# Security for Healthcare Data on Cloud

Kushan Shah
Dept. Of Computer Engineering, Nirma University, Ahmedabad, Gujarat
kushanmshah@yahoo.co.in

Vivek Prasad
Dept. Of Computer Engineering, Nirma University, Ahmedabad, Gujarat
vivek.prasad@nirmauni.ac.in

***Abstract*–A large amount of healthcare data is collected and stored online all around the world. This data can be used by doctors in prognosis, patients for their understanding, insurance companies, governments and many other institutions. Cloud acts as a platform where this data can be stored and can be accessed anywhere. It decreases the cost of processing, storing and transferring the data online to merchants. The data is stored in the form of EHR – Electronic Health Records which is available online anytime and can be updated as and when the patient undergoes any treatment of diagnosis. The report data along with the supporting data like x-ray images, scan images, patient private data and treatment procedure can be stored. The major issue that needs to be dealt with at this stage is the security of the above mentioned data. This data can be misused very easily and can cause harm to the individual if fallen victim to such schemes. Encrypting the data stored online is thus very important and the key for decrypting the data should only be made available to the doctors and concerned stakeholders. Contents of this paper mention various methods of encryption and also addresses security and privacy challenges in healthcare cloud by deploying a novel framework with CPRBAC (Cloud-based Privacy-aware Role Based Access Control) model[4].**

***Keywords* –** Encryption Techniques, Cloud, Privacy-aware, Control Model

## I. Introduction

Since Cloud computing serves the fundamental purpose of deploying services to the users around the world. Cloud service providers make virtual machine and processors available to the users to run their programs and systems on them. Cloud Service Providers (CSP) have increased in amount exponentially in the coming years. This has changed the way distributed system architecture is executed.

Most of the data can be warehoused offline on personal servers and data that needs to be accessed and changed from different places can be stored on the cloud. People leverage cloud to store data online, share their information, productive usage of services with fast access and low-cost of cloud on a remote server rather than invest in physical resources.

Even when the technology is moving forward and faster and coding ethics are being developed, data can still be stolen online which makes data security the most important paradigm. This paradigm addresses several issues concerning losing control of the process and systems if intrusion is experienced, data is stolen, illegal disclosure of malicious violation of intellectual property rights and protection of data ownership.

Deliverables of the system :

### *A. Electronic Health Records (EHR)*

Any clinic/doctor/hospital that a patient visits, patient information is recorded which includes the current health and vitals of the patient, previous treatments, ongoing medications and other important personal information. This patient health record is very important for the use of a doctor and hence is equally important that it be available online and immediately. It should also be updated and accessed as and when required. This helps in cases where the health record may be misplaced by the patient or when it is not available when required.

The EHR may also have multimedia files in them along with the text data which accounts for the scanned images of procedures, x-rays and other scan results. Any hospital that deals with such data processes thousands of requests for data and hence needs a large storage space, computation time and power.

### *B. Cloud Computing and Healthcare Industry[6]*

The main objective of this industry is to increase the reach of healthcare services to more and more number of people. These services are very crucial and can help every person who could find use of the service. Using cloud computing, the healthcare sector provides patient care by giving them a ubiquitous service at a higher performance and a lower cost. The more widespread the services becomes, more is the importance for it to be secure and fool proof. Certain privacy measures are thus taken by the system managers to make sure data preservation. Some of the privacy domains are mentioned in figure 1.

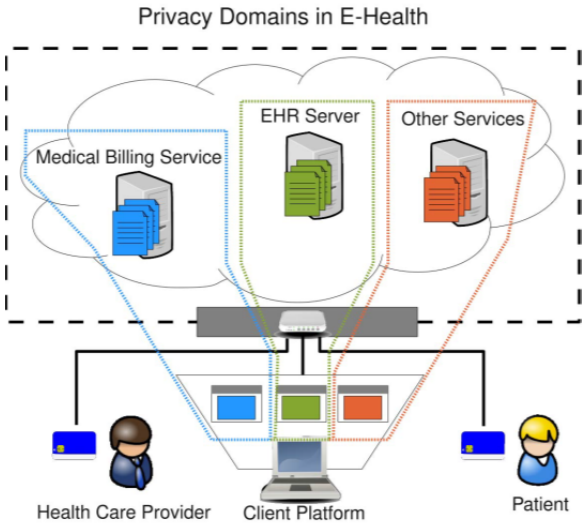


Figure.1. Privacy domains in E-Health

*C. Model for Data Preservation*

Ongoing research in this field have covered only some issues addressing the data security paradigm in cloud computing. Although these schemes are able to decrease the risk of malicious data modification attack, there are some drawbacks like relying on complex algorithms and infrastructure distribution services that pose unnecessary computation problems and increase communication overhead. Security procedure that is currently deployed consists of cryptographic keys that are assigns to the stakeholders only who have access to the data and the modules. To achieve a fine grained data protection scheme [5], certain abstractions like access condition, time constraint or data operation etc. are to be brought to use. A novel data protection model for the application of distributed health cloud scenario has been addressed here. The framework is based on Cloud-based Privacy-aware Role Based Access Control model (CPRBAC). This model is an extension of the existing Role Based (RBAC) model and modify some components including authorization delegation, resource sharing and sophisticated data protection techniques.

## II. Data Preservation concepts and strategies

*A. Securely Outsourcing Large-Scale System of Linear Equations*

This paper investigates converting large scale system processes on linear equation basis, which include the most used computational tools from various engineering disciplines to optimize real-world systems [5]. In this system there are three phases that are used to convert data into linear equations which are ProbTransform, ProbSolve and ResultVerify [1].

*B. Gradient-Descent for Privacy-Preservation*

This method is one of the most used methods in the field of privacy preservation. [2] A comparison between traditional Gradient-Descent method and Stochastic Gradient-Descent is carried out which shows that due to non linearly separated samples, the convergence rate is faster than the traditional method and thus is modified and better.

*C. Fully Homomorphic Encryption Using Ideal Lattices*

Fully homomorphic encryption is the method in which the whole process is undertaken without giving the user the access to it thus providing full data protection during the running time of the process[6]. This is achieved in 3 steps – Construction of an encryption scheme with decryption sequence, construction of public key using lattices, evaluation of the data without decrypting it.

*D. Communication Efficient Secure Linear Algebra[5]*

Linear Algebra plays a paramount role in cryptographic techniques. Solutions of efficient cryptographic algorithms is done through linear equation solving. Encryption and decryption technique results are formed using the same. Communication efficient and secure protocol was built in which two communicating entities study the differences in their input subspaces without displaying the result of their computations hence providing a level of abstraction.

*E. Paillier Cryptosystem Algorithm*

In this existing system [1], linear-equation outsourcing for a large-scale system is investigated. It shows how those equations can be securely solved over the infrastructure of a cloud.

The customer has a large scale linear equation problem Ax = b denoted as Φ = (A, b). Here the coefficient matrix A is a non-singular matrix and b is a right hand side vector.

An iterative method is used to solve the large set of linear equations which utilizes the additive homomorphic encryption scheme just like the Pailier cryptosystem algorithm. Successive approximate solutions to the linear equations can be found securely on cloud.

Then homomorphic encryption is used in encrypting the data in the cloud. This ensures data privacy. The result can be decrypted using the private key of the homomorphic encryption. Later when the result of the computation is retrieved from the cloud it can be verified using the secret key k. This ensures the authenticity of the cloud service provider.

The above mentioned systems still pose some disadvantages due to the complexity and impracticality of the algorithm to implement. To impose security on the Electronic Health Record the Pailier Cryptosystem scheme has been defined as in Figure 2.

The Paillier Cryptosystem is an asymmetric cryptosystem (public-private keypair) of the homomorphic encryption system. Homomorphic encryption preserves the structure of the input that is encrypted. Therefore, it can very well be applied on matrices.

In a similar manner, the same can be applied on matrices.

Steps in **Pailiier Cryptosystem Algorithm**

**Key generation phase:**

1. Two prime numbers p and q are randomly selected such that gcd( pq, (p-1)(q-1)) is equal to 1.
2. Compute n such that $n = p*q$.
3. Compute $\lambda$ such that $\lambda = lcm(p-1,q-1)$.
4. A random integer g is selected such that $g \in Z^*_{n^2}$.
5. Compute $\mu$ such that $\mu = (L(g^{\lambda} \bmod n^2))^{-1} \bmod n$ where $L(u) = (u-1)/n$.
6. The public key used for encryption is (n,g) and the private key used for decryption is $(\lambda,\mu)$.

**Encryption phase:**

1. m is the message to be encrypted where $m \in Z_n$.
2. A random integer r is selected such that $Z^*_n$.
3. Cipher text c is calculated as $c = g^m . r^n \bmod n^2$.

**Decryption phase:**

1. The cipher text is $c \in Z^*_{n^2}$.
2. The message m is got back using $m = L( c^{\lambda} \bmod n^2) . \mu \bmod n$.

Figure 2 : Pailier Cryptosystem Algorithm

Proposed system for Data encryption on cloud has been depicted in Figure 3. In this figure the different relationships and the communication diagram between the different stakeholders have been shown.

## III. CPRBAC model

To achieve a fine grained data protection scheme that is robust and secure at the same time, reduces computational complexity and communication overhead, this model is useful. In comparison to the other access control models, the proposed by L. Chen et al introduces 4 new features i.e *Organizations (Or), Conditions (Co), Obligations (Ob), purposes (Pu) [6].*

The said features would help in enriching the policy for a complicated set of usage requirements, authorization delegation, cross-realm role management, privacy-aware and active auditioning scheme [6].

The model is illustrated in figure 1. *Subject (S)* is an entity which accesses relevant Object. It can be a human being or a service application, and its attribute is used to determine a specific role. *Object (O)* represents any Information or data relating to the identified *S*, such as patient's electronic health record (EHR). *Role (R)* is a Functional entity associated some with specific authority and responsibility within an organization. For instance, the general practitioner role can access the EHR data of patients but has no permission to access their sexual life data or cancer history, but the surgeon role has both. *Operation (Op)* binds *O* and consists of a set of actions that a subject can execute (such as read or write privileges).

*Organization (Or)* is a domain identifier that defines R. In cloud environment, the set of *Ors* and *Rs* defines distributed role allocation. *Condition (Co)* is a prerequisite to be met before any *Op* can be executed, for example, EHR of patient A can be only disclosed to his/her specific practitioner when the current time is between 9AM and 5PM. *Purpose (Pu)* specifies the intended reason of the *Op*. One example is that a practitioner can access EHR of a patient only when the purpose is for emergency treatment. *Obligation (Ob)* is a function which must be executed before an *Op* is executed on *O* or after the execution. It usually connects the report service for auditing or logging process. These entities are the basic building blocks of our CPRBAC policy.

*Definition 1 :* The CPRBAC model is composed of the following components as in Figure 4 :

- Data Permission set is denoted as *DP is a subset of* $O \times Op \times Co \times Pu \times P(Ob)$ where $P(Ob)$ denotes the set of all subsets of *Ob*.
- Data Permission Assignment set is denoted as *DPA* is a subset of $R \times DP$ which is a many-to-many mapping relation of *R* to *DP*.

Figure 3: Definition 1

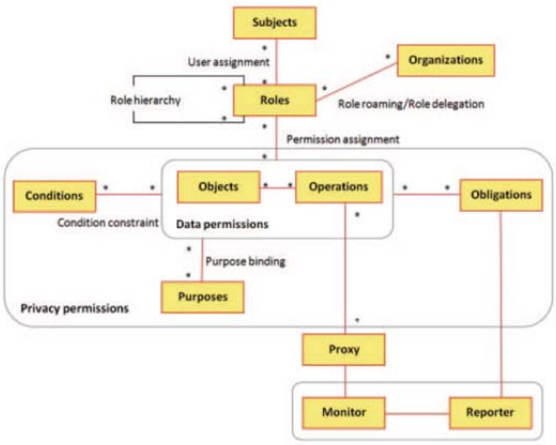


Figure 4: CPRBAC model

*Role assignment*is the procedure of authentication of user healthcare information and assigning the role of operations on data and making sure that it is valid. Attribute based role assignment (ABRA) is done in this model which indicates that the user is engaged in both the Subject and the Attribute Set [3]. ABRA is capable to satisfy dynamic role assignment based on context information rather than focusing on the relation between the subject and the role.

*Definition 2:* Attribute-Based Role Assignment. Let Subject be S which is defined as subject identifier, AttributeSet AS is a set of context information such as organization, domain, or time, etc., and the corresponding Role is R. Role assignment RA can be denoted as RA is a subset of $(S,AS)\times R$

Figure 5: Definition 2

*Role Hierarchy*reduces permission assignment costs [5]. Quite similar to object oriented programming, inheritance relationships roles are set. Role hierarchy is a partial order set i.e. a symmetric, reflexive and asymmetric relation.

*Definition 3:* Role Hierarchy

- Role Hierarchy $RH$ is a subset of $R \times R$, a partial order on roles
- If $R<R'$ then all subjects assigned to $R'$ are (implicitly) assigned to $R$
- If $R<R'$ then all permissions assigned to $R$ are (implicitly) assigned to $R'$
- If $R<=R'$ then implies that $P(R)$ *is a subset of* $P(R')$, $P(R)$ denotes the permission of $R$

Figure 6: Definition 3

*Role Delegation* is the process of delegating some work to other users in case some stringent conditions restrict the capabilities of that user. It is to be noted that the same can only be done in the same domain.

*Definition 4:* Role Delegation

Let original Subject be $OS$, delegated Subject be $DS$, the Role of OS is $R_{os}$, the Role of DS is $R_{ds}$, Condition C. The Role Delegation is expressed as $RD$ is a subset of $(OS,R_{os}) \times C \times (DS,R_{ds})$

Figure 7: Definition 4

*Role Roaming*in a multiple-domain healthcare environment on cloud refers to a role assignment process. A user in this organization can give access to another user by mirroring the role that he is assigned by using this scheme.

*Definition 5:* Role Roaming

Let $R_1$ be the role in Organization $O_1$, $R_2$ is the role or role set in Organization $O_2$, Role roaming scheme can be expressed as $RR$ is a subset of $(R_1,O_1) \times (R_2,O_2)$, which depicts $R_1$ in $O_1$can map to corresponding role or one of role set $R_2$ in $O_2$.

Figure 8: Definition 5

## IV. Outcomes of the Program

Even after reviewing the encryption algorithms and techniques we see that there is still a lot of computational time even while space complexity is saved by converting to bits. Time is also consumed while the result verification process. The time also changes with the change in key size. We also see from the performance charts that the performance of the enhanced algorithm is more than that of the standard algorithm. The security analysis and performance analysis of the CPRBAC model says that shows how data encryption is effective.

## V. Conclusion and Future Work

The techniques mentioned aim at storing a large amount of patient health data securely on cloud. The procedure is based on the conversion to pixel which makes the processing easier and decreases the amount of memory required to store at the same time makes it difficult for an intruder to understand the data. In this way the basic security of data on the basis of meaningful data being preserved is achieved. Time to time security analysis can be carried out to measure the level of abstraction that has been achieved.

A novel model for the same, the CPRBAC model was proposed for more controllability, traceability of data and authentication preservation to such resources online under a fine-grained data protection scheme. The encryption technique can be clubbed with the model to engage a state-of-the-art encrypted healthcare data management service.

Further, a system can also be developed for storing high level data like videos from ultrasound or other surgeries that could be used by the doctors, in a similar format so as to achieve a more dynamic data framework and more transparent working on the cloud.

## VI. References

[1] C.Wang, K. Ren, J.Wang, and Q. Wang, “Harnessing the cloud for securely outsourcing large-scale systems of linear equations,” IEEE Transactions on parallel and distributed systems, vol. 24, no. 6, pp. 1172-1181, 2013.

[2] S. Han, W. K. Ng, L. Wan, and V. C. Lee, “Privacy-preserving gradient-descent methods,” IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 6, pp. 884–899, 2010.

[3] Y. Shucheng, W. Cong, R. Kui, and L. Wenjing, "Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing" in INFOCOM, 2010 Proceedings IEEE, pp. 1-9, 2010.

[4] www.uts.edu.au

[5] Aiswarya, R., et al. "Harnessing healthcare data security in cloud." Recent Trends in Information Technology (ICRTIT), 2013 International Conference on.IEEE, 2013.

[6] Chen, Lingfeng, and Doan B. Hoang. "Novel data protection model in healthcare cloud." High Performance Computing and Communications (HPCC), 2011 IEEE 13th International Conference on.IEEE, 2011.