (EHRs) has been regulated in standards, such as HIPAA, several cloud providers are still not compliant with them. In order to secure healthcare data, the first step to be taken is to categorize the data in the Electronic Health Records (EHRs) in correspondence to its level of security sensitivity. The first category is Personal Identifiable Information (PII), such as patient records, normally saved in a relational database as structured data.

The second category is Healthcare data (PHI), which is typically consists of large media files such as radiology, CT scan, x-ray, and other types of video and images that conceal patient's identity. Such files are often stored in distributed storage. A medical record has some components that are classified by both individuals and organizations, such as HIPAA of U.S. Department of Health and Human Services (2005) and HITECH (Health Information Technology, 2009), as highly critical and should be disclosed only to the entities that have an explicit access right to them. This is because revealing such data can lead to unjustly show bias against an individual or refuse them chances that they otherwise entitled to. For example, knowing that a person is diabetic might negatively influence their professional growth, personal relationships, insurance cost, and employment opportunities. Outsourcing the storage of unencrypted information in the cloud, is of a high danger. For a highly sensitive data, such as Personal Health Informations (PHIs), locating them unencrypted, out of site, is considered against the law. However, to access data stored on a distance server, the Cloud providers need to access the primitive, i.e. un-encrypted, data. Most people do not have full confident on the Cloud providers for their sensitive healthcare data because there is no law regulating how they use this data and whether the patients have control over them. On the other hand, data encryption might counteract the advantages of cloud computing, unless the cloud service providers get the secret decryption key. Traditional cryptography is not a solution in this situation (. Patients may only want portions of their record made available to all doctors and specific portions to be available to specific users, i.e. insurance company. Patients can be given maximum control over their data by encrypting each portion of a patient’s record under a different policy. Access control policies should be active to ensure that accessing sensitive information is restricted only to parties that have a valid privilege. This feature can be provided by Attribute-Based encryption.

II. EASE OF USE

A. Advantages of Cloud-Based E-healthcare Systems

The more a healthcare center connects system information to a global computer network such as the Internet, the more it opens up access from around the world and facilitates data leaks. The need for an electronic health record should be protected from illegal users who may misuse this for a variety of purposes. Identity-based encryption is one of the best security solutions to protect e-Health record data. The algorithm deals with problems found in common cryptographic techniques using any thread as a public key. The system can enhance the security of health records by adding authentication procedures to connected servers . In this system, communication between this servers uses encrypted data using ABE, so that each server can perform the encryption and decryption process during the data exchange. Only servers with IDs can access and extract health record data. Currently, test results show performance relative to the speed of the algorithm used in the system . Cloud Storage is a computer model that stores data on the Internet or in the cloud. Cloud storage is delivered according to demand and capacity and costs that will leave the customer investing and managing their data storage infrastructure. This provides speed, scale, and durability. Below are some of the general advantages of cloud computing; in our case, we focus on E-health systems.

1) Ease of access using a ‘Web Browser’ with integrated Single-Sign-On (SSO). No requirement for VPN to access Cross Sites or Networks. Simplified Management and On-demand Scalability.
2) No Overhead Cost to maintain the physical infrastructure.
3) No Hardware post warranty charges for the physical infrastructure.
4) No Power Consumption.

One of the major schemes in healthcare systems is attribute-based encryption for data. Encryption provides high-class access control for every user and revocation, scalability, dynamic user management, and traceability

B. Challenges in Cloud Computing Cloud Computing

Challenges are always been there. Companies and organizations are aware of the values that cloud computing brings and are taking necessary steps towards the transition to the cloud environment. Like any new technology the adoption of cloud computing is also full of issues and other challenges. Some of them are:

1) Confidentiality: Confidentiality is a process or mechanism of safeguarding patient health data from unauthorized access from public or internal users. Unauthorized access is dangerous and can potentially result in data leakage and can even cause serious damage to businesses. With respect to the data size, the number of patients on devices increases, and there is a huge potential threat to the data to expose these to external parties. Confidentiality is important in the healthcare industry as the patient can be reluctant to give personal details to doctors if they are not confident with the confidentiality. By implementing access control and using encryption techniques, confidentiality can be achieved.

2) *Integrity:* Integrity is important factor to make sure that the data are not changed at any single point in time. The HIPAA Security illustrates that covered entities must implement procedures and policies to protect electronic healthcare information from improper destruction or alteration. Integrity can be achieved by a hashing mechanism or checksum for all the data. One of the best and accurate ways is by implementing block chain technology as it is merely impossible to change the hash of the data as it will change the entire chain if any of the hashes are changed.
3) *Availability:* The information must be available all the time. Business critical systems should be clustered or must have high availability to have maximum uptime without service interruptions.
4) *Data Violations:* Business Impact on Company Dignity and Trust for Customers or Partners. Degradation of intellectual property by competitors can lead to product outsourcing, financial discovery, and the occurrence of events and forensics.
5) *Wrong fix:* This is one of the most common cloud challenges. As cloud computing is a shared resource, any misconfiguration of the datacenter will lead to complete exploration of all the customer data hosted within the same datacenter.
6) *Lack of Security Technologies:* The biggest challenge during the transition to cloud computing is the implementation of appropriate security architecture to withstand cyberattacks. Unfortunately, this process remains a mystery for several organizations.
7) *Account Hijacking:* A key feature where attackers gain access to accounts, and serious or sensitive rights are exploited. Criminal attacks on sensitive data, cloud system exploits, or access to stolen signals can put these accounts at risk.
8) *Insider Threat:* Circumstances have been identified including malicious servers, employees saving sensitive data on their unprotected devices and programs, employees or other insiders who steal stolen emails exposed by malicious attacks on company assets
9) *Unsecured APIs:* Cloud computing providers develop a range of user software and APIs to allow customers to manage and interact with cloud services. The security and availability of standard cloud services are linked to the security of those APIs. Poorly designed APIs can lead to misuse or even worse, infringement of information. Exposed, broken, and hacked APIs have serious concerns about data breaches. Healthcare really needs to understand the safety requirements for designing and introducing visible connectors online

C. Related Work

Private data deduplication technique for storing private and personalized data was introduced and formalized by Wee Keong Ng SCE, Yonggang Wen SCE, and Huafei Zhu [1]. Where a private data deduplication protocol allows a client, who holds a private data which proves to a server who holds a summary string of the data that he/she is the owner of that data without revealing further information to the server. The security of the private data is formalized using private data deduplication techniques. This protocol is the first data deduplication for private data. An architecture in DupLESS: Server-Aided Encryption for Deduplicated Storage system was described by Mihir Bellare and Sriram Keelveedhi [2] which provides secured deduplicated storage resisting bruteforce attacks, and realize it in a system called DupLESS (Duplicate less Encryption for Simple Storage). It provides more secure, easily-deployed solution for encryption that supports deduplication. In DupLESS, clients encrypt under message-based keys obtained from a key-server via an oblivious PRF protocol. Clients authenticate themselves to the Key-Server, but do not leak any information about their data to it. As long as the Key-Server remains in- accessible to attackers, we ensure high security. It enables clients to store the encrypted data and achieves strong confidentiality of data.DupLESS is more of feature compatible with the API commands in the system. The Deduplication subroutine enables fine grained control over the files and will be deduplicated, for exception the data in the personal files will not be deduplicated. Thus DupLESS provides security that is significantly better than current, convergent encryption based deduplicated encryption architectures. Deduplication is known to eoectively eliminate duplicates for Virtual Machine (VM) image storage which was explained in RevDedup: A Reverse Deduplication Storage System Optimized for Reads to Latest Backups by Chun-Ho Ng and Patrick P. C. Lee in June 28, 2013 [3]. Where it introduces fragmentation that will degrade read performance. RevDedup is introduced here, which is a deduplication system that optimizes reads to latest VM image backups is using an idea called reverse deduplication. RevDedup removes duplicates from old data, RevDedup achieves high deduplication eβciency with around 97% of saving, and high backup and read throughput on the order of 1GB/s. Many backup solutions are made by disk-spaced storage systems which has better I/O performance than other traditional storage systems. Deduplication is mainly studied in contentaddressable backup systems. It is also shown to provide spaceeβcient VM image storage given that VM images have signi¿cant content similarities. Here Deduplication mainly focuses on optimizing storage efficiency and performance. RevDedup exploits content similarities of VM images using a hybrid of inline and out-of-order deduplication approaches.

It applies coarse-grained global deduplication (inline) to dioerent VMs and removes any duplicates on the path, and further applies ¿ne-grained reverse deduplication (out-oforder) to dioerent backup versions of the same VM and removes any duplicates from old backup versions. Thresholdbased block removal mechanism is used, that combines holepunching to remove duplicate blocks of old backup versions and segment compaction to compact data segments without duplicate blocks to reclaim contiguous space. A novel encryption scheme that guarantees semantic security for unpopular data and provides weaker security and better storage and bandwidth benefits for popular data was introduced in A Secure Data Deduplication Scheme for Cloud Storage by Jan Stanek Alessandro Sorniotti, Elli Andreoulaki, and Lukas Kencl [4]. Data deduplication is made effective for popular data, whereas semantic security encryption protects the unpopular data. The effectiveness of storage efficiency functions such as compressions and deduplication is and objective for both storage provider and customer. Data deduplication proves that multiple uploads for the same content only consumes the network bandwidth and storage space for single upload. Deduplication is continuously used by many cloud providers as well as various cloud services to eliminate duplicated data. With the growing data size of cloud computing, a reduction in data volumes could help providers reducing the costs of running large storage system and saving energy consumption. So data deduplication techniques have been brought to improve storage efficiency in cloud storages which was explained in Dynamic Data Deduplication in Cloud Storage by Waraporn Leesakul, Paul Townend [5].With the dynamic nature of data in cloud storage, data usage in cloud changes overtime, some data chunks may be read frequently in period of time, but may not be used in another time period. A dynamic deduplication scheme for cloud storage, which aims to improve storage efficiency and maintaining redundancy for fault tolerance is used. Data deduplication is a technique whose objective is to improve storage efficiency. With the aim to reduce storage space, in traditional deduplication systems, duplicated data chunks identify and store only one replica of the data in storage. Logical pointers are created for other copies instead of storing redundant data. Deduplication can reduce both storage space and network bandwidth. Deduplication in cloud storages requires a dynamic scheme which has the ability to adapt to various access patterns and changes the user behavior in cloud storages. Cloud storage services commonly use deduplication technique for eliminating redundant data by storing only a single copy of data of each file of data block which was explained in Side channels in cloud services, the case of deduplication in cloud storage by Danny Harnik [6]. Privacy implication of cross-user deduplication is used in the system and it will be demonstrated as a side channel which reveals the information about the contents of the files of other users. High savings are offered by cross-user deduplication. Data deduplication strategies are categorized into two types they are 1) File-level deduplication 2) Block-level deduplication. In file-level deduplication only single copy of the file is stored and two or more files are identified as identical if they have the same hash value. In Block-level deduplication the data file is segmented into blocks and only single block will be stored. The system could use either fixed size block or variable sized blocks. The effectiveness of deduplication depends on multiple factors such the type of data, the retention period and the number of users. By applying this source based deduplication approach, client will be able to easily identify whether a certain file or block is deduplicated. This can be done by either examining the amount of data transferred over the network, or by observing the log of the storage software.

III. THE PROPOSED MODEL

E-Healthcare is an emerging field of medical informatics, referring to the delivery of health services and information using the Internet and related technologies. EHealthcare is the most important revolution in the healthcare society recently. E-healthcare system is now being globalized.

A. Overview of Model

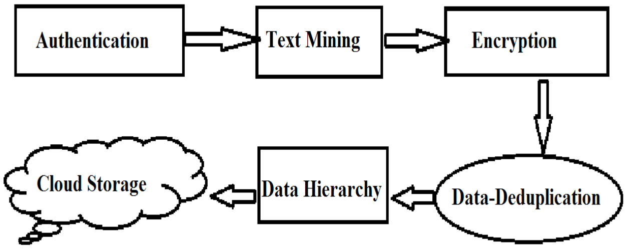

Figure 4.1 System Overview

B. *Authentication*

Indirectly authorized physicians and unauthorized persons cannot correctly distinguish the identities of the patients from each other. Only the physicians directly authorized by the patients can access the patients' personal health information and authenticate their identities simultaneously. The physicians and research staff indirectly authorized by patients cannot authenticate the patients' identities but recover the personal health information.

C. *Text Mining*

It is well known that the characteristics of one specific disease would vary as the health deteriorating status and/or the recovering status develops. With a certain dosage during a course of treatment, some specific vital signs such as the body temperature, the blood pressure, the leucocyte count and the blood platelet count possess their own regularities in each time period. Therefore, it is required to compare the dynamically collected personal health information (PHI) from the patient with the experience PHI template for one specific disease, each of which is represented by a vector of multiple elements representing the values of vital signs for each time period, to decide whether the patient's health condition is deteriorating or recovering.

D. *Encryption*

Attribute-based encryption (ABE) is a concept of public-key cryptography. In traditional public-key cryptography, a message is encrypted for a specific receiver using the receiver's public-key. Identity-based cryptography and in particular identity-based encryption (IBE) changed the traditional understanding of public-key cryptography by allowing the public-key to be an arbitrary string, e.g., the email address of the receiver. Attribute Based Encryption (ABE) goes one step further and defines the identity not atomic but as a set of attributes, e.g., roles, and messages can be encrypted with respect to subsets of attributes (key-policy ABE - KP-ABE) or policies defined over a set of attributes (cipher text-policy ABE - CP-ABE). The key issue is, that someone should only be able to decrypt a cipher text if the person holds a key for "matching attributes" (more below) where user keys are always issued by some trusted party.

E. *Data-Deduplication*

Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. Initially the name of the file is compared with the other available files in the cloud. After which the data inside the file is compared with the data that is already present. If either of which match with the new file, then the file will not be allowed to be saved in the cloud system. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication can take place at either the file level or the block level. For file level deduplication, it eliminates duplicate copies of the same file. Deduplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files.

F. *Cloud Storage*

Cloud computing is now the hot spot of computer business and research. Several trends are opening up the era of Cloud Computing, which is an Internet-based development and use of computer technology. Cloud Storage usually contains business-critical data and processes, hence high security is the only solution to retain strong trust relationship between the cloud users and cloud service providers. Thus to overcome the security threats, multiple cloud storage is enhanced. Thus the common forms of data storage such as files and databases of a specific user is split and stored in the various cloud storages (e.g. Cloud A and Cloud B).

G. *Data Hierarchy*

The layered model of access structure to solve the problem of multiple hierarchical files sharing is proposed. Layered model also improves the level of security at each layer and the data present will be highly authenticated. The files are encrypted with one integrated access structure. we also formally prove the security of FH-CP-ABE scheme that can successfully resist chosen plaintext attacks (CPA).An attending physician needs to access both the patient's name and his medical record in order to make a diagnosis, and medical researcher only needs to access some medical test results for academic purpose in the related area, where a doctor must be a medical researcher, and the converse is not necessarily true. Suppose that the patient sets the access structure of m1 as: T1 {("Cardiology" AND "Researcher") AND "Attending Physician"}. Similarly, m2 is termed as: T2 {"Cardiology" AND "Researcher"} the information needs to be encrypted twice if m1 and m2 are encrypted with access structures T1 and T2, respectively. The two structures could be integrated into one structure T. the computation complexity of encryption and storage overhead of cipher text can be reduced greatly.

H. Experimental Analysis

To protect patient data confidentiality, privacy preserving techniques are implemented to secure the PHI (Personal Health Information), and also to share the data to the admin. The layered model of access structure to solve the problem of multiple hierarchical files sharing is implemented in order to increase the efficiency of the system and also to improve the security constraints per layer of the structure. Admin plays a major role by handling all of the system data that is stored and retrieved every now and then. Hierarchical file sharing provides more security to the confidential information that is being stored in the system cloud. The files are encrypted with one integrated access structure which would reduce the encryption cost and increase the storage space. For every encrypted file a separate key is generated without which the file cannot be decrypted. The authorized duplicate check for this file can be performed by the user with the public cloud before uploading this file. The results of duplicate check is made by deduplication methodology which performs name check and also data check. Based on the results made from the data check the user uploads this file on the cloud or runs it directly on the system. The encrypted files will be uploaded into the cloud, if the user request match with the image then the file can be decrypted and downloaded. The private keys for privileges will not be issued to users directly, which will be kept and managed by the private cloud server instead. Image based authentication is enhanced. The Clinical Document Architecture (CDA) is described for CDA document generation and integration service based on cloud computing, through which hospitals are enabled to conveniently generate CDA documents without having to purchase proprietary software.

REFERENCES

[1] "Private Data Deduplication Protocols in Cloud Storage" Wee Keong Ng SCE, Yonggang Wen SCE, Huafei Zhu

[2] "DupLESS: Server-Aided Encryption for Deduplicated Storage" Mihir Bellare and Sriram Keelveedhi, University of California, San Diego; Thomas Ristenpart, University of Wisconsin—Madison.

[3] "RevDedup: A Reverse Deduplication Storage System Optimized for Reads to Latest Backups" Chun-Ho Ng and Patrick P. C. Lee The Chinese University of Hong Kong, Hong Kong Technical Report June 28, 2013.

[4] "A secure data deduplication scheme for cloud storage" Jan Stanek Alessandro Sorniotti, Elli Andreoulaki, Lukas Kencl.

[5] "Dynamic Data Deduplication in Cloud Storage" Waraporn Leesakul, Paul Townend, Jie Xu School of Computing University of Leeds, Leeds, LS2 9JT United Kingdom.

[6] "Side channels in cloud services, the case of deduplication in cloud storage" Danny Harnik IBM Haifa Research Lab Benny Pinkas Bar Ilan University Alexandra Shulman-Peleg IBM Haifa Research Lab

[7] "Memory Deduplication as a Threat to the Guest OS" Kuniyasu Suzaki, Kengo Iijima, Toshiki Yagi, Cyrille Artho National Institute of Advanced Industrial Science and Technology.

[8] "Hierarchical Attribute-Based Encryption for FineGrained Access Control in Cloud Storage Services"Guojun Wang, Qin Liu School of Information Science and Engineering Central South University Changsha, Hunan Province, P. R. China, 410083 Jie Wu Dept. of Computer and Information Sciences Temple University Philadelphia, PA 19122, USA.

[9] "Scalable and Secure Sharing of Personal Health Records in Cloud Computing using Attribute-based Encryption" Ming Li Member, IEEE, Shucheng Yu, Member, IEEE, Yao Zheng, Student Member, IEEE, Kui Ren, Senior Member, IEEE, and Wenjing Lou, Senior Member, IEEE.

[10] "Privacy Preserving EHR SystemUsing Attribute-based Infrastructure" Shivaramakrishnan Narayan, Martin Gagné and Reihaneh Safavi-Naini Department of Computer Science University of Calgary, Alberta, Canada {snarayan,mgagne,rei}@ucalgary.ca