Towards Healthcare Data Security inCloud Computing

Huda Elmogazy[1], Omaima Bamasak[2]

[1]*Department of Computer Science,Faculty of Computing and Information Technology, King AbdulAzizUniversity, Jeddah, Saudi Arabia, helmogazy@kau.edu.sa*
[2]*Department of Computer Science,Faculty of Computing and Information Technology , King AbdulAziz University, Jeddah, Saudi Arabia, obamasek@kau.edu.sa*

Abstract

Healthcare data has stringent security requirements for confidentiality, availability to authorized users, and traceability of access. The focus of this study is to investigate on these requirements and propose a solution for healthcare cloud providers that will help in protecting patient' data they host, which is of high importance. The focus will be on specific cloud computing healthcare security concerns and how cloud homomorphic encryption with splitting key and key delegation can help in meeting healthcare regulatory requirements. The suggested technique is based on FHE algorithm with key delegation to ensure data confidentiality, authentication, integrity, and availability in a multi-level hierarchical order. This will enable the healthcare provider to apply/omit any access rule in any order, especially in medical research environment.

Index Terms: *Cloud Computing; Cryptography; Threats; Homomorphic encryption.*

-- *** --

1. INTRODUCTION

Cloud Computing represents one of the most significant shifts in information technology that have been witnessed recently. Cloud computing offers potential benefits including cost savings and improved business outcomes.Sun Microsystems calls cloud computing the next generation of network computing. Google describes the cloud as "the collective power of thousands of computers that serve information to you from far-away rooms distributed around the world."

In modern healthcare environments, there is a strong need to create an infrastructure that reduces time-consuming efforts and costly operations to obtain a patient's complete medical record and uniformly integrates this heterogeneous collection of medical data to deliver it to the healthcare professionals. Electronic health records (EHRs) [1], [2]have been widely adopted to enable healthcare providers, insurance companies and patients to create, manage and access patients' healthcare information from anywhere, and at any time.

As a result, healthcare providers are more willing to shift their systems to clouds that can remove the geographical distance barriers among providers and patients. With cloud computing, different doctors can access a patient's health records even if they're miles apart. These physicians need not talk over the phone to request for a transfer of the health records. They will just have to access them in the clouds. Data privacy and security are two of the main reasons why healthcare takes the slow route towards the adoption of these new technologies. Healthcare data security has been around for a long time, but as cloud computing gains more and more attention, healthcare providers are aiming at utilizing cloud's advantages to their benefit. However, these advantages come at a cost of various information security risks that need to be carefully considered. Risks vary depending on the sensitivity of the data to be stored or processed, and how the chosen cloud vendor has implemented their specific cloud services.

In order to be attractive to healthcare environment, cloud computing should provide safeguards necessary to satisfy HIPAA(Health Information Portability and Accountability Act) and other privacy and security requirements. Although there exist healthcare regulations for Electronic Health Records (EHRs), such as HIPAA which is recently amended to incorporate business associates in 2009 [3], several cloud providers are not covered entities by them.

The first step to secure healthcare data is to classify the information of the Electronic Health Records (EHRs)according to its security sensitivity. The first type is Personally Identifiable Information (PII), such as patient records, which is often stored in a relational database as structured data. The second type is Healthcare data, which is mainly comprised of large media files such as x-ray, radiology, CT scans, and other types of video and images that does not reveal patient's identity. Such files are often stored in distributed storage.

To put everything online “in the cloud," unencrypted, is a big risk. For certain types of data, such as Electronic Health Records (EHRs), storing them off-site unencrypted may be illegal.However, to process data located on a remote server, the Cloud providers need to access the raw data, i.e. not encrypted. Most people do not fully entrust the Cloud service providers for their sensitive healthcaredata because there is no governance about how this information can be used by them and whether the patients actually control their information. On the other hand, encrypting one's data seems to nullify the benefits of cloud computing, unless I give the cloud my secret decryption key.Ordinary cryptography is no help in this situation. Suppose that healthcare providers want to delegate the ability to process their data, without giving away access to it. The technique that makes this magic trick possible is called fully homomorphic encryption, or FHE. Fully homomorphic encryption scheme has been suggested in the literature for data privacy provision, which allows a Cloud provider who does not have the secret decryption key to perform computation on data (still encrypted) and produce results, even when the function of the data is very complex. This means that Cloud provider can perform complicated processing on data without being able to see it.

There are fields in a medical record that are considered by both individuals and legislation (such as HIPAA) as highly sensitive and private and should not be made available to all that have access to theElectronic Health Records (EHRs), legitimate or otherwise. However in medical research environment, some research agents need the healthcare data for their medical-related research. This case demonstrates that Electronic Health Records (EHRs) systems must not employbulk encryption to protect sensitive information stored in one's medical record from unauthorized or malicious access.Moreover, Access control privileges must be in place to guarantee that access to sensitive information is limited only to those entities that have a legitimate need-to-know privilege allowed.

The rest of this paper is organized as follows: in Section II, weare introducing the concept ofcloud computing. In Section III, we present an overview about security issues in cloud computing. In Section IV,we discuss background technologies about using cryptography in cloud computing. In section V, we outline the proposed healthcare cloud computing framework. Finally, Section VI concludes this paper and discusses our future direction.

2. BACKGROUND

Cloud computing, as a model for IT services, is defined by the National Institute of Standards and Technology (NIST) [4]as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”.
The cloud model is composed of five essential characteristics**,** three service models, and four deployment models.

2.1 Essential Characteristics

On-demand self-service.A consumer can unilaterally be provided with computing capabilities, such as server time, network storage or user email accounts, as needed automatically without requiring human interaction with each service provider.
Broad network access.Capabilities are available over the network and accessed through standard mechanisms that promote usage by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).
Resource pooling.The provider’s computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, and network bandwidth.

Rapid elasticity.Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

Measured service.Cloud systems automatically control and optimize resource use by leveraging a metering capabilityat some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

2.2 Deployment Models

Depending on infrastructure ownership, there are four deployment models of cloud computing each with its merits and demerits. This is where the security issues appear.

Private cloud.The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

Community cloud.The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

Public cloud.The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

Hybrid cloud. The cloud infrastructure is a composition of two or more distinct cloud infrastructures (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).

2.3 Service Models

Software as a Service (SaaS).The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings. Example SaaS vendor services include Salesforce.com Customer Relationship Management (CRM), Google Docs and Google Gmail. Microsoft Office 365 (formerly called Business Productivity Online Suite) consists of Microsoft Office Web Apps, Microsoft Exchange Online, Microsoft SharePoint Online, Microsoft Dynamics CRM Online and Microsoft Lync.

Platform as a Service (PaaS). The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider . The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, rather has control over the deployed applications and possibly configuration settings for the application-hosting environment. Example PaaS vendor services include Google App Engine, Force.com, Amazon Web Services Elastic Beanstalk, and the Microsoft Windows Azure platform.

Infrastructure as a Service (IaaS).The capability provided to the consumer is to provide processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls). Example IaaS vendor services include Amazon Elastic Compute Cloud (EC2), GoGrid and Rackspace Cloud.

3. SECURITY ISSUES IN CLOUD COMPUTING

Despite the enormous potential and rapid growth, privacy, security and trust for cloud remain areas of concern and uncertainty, and the resulted risks are needed to be better understood. A risk management process must be used to balance the benefits of cloud computing with the associated security risks. Cloud Security

Alliance" (CSA) provides needed context to assist organizations in making properrisk management decisions regarding their cloud adoption strategies. CSA identifies the following top seven threats:[5]

- Abuse and Nefarious Use of Cloud Computing
- Insecure Application Programming Interfaces
- Malicious Insiders
- Shared Technology Vulnerabilities
- Data Loss/Leakage
- Account, Service & Traffic Hijacking
- Unknown Risk Profile

CSA also identifies the following security service categories as the most interesting to experienced industry consumers and security professionals:[6]

- Identity and Access Management (IAM)
- Data Loss Prevention (DLP)
- Web Security
- Email Security
- Security Assessments
- Intrusion Management
- Security Information and Event Management (SIEM)
- Encryption
- Business Continuity and Disaster Recovery
- Network Security

Our research will focus on the use of encryption to defeat threats addressed as:

- Failure to meet Regulatory Compliance requirements
- Mitigating insider and external threats to data
- Intercepted clear text network traffic
- Clear text data on stolen / disposed of hardware
- the risk or and potentially enabling cross-border business opportunities
- perceived risks and thus enabling Cloud's Adoption by government

4. LITERATURE REVIEW

4.1 Modern Cryptography

Modern Encryption algorithms (such as RC4, RC6, MARS, AES, DES, 3DES, Two-Fish, and Blow-Fish) still play the main role in data security of cloud computing. The evaluation has been performed for those encryption algorithms according to randomness testing by using NIST statistical testing in cloud computing environment (Amazon EC2) [7]. From simulation results, the authors concluded that no strong indications of statistical weaknesses for the eight modern encryption algorithms when applied in cloud computing environments.

A hybrid encryption technique (uses RSA, 3-DES and Random Number generator algorithm) is suggested to enhance the security of cloud database [8]. This technique provides the flexibility in range and sequence to the user's choice. This is because a user can apply all of the three encryption methods or omit any in any order. Even if the user does not select any encryption technique, the random number algorithm will still be implemented by default, thus providing at least a single level security. The opted sequence will also be stored in the database so that the decryption can be possible. The negative effect of this scheme is that it creates an overhead on the query performance due to the multilevel nature of encryption and decryption. Also the computation time increases as the size of data increases.

Elliptic Curve Cryptographywas proposed to explore data security (confidentiality and authentication of data) between clouds [9]. Elliptic curve cryptography [ECC] is a public-key cryptosystem in whichevery user has a public and a private key. Public key is used for encryption/signature verification. Private Key is used for decryption/signature generation. Elliptic curves are used as an extension to other current cryptosystems, i.e. Elliptic Curve Diffie-Hellman Key Exchange and Elliptic Curve Digital Signature Algorithm.

4.2 Homomorphic cryptography

The fully homomorphic encryption, or FHE is not a new idea as it has been, for many years, viewed as a fantasy that would never come true. Rivest, Adleman, and Dertouzos [10] suggested that fully homomorphic encryption may be possible in 1978, shortly after the invention of the RSA cryptosystem [11], but were unable to find a secure scheme for its realization.Thiswas changed in 2009, with a breakthrough discovery by Craig Gentry [12, 13], who was then a graduate student at Stanford University. (He is now at IBM Research.) Since then, further refinements and more new ideas have been coming at a rapid pace[14].

Before trying to explain how homomorphicencryption works, we should explain the word homomorphic [15]. TheGreek roots translate as same shape orsame form, and the underlying idea isthat of a transformation that has thesame effect on two different sets ofobjects. The concept comes from theesoteric world of abstract algebra, butwe can offer a more homely example,where the two sets of objects are thepositive real numbers on the one handand their logarithms on the other. Thenmultiplication of real numbers andaddition of logarithms are homomorphicoperations. For any positive realnumbers x, y and z, if x·y = z, thenlog(x)+log(y) = log(z). This homomorphismoffers two alternative routes tothe same destination. If we are given xand y, we can multiply them directly; orwe can take their logarithms, then add,and finally take the antilog of the result.In either case, we wind up with z. Homomorphic cryptography offersa similar pair of pathways. We can doarithmetic directly on the plaintext inputsx and y. Or we can encrypt x andy, apply a series of operations to theciphertext values, then decrypt the resultto arrive at the same final answer.

In encrypted computation, the user specifies encrypted inputs to a program, and the server computes on encrypted inputs to produce an encrypted result. This encrypted result is sent back to the user who decrypts it to get the actual result.

The fully homomorphic encryption (FHE) scheme, means that there are no limitations on what manipulations can be performed[16] . The fully homomorphic encryption (FHE) scheme allows a worker that does not have the secret decryption key to compute any result of the data (still encrypted), even when the function of the data is very complex.

Homomorphic encryption scheme, based on Residue Number System (RNS), was proposed to enhance the security [17]. In HORNS scheme, a secret is split into multiple shares on whichcomputations can be performed independently.Efficiency is achieved through the use of smaller shares. HORNS scheme depends on the RNS property that creates multiple shares of a data and the operations on these shares are homomorphic. Security is enhanced by not allowing the independent clouds to collude.

4.3 Searchable Encryption

Searchable encryption is a broad concept that deals with searches in encrypted data .The goal is to outsource encrypted data and be able to conditionally retrieve or query data without having to decrypt all the data. There are two approaches to the searchable encryption. The first approach is to use symmetric encryption [18,19], whereas thesecond approach for is to use asymmetric encryption [20, 21, 22].

4.4 Attribute Based Encryption

In the Attribute Based Encryption ABE [23], the attributes and policies associated with the message and the user decide which user can decrypt acipher text. A central authority will create secret keys for the users based on attributes/policies for each user.Users in the system have attributes; users receives a key ("or key bundle") from an authority for their set of attributes.Cipher text contains a policy (a Boolean predicate over the attribute space). If a user's attribute set satisfies thepolicy, he can use his key bundle to decrypt the cipher text. Multiple users cannot pool their attributes together.

5. HEALTHCARE DATA SECURITY IN THE CLOUD

The healthcare industry is shifting toward an information-centric care delivery model, enabled in part by open standards that support cooperation, collaborative workflows and information sharing. Services delivered by cloud computing will evolve to support a wide variety of healthcare processes. Cloud computing provides an infrastructure that allows hospitals, medical practices, insurance companies, and research agents to tap improved computing resources at lower initial capital outlays. Healthcare IT managers understand that cloud computing has its potentials, but many are concerned about privacy and security issues and are delaying plans to move their critical patient data onto cloud-based systems. In this section, we will focus on specific cloud computing healthcare security concerns. Patient privacy is considered as the most serious problem in cloud computing and encryption is thought to be the sole solution. However, several questions need to be answered to address this issue:

1. How to make processing over encrypted data?
2. How to search over encrypted data?
3. How to grant partial access?

Our scenario willdescribe a proposed centralized secure data sharing framework for healthcare cloud-based EHR and, hence, addressesthe above mentioned questions. Let us assume that we have three basic organizations A, B and C in our healthcare industry, as shown in figure 1.

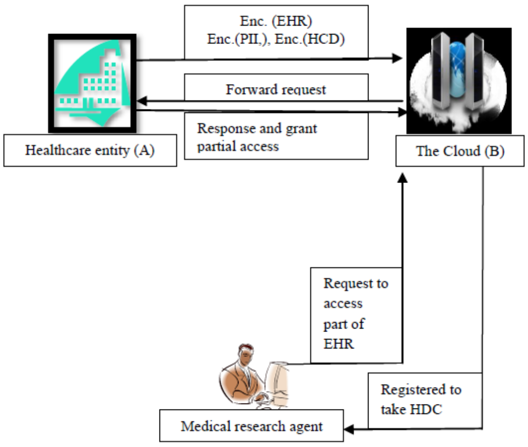

Figure1: The centralized secure data sharing framework for cloud

Healthcare entity (A)from various domains such as primary care, pharmacy, clinic lab and emergency care hosts their EMRs in cloud (B) to achieve lower operation cost, higher interoperability, and ubiquitous service delivery and so on.The medical research agent (C)needs some healthcare data for its medical-related research.The healthcare entity(A)is the owners of EHRs who specify access control policies to control who can access which portions of their EHRs, with various roles. To satisfy HIPAA and other privacy and security requirements,the medical research agent C should not access Personally Identifiable Information (PII). Administrators perform administrative functions such as activating or deactivating users, and registering or de-registering medical research agent.

The proposed solution is the using of homomorphism cryptographywith Attribute Based Encryption. In this scenario, we suggest thatEHR is stored as a Hierarchical record; each part of patient record (i.e. Personally Identifiable Information and Healthcare data) is encrypted separately using different keys. The index of EHR should not leak any sensitive information of a patient. Furthermore, the index should be effective and efficient to speed up the search of various types of EHR records and easy to expand when new records are added.

However, data encryption is only one part of the equation. The next challenging issue healthcare providers are facing is the issue of the encryption keys, and how to effectively and securely manage encryption keys in the cloud without sacrificing patients' privacy and meet regulatory compliance. Best practice for an effective and secure cloud key management is split-key encryption. A secret is split into multiple shares on which computations can be performed independently. It allows healthcare providers to manage encryption keys in the cloud, yet at the same time to split the encryption key, so customers (for example a hospital using medical applications hosted in the cloud) are the only ones who control their "half key", and therefore patient data is never visible to the cloud provider.

Now we will describe the general flow in our usage scenario,as shown in figure 1.The setup step includes that all EHRs are encrypted using a combination levels of FHE,ABEand index search algorithm before going to the cloud B.Only authorized entities can obtain the access to the authorized portions of the encrypted EHR data through identification and authorization. User will login into the system through web application. The user identity will be checked against login database system, which will verify on the basis of attribute authentication system.The medical research agent C will request to access some parts of EHRs to investigate their research. This request is forward to healthcare entity to gain access privileges to certain healthcare data for certain time. Then this access will be revoked.

6. CONCLUSION

Researches have shown that FHE is possible;however, there is still work to be done toward making FHE trulypractical. For future work, we intend to further investigate and implement the scheme in the context.As the proposed scheme is in stage of development hence actual results will be shared in future publication.

REFERENCES

[1] C. DesRoches, E. Campbell,S. Rao, K. Donelan,T. Ferris, A. Jha, R. Kaushal, D. Levy,S. Rosenbaum,A. Shields,"Electronic health records in ambulatory care - a national survey of physicians", New England Journal of Medicine, 359(1), pp. 50-60, 2008

[2] M. Eichelberg, T. Aden, J. Riesmeier, A. Dogac, G. Laleci, "A survey and analysis of electronic healthcare record standards", ACM Computing Surveys (CSUR), 37(4), pp. 277-315, 2005.

[3] The health insurance portability and accountability act. http://www.hhs.gov/ocr/privacy/hipaa/understanding/index.html, last accessed in 21/8/2013

[4] P. Mell and T.Grance, "The NIST Definition of Cloud Computing", October 7, 2009, version 15, National Institute of Standards and Technology (NIST), www.csrc.nist.gov.

[5] “Security Guidance for Critical Areas of Focus in Cloud Computing”, April 2009, presented by Cloud Security Alliance (CSA).(www.cloudsecurityalliance.org/guidance)

[6] Cloud Security Alliance (CSA), "Cloud security alliance secaas defined categories of service", 2011, (https://cloudsecurityalliance.org/wp-content/uploads/2011/09/SecaaS_V1_0.pdf)

[7]S. El-etriby, E. M. Mohamed, H. S. Abdul-kader, “Modern Encryption Techniques for Cloud Computing”, ICCIT, 2012.

[8]A.Kaur, M. Bhardwaj, “Hybrid Encryption for Cloud Database Security”, IJESAT, Vol-2, Issue-3, pp. 737 – 741, May-Jun 2012.(http://www.ijesat.org)

[9] V. Gampala, S. Inuganti, S.Muppidi, "Data Security in Cloud Computing with Elliptic Curve Cryptography”, IJSCE, Vol. 2, Issue 3, ISSN: 2231-2307, July 2012.

[10] R. L. Rivest, L. M. Adleman, and M. L. Dertouzos. On data banks and privacy homomorphisms.In Foundations of Sec. Comp., pp.169-180, 1978.

[11] R. L. Rivest, A. Shamir, and L. M. Adleman. A method for obtaining digital signatures and public-key cryptosystems.Commun. ACM, 21(2):pp.120-126, 1978.

[12]C. Gentry. A fully homomorphic encryption scheme.PhD thesis, Stanford University, 2009, www.crypto.stanford.edu/craig.

[13] C. Gentry,"Fully homomorphic encryption using ideal lattices", In M. Mitzenmacher, editor, STOC, pp. 169- 178. ACM, 2009.

[14] M. Tebaa, S. El hajji, Abdellatif El ghazi, “Homomorphic Encryption Applied to the Cloud Computing Security” , World Congress on Engineering (WCE) Vol. 1, July 2012.

[15] B. Hayes, ”Alice and Bob in Cipherspace”, American Scientist, Vol. 100, pp. 362-367, Sep–Oct 2012, www.americanscientist.org

[16] C. Gentry, "Computing Arbitrary Functions of Encrypted Data", ACM, Vol. 53 Issue 3, pp. 97-105, March 2010

[17] M. Gomathisankaran , A. Tyagi , K.Namuduri “HORNS: A Homomorphic Encryption Scheme for Cloud Computing using Residue Number System”, CISS, March 2011.

[18] D.X. Song, D. Wagner, A. Perrig,"Practical techniques for searches on encrypted data", In Proceedings of 21st Symp. on Security and Privacy (S&P), Berkeley, California, May 2000, pp. 44–55. IEEE Computer Society, Los Alamitos, 2000.

[19]R. Curtmola, J. A. Garay, S. Kamara, R. Ostrovsky,"Searchable symmetric encryption: improved definitions and efficient constructions",In Proceedings of 13th ACM Conference on Computer and Communications Security (CCS '06) , pp. 79–88, 2006.

[20] D. Boneh, G. Di Crescenzo, R. Ostrovsky, G. Persian, "Public key encryption with keyword search", In EUROCRYPT 2004. LNCS 3027, pp. 506–522. Springer, Heidelberg, 2004.

[21] G.D.Crescenzo, V. Saraswat,"Public key encryption with searchable keywords based on Jacobi symbols", In Proceedings of the 8th International Conference on Progress in Cryptology (INDOCRYPT'07), Springer-Verlag, Berlin, Heidelberg, pp. 282-296, 2007.

[22] H.S. Rhee, J. H. Park, W. Susilo, D. H. Lee, "Improved searchable public key encryption with designated tester". In ASIACCS, pp. 376–379. ACM, New York, 2009.

[23] S. Kaur, "Cryptography and Encryption In Cloud Computing", VSRD International Journal of CS & IT, Vol. 2 (3), pp. 242-249, 2012.