2015

Health Data in Cloud Environments

Doan Hoang
University of Technology Sydney, Doan.Hoang@uts.edu.au

Dang Thanh Dat
University of Technology Sydney, Thanh.D.Dang@student.uts.edu.au

HEALTH DATA IN CLOUD ENVIRONMENTS

Doan Hoang, School of Computing and Communications, University of Technology Sydney, Australia, Doan.Hoang@uts.edu.au

Dang Thanh Dat, School of Computing and Communications, University of Technology Sydney, Australia, Thanh.D.Dang@student.uts.edu.au

Abstract

The process of provisioning healthcare involves massive healthcare data which exists in different forms on disparate data sources and in different formats. Consequently, health information systems encounter interoperability problems at many levels. Integrating these disparate systems requires the support at all levels of a very expensive infrastructures. Cloud computing dramatically reduces the expense and complexity of managing IT systems. Business customers do not need to invest in their own costly IT infrastructure, but can delegate and deploy their services effectively to Cloud vendors and service providers. It is inevitable that electronic health records (EHRs) and healthcare-related services will be deployed on cloud platforms to reduce the cost and complexity of handling and integrating medical records while improving efficiency and accuracy. The paper presents a review of EHR including definitions, EHR file formats, structures leading to the discussion of interoperability and security issues. The paper also presents challenges that have to be addressed for realizing Cloud-based healthcare systems: data protection and big health data management. Finally, the paper presents an active data model for housing and protecting EHRs in a Cloud environment.

Keywords: EHR, Big health data, EHR format, EHR security, EHR standard, EHR interoperability.

1 INTRODUCTION

Electronic health record (EHR) stores digitally healthcare information about an individual's lifetime with the purpose of supporting continuity of care, education and research, and ensuring confidentiality at all times (Khudair 2008). EHR can be integrated in the Health Information Systems (HIS) which allow patients to manage and control their health records through the internet. This makes it possible for them to easily access their health data and share them with health care providers, insurance practitioners, researchers and family members. The health data in an EHR can be considered as a complete and accurate summary of an individual's medical history and health status. EHR has many functions and includes many types of data. The amount and quality of information available to health care professionals in the data has an impact both on the outcomes of patient care and the continuity of care.

A patient often receives medical treatment from different health professionals in various organizations over his/her lifetime. It is important to share this information to increase the quality of care and to decrease its cost. Therefore, providing the interoperability of electronic healthcare records (EHRs) is on the agenda of many national and regional initiatives.

Currently, EHR information is stored in various proprietary formats ranging from unstructured data to structured data. Typical formats include structured document-based storage such as relational database, XML and unstructured file types such as text, image, video, and sensor data. Furthermore, the data may not conform to an open standard. These result in severe interoperability problems associated with different EHRs. Furthermore, privacy and security issues also present challenges for HIS providers since EHR is deployed in multi-user environment and sensitive information could be exposed to the public by criminal attacks or accidental information breaches.

Integrating these disparate systems requires the support at all levels of a very expensive infrastructures. Cloud computing dramatically reduces the expense and complexity of managing IT systems (Schubert & Jeffery 2012). An increasing number of business customers are shifting their services and applications to Cloud computing since they do not need to invest in their own costly IT infrastructure, but can delegate and deploy their services effectively to Cloud vendors and service providers. It is inevitable that electronic health records (EHRs) and healthcare-related services will be deployed on cloud platforms to reduce the cost and complexity of handling medical records while improving efficiency and accuracy.

This paper presents a review of EHR including definitions, EHR file formats, and structures. The paper discusses interoperability and security issues related to EHR and EHR systems. The paper also presents challenges that have to be addressed for realizing Cloud-based healthcare systems, and finally it introduces an active data model for encapsulating EHRs with the aim of housing and protecting them in a Cloud environment.

The rest of the paper is organized as follows, section II provides and discusses various EHR definitions. Section III presents a comprehensive review of current structured and unstructured data EHRs. Issues pertaining to EHR are discussed in section IV. Section V discusses challenges for future research including Cloud-based health systems. Section VI presents an active data cube model and section VII summarizes the contribution of the paper.

2 DEFINITIONS AND CONTENTS AND STRUCTURES

2.1 EHR definition

The Health Information and Management Systems Society (HIMSS) (HIMSS) defines an electronic health record as, "a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics,

progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports. The Institute of Medicine (IoM), USA, states that an essential technology for health care is the Computer based Patient Record (CPR). According to the IoM, "A computer-based patient record (CPR) is an electronic patient record that resides in a system specifically designed to support users by providing accessibility to complete and accurate data, alerts, reminders, clinical decision support systems, links to medical knowledge, and other aids." (Roberts 1999). The International Organization for Standardization (ISO) defines it as, "a repository of patient data in digital form, stored and exchanged securely, and accessible by multiple authorized users. It contains retrospective, concurrent, and prospective information and its primary purpose is to support continuing, efficient and quality integrated healthcare (Häyrinen et al. 2008).

Explicitly, a useful EHR comprises of five essential components: personal information, medical history reports, health examination, nursing data, and administrative medical and health service records.

- **Personal information** presents the patients' general demographic information such as Patient name, Previously registered name/maiden name, Individual identifier/medical record number, Universal patient health number, Gender, Race, Address, Telephone number, Date of birth, Organization, Admission date, Discharge date, Legal authenticator, Authentication date, Transcriptionist/data enterer and Transcription date.
- **Medical history reports** present a whole picture about a patient's medical history including past medical history, medications, social history, procedure history and allergies.
- **Health examination** includes physical examination, diagnostic findings, assessment and plan, and Operative Report.
- **Nursing data** contains daily charting, physical assessments, and admission nursing notes. Daily charting includes patients' daily functional activities such as vital signs, food, elimination, mobility and patient teaching. Physical assessment comprises all kinds of status assessments. Admission nursing note contains information on allergies, health behavior, physical assessment, discharge planning and initial care plan.
- **Administrative medical and health service records** include medical administration information such as admissions records, referrals records, consultation records and bill payer records.

2.2 EHR Structures

EHR structures vary greatly since data is recorded in HIS by different groups of health care professionals together with the different legislation, standards and health practices of different countries, states and health care facilities. Xu et al. (2009) used XML to presents an EHR data structure which is hierarchical tree structure. The EPR data represented in tree structure can include patient's basic information, hospital entering records, records of the first disease course, surgical records and records of discharged from hospital (Figure 1). The record starts with the root node which could have one or more sub-node, and each sub-node serves as one case that presents for a component. Hierarchical tree structure is desirable and flexible in defining structures since it can be described by XML stored by XML database, and allows fast retrieval. Lj et al. (2000) also uses XML to present EHR structure at two levels as shown in Figure 2. The GEHR Object Model (or GOM) defines the data structures representing the electronic health record.

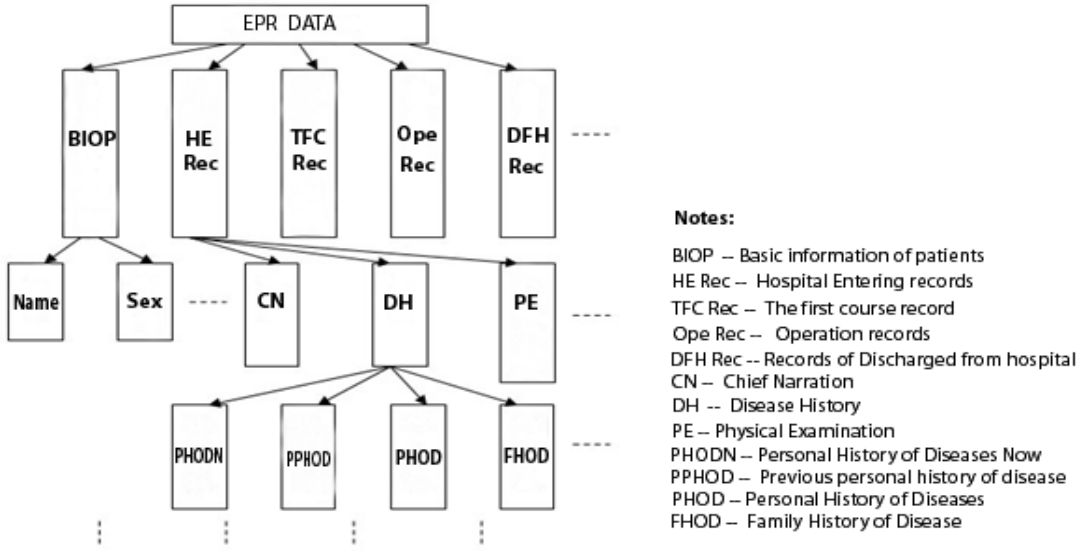

Figure 1. Data Structure of Electronic Patient Record (EPR) (Xu et al. 2009)

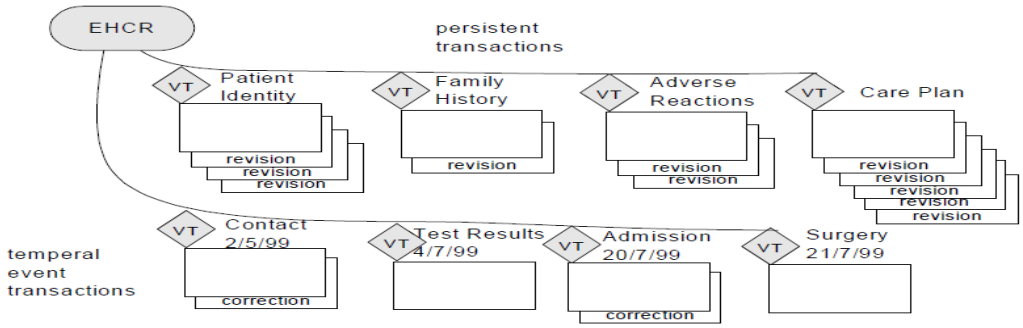

Figure 2. Versioned transaction view of the EHR (Lj et al. 2000)

At the root, EHR is a container to store a collection of transactions. Persistent transactions contain information which remains valid in the long term, such as family history, chronic conditions etc. Event transactions are used for information whose validity is relatively short-lived, such as the test results, a contact with a health care professional or a hospital admission.

3 STRUCTURES AND FORMATS

Currently, EHR information is stored in various proprietary formats. Typical formats include structured document-based storage in various formats such as database tables and structure XML files, and unstructured document storage in text, image, or video formats and digitized hardcopies maintained in a typical classical document management system. All possible formats of EHR, however, can be categorized in unstructured data and structured data as shown in Figure 3.

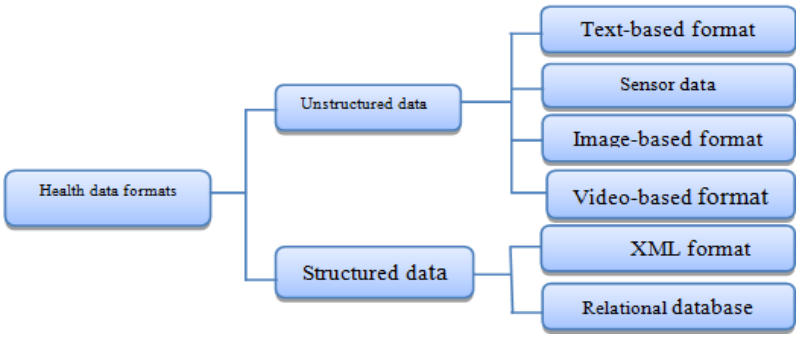

Figure 3. Categories of health data formats

3.1 Unstructured data

Unstructured EHR is an electronic health record which does not follow any patterns and structures. It allows users to store health-related information, such as case histories, health examination results, daily records and so forth in unstructured forms such as text, images, sensor data, etc. Using Unstructured EHRs offers users some benefits. Firstly, it supports variety of file formats which make it possible to store different types of health data such as text, images and video. Secondly, it enables doctors collecting information while interacting with patients. By using special notes, they can write the notes and detail records when they are diagnosing patients.

3.1.1 Text-based electronic health record

Unstructured data is commonly used in storing data collected from sensor networks such as text, log files. Moreover, there are many unstructured health data such as radiology reports, discharge summaries which are still stored in free-text format without any formal convention. Clinical narratives, such as radiology and pathology reports, are a growing electronically available source of information. Clinical texts are commonly dictated and transcribed by a person or speech recognition software, or are directly entered in text form by physicians.

3.1.2 Image-based electronic health record

Image-based EHR systems can deal with a wide array of medical data types including multi-dimensional medical images personal health records (PHRs) base on web such as PET-CT, MRI, Xray, medical videos, text-based data and metadata, and supplementary notes. Image-based electronic health record could store a collection of CT or MRT image which is useful for radiologists and other specialists such as surgeons. Imaging report that is widely needed and whose broad accessibility is vital for achieving optimal patient care. However, each image format defines its own unique metadata format. Existing image metadata fields are limited and not extensible to individual needs.

3.1.3 Video-based electronic health record

A Dietary Data Recording System (DDRS) which consists of an electronic data collection device and the software and protocols necessary to support data capture and calculation of nutrient intake was proposed in (Junqing et al. 2011). A video camera and a laser-generated grid of distances to food surfaces are employed for calculating the food intake. DDRS includes three components namely the mobile application, the web service, and the volume measurement and use the client-server model where data is collected by smartphone and then sent to a server via a wireless network.

3.1.4 Unstructured Sensor Data

EHR data gathered from different sensors can be stored in files such as text files, csv files. This creates unstructured sensor files including a huge amount of health records captured from patients such as

heart rate and blood pressure. EHR based on sensor data enables collecting patient health information continuously due to the use of sensor devices. Consequently, patients can achieve continuity care because each health record is analysed and alerts may be sent to physicians and healthcare providers if there are indications of an emergency.

3.2 Structured data

Structured EHR is an electronic health record in which health-related information is stored in structured forms. In this category, medical information is collected using designated forms and interfaces which make it possible to index and search information quickly. It can be easily converted to different formats allowing the exchanging information.

3.2.1 XML file format

Extensible Markup Language (XML) is a simple, flexible and a standard meta-language recommended by the World Wide Web Consortium (W3C) as a mechanism for structuring data for large-scale electronic publishing. It is a system independent and programming language independent description language that is especially suitable to describe any type of data of known or unpredicted data structures. It uses hierarchical, object-oriented structure methods of description, which is well suited to describe the complex contents of medical records.

3.2.2 Relational database

Relational database is known as the most popular structure to store health records in healthcare systems. Accessing information in the database is through the simple and almost universal structured query language (SQL). Several standard Relational Database Management Systems (RDBMS) using any SQL-based software such as Oracle, Sybase or MySQL are used as databases in HIS. In RDBMS, EHR contents are stored in tables whereas EHR elements are mapped into corresponding fields and data types. In addition, relevant relationships among tables are also created by linking identified fields as primary keys in order to perform queries. Vucetic et al. (2011) designed and developed a health information system using central database The scheme describes crucial tables such as Patient, Health_id_card, doctor, patient_alergy, patient_surgery, check, and etc. Relational database uses key-based storage which enables fast retrieval by using queries and updating tools. However, using relational database method to store EHS has two major drawbacks: a) relational database cannot cope with frequently changes in the Healthcare Information System as it relies on static designs that do not allow data type extensions and b) the size limit of database tables prevents storage of big data.

4 MAJOR ISSUES IN EHR AND IMPLEMENTING EHR SYSTEMS

As discussed earlier, health information system encounter problems at many levels: interoperability at the data structures, communication protocols, security measures, management and scalability, and standardizations. This section discusses two major issues facing the E-Health community: interoperability and security and privacy.

4.1 Interoperability

Interoperability is a barrier to load, store and transfer health data in distinctive organizations and sites due to healthcare providers, e-health researchers and developers adopt EHR at different levels. Firstly, many nations have adopted with the implementation of HIS but the deployments are different from country to country. The wealthiest nations implement their own EHR systems or commercial HIS while developing nations employ open source solutions, producing disparate EHR systems. Consequently, interoperability among systems is difficult since each system relies on different EHR structures. In addition, HIS are developed using different EHR standards but some standards do not provide support for exchanging information. Especially, these standards use not only their own

protocols for extract information but also their own EHR structures for storing data. Furthermore, operating data on one site cannot run on another unless the HIS implements the same framework.

De la Torre et al. (2013) presented the analysis on the use of open-source and commercial EHR solutions in many nations. The results showed that these nations employ variety of software for EHR management. Sood et al. (2008) indicated that there are gaps among developed and developing countries in the deployment of EHR. Developed nations have national strategies for developing e-health network and proposed their own EHR standard structures while developing nations are adopting e-health in initial stages by building HIS for local healthcare professions. Consequently, these systems find it difficult to obtain the interoperability due to disparate resources and healthcare infrastructures.

Interoperability can be achieved at levels by integrating EHR into unified standards or compatible standards. However, this is difficult as HIS implementations differ greatly not only from country to country but also among healthcare providers.

4.2 Security and privacy

As EHR contains sensitive health data of a person, it must be protected from unauthorized accesses. Security and privacy are thus crucial in any HIS. Numerous approaches have been proposed and used for preserving the privacy and security of EHR. However, there is no clear classification of the security approaches. For simplicity, we classify these approaches into three categories: policy-based authorization, cryptographic, and policy based authorization and cryptographic. The policy based authorization allows data to be accessed with the role based access control. The cryptographic approaches employ encryption schemes and cryptographic primitives. Policy based authorization and cryptographic approaches combine the policy-based and cryptographic approaches.

4.2.1 Policy based authorization approaches

Policy based authorization approaches ensure that EHR can only be accessed by authorized users who received permission rights to access patient data in HIS. These approaches employ different policy mechanisms such as role based access control, time-bound access, task based access control or even a combination of these policies to define and assign roles for users. As a result, with both valid system identity credential and access credential, users can legally obtain corresponding patient data from EHR databases without disclosure sensitive information. Chen and Hoang (2011) proposed a novel framework with Cloud-based Privacy-aware Role Based Access Control model for controllability, traceability of data and authorized access to system resources in healthcare environment. Narayanan and Giine (2011) introduced Task Role-Based Access Control in which roles are used to support passive access control and tasks are used to support active access control. Policy based authorization approaches allow users accessing their EHR if they and their roles meet the systems' policies. However, health data are not encrypted and hence these approaches cannot protect sensitive data against malicious modification or intrusive attacks.

4.2.2 Cryptographic approaches

The Cryptographic approaches used to secure EHR are based on encryption schemes such as Public Key Encryption, Symmetric Key Encryption, and Attribute Based Encryption (ABE) and its variants. In these approaches, EHR contents are stored in encrypted forms to provide data confidentiality. The concept of Attributed-Based Encryption (ABE) was proposed by Sahai and Waters (2005). In this scheme, user keys and cipher texts are labelled by using the sets of descriptive attributes. Bethencourt et al. (2007) proposed Cyphertext-policy ABE to protect encrypted data even if the storage server is untrusted. Cryptographic approaches prove that EHR can be secured with high level by using encryption schemes. EHR will be stored in encrypted forms which ensure to avoid the disclosure to sensitive health information. Nevertheless, computing overhead and key management are challenges for applying these solutions. In fact, the lack of defined roles to access data in several schemes makes

it impossible to deploy the schemes in multi-users environment where users can own one or more roles in the system.

4.2.3 Policy based authorization and cryptographic approaches

The combinations of policy based authorization and cryptographic approaches ensure data confidentiality by defining roles for clients to access EHR and storing EHR in encrypted forms. These result in appropriate security frameworks as approaches embedded in HIS in order to serve for security requirements. In order to access EHR, users need to obtain corresponding roles which describe permissions such as read, write to EHR. These roles have to be authorized by HIS or third parties before users receive keys for decryption processes. (Kathryn Garson & Adams 2008) proposed Policy Based Encryption which builds on the idea of encrypting under an arbitrary string. In order to access the document users need to be authenticated by the system and receive a decryption key associated with their role.

These approaches provide high security level and meet with requirements when healthcare move to new infrastructures such as cloud computing. However, it is needed to consider that the approaches are complex and require systems implementing on the same framework or adaptable framework while majority of current HIS are developed based on different sources and providers.

5 CHALLENGES TO BE ADDRESSED IN FUTURE RESEARCH ON HEALTH CLOUDS

It is anticipated that electronic health records and healthcare-related services will be deployed on cloud platforms to reduce the cost and complexity of handling medical records while improving efficiency and accuracy. However, users and health service providers are reluctant to adopt Cloud computing because of two main concerns: protection of sensitive data and efficient access of health data. Latency in retrieving relevant information from hundreds of millions of health records over a large number of geographically distributed data servers could very well be the difference between life and death for a patient (Tancer & Varde 2012). These present serious barriers for wide adoption of cloud computing in the E-Health regardless of the reduction in costs.

5.1 Protection health data in Health Cloud Environments

Health records contain sensitive information that must be protected. This means that the data should only be accessed by authorised handlers AND that every action on the data has to be accounted for. The users are concerned about their electronic health record (EHR) not fully protected as they lose the control of the data once it is transferred to the Cloud.

In the Cloud Service Providers' environments, the main concern is that patients might lose control of their own EHRs. They might not know who can gain access to their sensitive data, how the EHRs are processed, what details are disclosed to others, and whether the security procedure and privacy protection conform to defined service level agreements (SLAs). Existing work lacks appropriate schemes to protect users' sensitive EHRs from illegal disclosure or malicious violation by employees within cloud service providers (CSPs). Such an incident has occurred: an employee stole the US Department of Veterans Affairs database that contained sensitive personal health information of 26.5 million military veterans and took it home without authorization (Virvilis et al. 2011).

Traditional security and protection mechanisms are not sufficient to protect data: When user data is passed onto a Cloud, the user no longer has control over the data and relies on the Cloud provider to assure them that their data is in a safe hand. Traditional security and protection mechanisms used by cloud providers such as encryption, authentication protocols, and digital signature mechanisms are not sufficient to address the above challenges and concerns, especially in big data context with increasingly complex health data formats. For example, Virvilis et al. (2011) introduced a cloud

provider agnostic protocol combining hybrid encryption and message authentication code technologies to preserve data confidentiality and integrity for outsourcing both static and dynamic data to third parties. Shucheng et al. (2010) proposed a scheme that achieves fine-grained access control, data confidentiality, and scalability by key policy attribute-based encryption and proxy re-encryption and lazy re-encryption techniques. These traditional concepts and techniques do not really protect data as they cannot answer when the data is accessed, what has been altered, where it is moved, by whom, furthermore, they inevitably bring in complexity and performance bottleneck due to key distribution, sophisticated encryption and decryption.

Existing work lacks scheme to prevent intrusion, data leakage: Existing work lacks appropriate schemes to efficiently prevent intrusion attack, data leakage, and deliver controllability and transparency to data owners. The related works focusing on policy-driven frameworks (Dan Lin & Squicciarini 2010; Pankaj Goyal & Mikkilineni 2009; Takabi & Joshi 2012) play a role of adjudication in deciding the legality of access in the external layer of data storage But cloud data is still in danger if adversaries have sufficient hacking skill or can leverage elevation of privilege to bypass access control service. Trust computing related technologies (Abawajy 2011; Firdhous et al. 2011; Cong et al. 2010) administrate cloud data security through trusted third party management systems. These schemes can efficiently increase the transparency of data usage and ensure that data is not being compromised or leaked by CSPs. However, these frameworks still disclose vulnerability once the hosted third party security services are compromised. Often, in a cloud environment with a complex and dynamic hierarchical service chain, data handling may be delegated from one CSP to another for business reasons . These CSPs do not always employ the same protection schemes and standards (Foster et al. 2008) and data may lose its protection on the new cloud hosts. Therefore, ensuring healthcare cloud data security, mitigating users' concerns, and encouraging broader adoption of cloud computing in healthcare sector requires alternative methodologies and technologies.

5.2 Efficient management of big data in Health Clouds

Currently, EHRs are stored in various proprietary formats through a multitude of medical information systems available on the market. There is an increase in the requirements of accessing information anywhere and anytime and exchange of health information and this drives the need to address the efficient management of health data. Efficient health data management must not only reduce the costs but also address protection, access speed, availability, and mobility of data.

Along with science data, financial data, social network data, health data has also entered big data era. Ultra-high "volumes" of health data collected from patients, hospitals, and research organizations not only improves healthcare quality and aid clinicians and researchers but also helps transform the way the hospital is managed with cloud storages (Lewis 2013), resulting in a reduction of complicated procedures, efficient use of resources, and operational costs. Big data analytics have huge potential to better match health care provision with need in e-Health (Tancer & Varde 2012). When health data reaches big data level its management presents higher level of complexity and difficulty. Various types of EHRs may require different security mechanisms that will inevitably cause additional overhead. Improvements and reductions in costs would not be very useful if these made healthcare professionals worse at their jobs by not being able to access relevant data timely. Latency in retrieving relevant information from hundreds of millions of health records over a large number of geographically distributed data servers could very well be the difference between life and death for a patient (Tancer & Varde 2012). Managing data usage in a big data context becomes a huge challenge and presents another barrier for wide adoption of cloud computing in healthcare industry regardless of the reduction in costs.

Though much work has been invested into distributed data bases, management of big data has not been addressed. Particularly, the challenges for health data management and related performance issues, new means of distribution and processing of data are needed. New data distribution models and programming models must address aspects of distribution, parallelism and replication. MapReduce

(Dean & Ghemawat 2008) is an approach developed by Google for performing large-scale data analysis over its Google File System. It has received much attention because it provides a simple model through which users can express relatively sophisticated distributed programs. MapReduce programs consist only of two functions, called Map and Reduce, to process key/value data pairs. It permits data to be in any arbitrary format and supports flexible data distribution. Dynamo (Giuseppe DeCandia et al. 2007) is another highly available key/value storage system that has been used by Amazon's core e-commerce platforms. Data is partitioned and replicated using consistent hashing. Dynamo can be characterised as a special form of distributed hash table (DHT) peer-to-peer systems, where each node maintains enough routing information locally to route a request to the appropriate node that holds the requested data directly. Addressing the above issues and challenges is essential to realizing the potential of healthcare cloud.

6 ACTIVE DATA CUBE MODEL FOR EHR PROTECTION

This section introduces our active data cube model for encapsulating an EHR for housing and protecting it in a Cloud environment. Cloud data can be classified as *structured* or *unstructured data* in terms of management type. The term *structured data* refers to data with an identifiable structure. The most common instance of the *structured data* is the database system, where data is stored based on predefined features and properties and is also searchable by data type with access interfaces. Conversely, *unstructured data* normally has no identifiable structure that refers to any data type. Media data, documents, and complex designated data formats like the EHR are considered *unstructured data*.

To protect these various data types in the outsourced cloud environment, *structured data* management typically interfaces with data by using secure connection interfaces. *Unstructured data* strongly relies on third party security mechanisms or encryption. Once third party services are compromised, *unstructured data* would be vulnerable to violation and tampering. In this work, we concentrate on protection mechanisms for *unstructured data.*

Instead of paying attention to attacks or violations, we focus on the target data and equip it with self-describing and self-defending capability. We introduce an Active Data Cube (ADCu) structure as a deployable data protection unit encapsulating sensitive data, networking, data manipulation, and security verification functions within a coherent data structure. A signed ADCu encloses dynamic information-flow tracking mechanisms that can precisely monitor the inner data and its derivatives. Any violations of policy or tampering with data would be compulsorily recorded and reported to data owners via the notification mechanisms within ADCu. This strong enforcement also triggers log information collection procedure (which records every access to the specific data items) that will be utilized to provide transparency and accountability of data usage via disseminating the log information to data owners.

As shown in Figure 4, our proposed ADCu consists of a *shell* and a *core*. The *shell* is equipped with active tamper-proofing codes and is executed prior to the *core* when the ADCu is triggered. The ADCu is associated with a runtime environment.

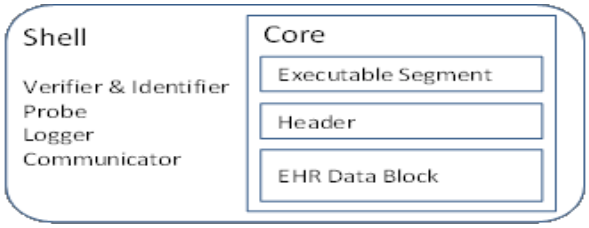

Figure 4. Structure of an Active EHR (ADCu)

At the entry point to the *shell*, the scripts invoke *verifier and identifier* to certify the validity of the request in terms of the format of request parameters and contents via request identification and

verification process. In our context, a permitted access request conforming to the configured policy is issued a certification which signifies that the access is authorized by the Cloud-based Privacy-aware Role Based Access Control (CPRBAC) service which was proposed in our previous work (Chen & Hoang 2011). Another significant component of the *shell* is the *logger* module, utilized to record significant checkpoints during transactions, data operation outcomes, and even errors when the ADCu throws exceptions. The *logger* is required to record significant intermediate information. During a single transaction, all log records marked with regular *Priority* level are stored temporarily in memory for performance consideration. Once the data operation finishes, the *logger* leverages the *communicator* in the *shell* to upload the log records to ADCu's external supervisor. However, a log record marked with an emergency tag will be immediately triggered by the *probe*, which then notifies the *communicator* to raise an exception. *TimeStamp* uses the Network Time Protocol to take into account the fact that cloud resources may be distributed across different time zones. Each ADCu's log information is transparent to its data owner. When the log records are stored in cloud, they are encrypted using the RSA encryption to avoid possible information leakage. Only the data owner has the corresponding key to disclose those records. In addition, sending out the log information of data usage rather than storing it inside the ADCu is activated to maintaining the light-weight feature. Increasing log information could raise the cost of storage, dissemination, or replication of the ADCu.

Each ADCu has a corresponding *supervisor* deployed in the same domain, which takes charge of monitoring external data operations (such as move, and copy) that cannot be detected by the internal probe inside the ADCu, and communicating with its ADCu. If the ADCu cannot establish a proper network connection or cannot contact its supervisor, it would switch to the termination state to avoid offline attack.

A *probe* in the *shell* is triggered by three types of activity: program exception, inconsistent checksum in data blocks, and verification failure of zero-knowledge proof procedure.

Once the verification and identification procedure succeeds, the *shell* delegates control to the data *core*. The *core* of ADCu is wrapped by an Executable Segment (ES), a header, and data blocks. We leverage dynamic delegation approach in the ES to call back the *shell* to trigger the *core* and execute data operations. The *header* refers to a manifest specifying the basic information of supplementary data residing at the beginning of *data blocks*.

Work has been progressed in making used of this active data cube model for our earlier proposed Mobile Cloud for Assistive Healthcare (Hoang & Chen 2010).

7 ON THE APPLICATIONS OF ACTIVE DATA CUBE MODEL

Clearly, cloud computing has become an alternative IT infrastructure for businesses and governments, providing digital services including healthcare services. We introduced our Mobile Cloud for Assistive Healthcare (Hoang & Chen 2010) and are implementing this Active Data Cube model for encapsulating health records within the Mobile Cloud (Chen et al. 2014).

It becomes clear that more and more we have to deal with bigger and bigger amount of data as well as its mobility within cloud environments. Our next step of investigation will concentrate on several issues: 1) investigating mechanisms and techniques for protecting data mobile within and around Inter-cloud environments, and 2) investigating suitable data structure for big health data so that efficient storing, accessing and processing can be performed.

8 CONCLUSION

This paper considers issues and challenges pertaining to EHR that are concerns of e-Health researchers and developers who aim to produce more efficient and effective EHR systems. It presents a brief review of EHR including definitions, EHR file formats, and structures to provide a common understanding of EHRs and the complexity of EHR systems. The paper discusses several important

issues related to EHR and EHR systems and presents challenges that have to be addressed for realizing and wider adoption of Cloud-based healthcare systems. The paper presents an Active Data Cube model that can be used to protect health data in an outsourced environment such as Cloud.

References

Abawajy, J. (2011). Establishing Trust in Hybrid Cloud Computing Environments. IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). 118-125.

Bethencourt, J., Sahai, A. and Waters, B. (2007). Ciphertext-Policy Attribute-Based Encryption. IEEE Symposium on Security and Privacy (SP '07). 321-334.

Chen, L. and Hoang, D. B. (2011). Towards Scalable, Fine-Grained, Intrusion-Tolerant Data Protection Models for Healthcare Cloud. IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). 126-133.

Chen, L., University of Technology, S. F. o. E. and Technology, I. 2014. *Achieving Trust-oriented Data Protection in the Cloud Environment*.

Cong, W., Qian, W., Kui, R. and Wenjing, L. (2010). Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing. Proceedings of IEEE INFOCOM. 1-9.

Dan Lin and Squicciarini, A. (2010). Data protection models for service provisioning in the cloud. Proceedings of the 15th ACM symposium on Access control models and technologies. 1809872: ACM, 183-192.

De la Torre, I., Martinez, B. and Lopez-Coronado, M. (2013). Analyzing open-source and commercial EHR solutions from an international perspective. IEEE 15th International Conference on e-Health Networking, Applications & Services (Healthcom). 399-403.

Dean, J. and Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Commun. ACM,* 51**,** 107-113.

Firdhous, M., Ghazali, O. and Hassan, S. (2011). A trust computing mechanism for cloud computing with multilevel thresholding. 6th IEEE International Conference on Industrial and Information Systems (ICIIS). 457-461.

Foster, I., Yong, Z., Raicu, I. and Shiyong, L. (2008). Cloud Computing and Grid Computing 360-Degree Compared. GCE '08 Grid Computing Environments Workshop. 1-10.

Giuseppe DeCandia, Deniz Hastorun, Madan Jampani, Gunavardhan Kakulapati, Avinash Lakshman, Alex Pilchin, Swaminathan Sivasubramanian, Peter Vosshall and Vogels, W. (2007). Dynamo: amazon's highly available key-value store. *SIGOPS Oper. Syst. Rev.,* 41**,** 205-220.

Häyrinen, K., Saranto, K. and Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *International Journal of Medical Informatics,* 77**,** 291-304.

HIMSS. *Healthcare Information and Management Systems Society* [Online]. Available: http://www.himss.org/.

Hoang, D. B. and Chen, L. (2010). Mobile Cloud for Assistive Healthcare (MoCAsH). IEEE Asia-Pacific Services Computing Conference (APSCC). 325-332.

Junqing, S., Sundara-Rajan, K., Lindsey, L., Mamishev, A., Johnson, E., Teredesai, A. and Kristal, A. (2011). A pervasive Dietary Data Recording System. IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). 307-309.

Kathryn Garson and Adams, C. (2008). Security and privacy system architecture for an e-hospital environment. Proceedings of the 7th symposium on Identity and trust on the Internet. 1373306: ACM, 122-130.

Khudair, A. A. (2008). Electronic health records: SAUDI PHYSICIANS' PERSPECTIVE. 5th IET Seminar on Appropriate Healthcare Technologies for Developing Countries (AHT 2008). 1-7.

Lewis, F. 2013. *Industry view: Big data and health* [Online]. Available: http://www.ehi.co.uk/resources/industry-view/126 [Accessed 14/05/2014 2014].

Lj, B., A, G. and H, S. (2000). Describing Electronic Health Records Using XML Schema. *XML Asia Pacific*.

Narayanan, H. A. J. and Giine, M. H. (2011). Ensuring access control in cloud provisioned healthcare systems. IEEE Consumer Communications and Networking Conference (CCNC). 247-251.

Pankaj Goyal and Mikkilineni, R. (2009). Policy-Based Event-Driven Services-Oriented Architecture for Cloud Services Operation; Management. IEEE International Conference on Cloud Computing, 2009 (CLOUD '09). 135-138.

Roberts, R. (1999). R. S. Dick, E. B. Steen and D. E. Dether (EDS), The Computer-based patient record: an essential technology for health care. Revised edition. Washington DC: Institute of Medicine, National Academy Press, 1997. ISBN 0-309-05532-6, 234 pages. £28.95. *The International Journal of Health Planning and Management,* 14**,** 74-75.

Sahai, A. and Waters, B. 2005. Fuzzy Identity-Based Encryption. *In:* CRAMER, R. (ed.) *Advances in Cryptology – EUROCRYPT 2005.* Springer Berlin Heidelberg.

Schubert, L. and Jeffery, K. (2012). Advances in clouds. *Report of the Cloud Computing Expert Working Group. European Commission*.

Shucheng, Y., Cong, W., Kui, R. and Wenjing, L. (2010). Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing. Proceedings of IEEE INFOCOM. 1-9.

Sood, S. P., Nwabueze, S. N., Mbarika, V. W. A., Prakash, N., Chatterjee, S., Ray, P. and Mishra, S. (2008). Electronic Medical Records: A Review Comparing the Challenges in Developed and Developing Countries. Proceedings of the 41st Annual Hawaii International Conference on System Sciences. 248-248.

Sundareswaran, S., Squicciarini, A., Lin, D. and Shuo, H. (2011). Promoting Distributed Accountability in the Cloud. IEEE International Conference on Cloud Computing (CLOUD). 113-120.

Takabi, H. and Joshi, J. B. D. (2012). Policy Management as a Service: An Approach to Manage Policy Heterogeneity in Cloud Computing Environment. 45th Hawaii International Conference on System Science (HICSS). 5500-5508.

Tancer, J. and Varde, A. S. (2012). Cloud technology and EHR data management. IEEE 6th International Conference on Information and Automation for Sustainability (ICIAfS). 112-117.

Virvilis, N., Dritsas, S. and Gritzalis, D. 2011. A Cloud Provider-Agnostic Secure Storage Protocol. *In:* XENAKIS, C. and WOLTHUSEN, S. (eds.) *Critical Information Infrastructures Security.* Springer Berlin Heidelberg.

Vucetic, M., Uzelac, A. and Gligoric, N. (2011). E-Health Transformation Model in Serbia: Design, Architecture and Developing. International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). 566-573.

Xu, Y., Chen, J. and Peng, H. (2009). Research of Electronic Patient Record Based on XML. International Conference on Management of e-Commerce and e-Government, 2009 (ICMECG '09). 219-222.