A Secure Real Time Data Processing Framework for Personally Controlled Electronic Health Record (PCEHR) System

Khandakar Rabbi[1(✉)], Mohammed Kaosar[2], Md Rafiqul Islam[1], and Quazi Mamun[1]

[1] School of Computing and Mathematics, Charles Sturt University, Bathurst, Australia
{krabbi,mislam,qmamun}@csu.edu.au
[2] Department of Computer Science, Effat University, Jeddah Saudi Arabia
mkaosar@effatuniversity.edu.sa

Abstract. An era of open information in the healthcare is now underway. This information can be considered as 'Big data', not only for its sheer volume but also for its complexity, diversity, and timeliness of data for any large eHealth System such as Personally Controlled Electronic Health Record (PCEHR). The system enables different person or organization to access, share, and manage their health data. Other challenges incorporated with the PCEHR data can be very excessive to capture, store, process and retrieve the insight knowledge in real time. Various PCEHR frameworks have been proposed in recent literature. However, big data challenges have not been considered in these frameworks. In this paper, we argue the PCEHR data should be considered as big data and the challenges of big data should be addressed when to design the framework of the PCEHR system. In doing so, we propose a PCEHR framework, which deals with real time big data challenges using the state-of-the-art technologies such as Apache Kafka and Apache Storm. At the same time the proposed framework ensures secure data communication using cryptographic techniques. Using a qualitative analysis, we show that the proposed framework addresses the big data challenges.

Keywords: PCEHR · Big data · Apache kafka · Apache storm · Big data security

1 Introduction

On 8 March 2014, the Malaysian Airlines flight MH370 was scheduled from Kuala Lumpur to Beijing and lost contact with the air traffic control about an hour after it took off. Within few weeks of this incident took place, the search text "Malaysian airlines MH370 missing" in Google returned about 160,000,000 results. As the news was updated very frequently, anyone could view the latest news from the result filtering option and could see the news coming from last 24 h. This huge list of result can be categories by 'Visited pages', 'Not yet visited', 'Reading level' and so on. This filter is a good example of so call "Big data Processing" where all data are coming from multiple, heterogeneous and anonymous places and they have a complex relationship which is evolving and growing each second.

J. Tian et al. (Eds.): SecureComm 2014, Part II, LNICST 153, pp. 141–156, 2015.
DOI: 10.1007/978-3-319-23802-9_13

Big data is one of the current and future research trends [2, 3]. Big data can be characterised by their *volume*, *velocity*, and *variety* (3V) [1]. Here volume refers to the size of *big data*, where velocity refers to the speed of data, and variety indicates the various sources of data [4]. In some cases there could be an extra feature, depending on the requirements, which can be any of *Value*, *Variability* or *Virtual* [1]. In general, *big data* is a collection of diversified large data sets which is extremely difficult or nearly impossible to process using the traditional data processing and management techniques [1]. *Big data* is also formidable to capture, cure, analyse and visualise using the existing technologies [1]. Thus, within current technology limitation there are few challenges in *big data* including *Storage*, *search*, *sharing*, *analysis*, *visualization*, *capture*, *security*, *privacy* and *curation*.

The increasing volume of the data generated in the eHR systems indicates that the healthcare organizations will not be able to complete analysing these real time data [19]. Moreover, about 80 % of overall medical data is unstructured and clinically important [19]. These huge amount of data is retrieved from multiple sources like EMR, lab and imaging systems, physicians prescriptions, medical correspondence, insurance claims even Customer Relationship Management (CRM) systems and finances [19]. Thus, the verity of data is great in numbers. Since each person has medical records therefore the volume and velocity is also large. Figure 1 shows the number of people register in PCEHR from July 2012 to July 2013.

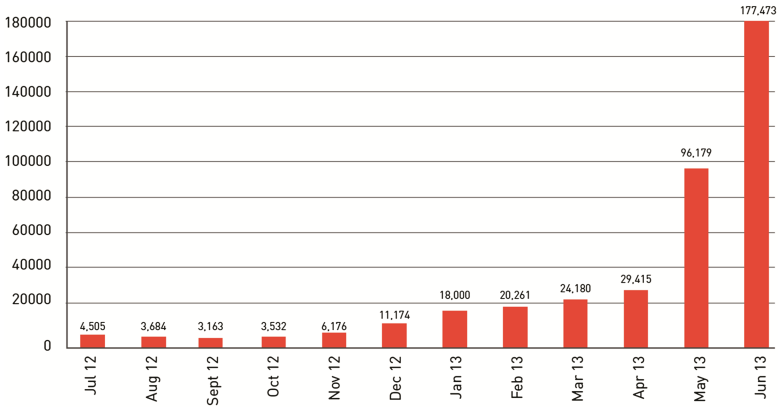

Fig. 1. Number of people registered in the PCEHR [21]

Illustrating Fig. 1, for the volume of the data in PCEHR; it can be considered as *big data*. A figure of Austrian Institute of Technology shows in Fig. 2 which state a collection of components are interconnected for processing for a eHealth platform [20].

Big data application includes healthcare, medical and government services where these data are often shared or released to the third party partners or public to analyse the insight knowledge [5]. Recently, National E-Health Transition Authority Australia, introduces PCEHR (Personally Controlled Electronic Health Record) to manage individuals health record and allow them to authorise other individuals who is eligible to view their electronic health records (eHR) [22]. It helps individual, their doctors and

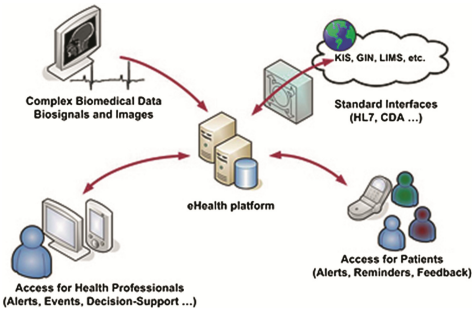

Fig. 2. eHealth platform [20]

other health care providers to view their medical record and provide the best possible medical care [6, 22].

Many different relevant works have been published over the past few years [6–16]. In most of them, authors around the world try to improve the system usability through different surveys. Some of the papers described and improved the system without evaluating the system as a *big data* platform. Thus a very important aspect of research has been missing until now.

Since data generated in PCEHR is a considered as *big data*, some immediate research questions are to be addressed in which not much significant research work has been done, such as:

- How to store and process the big amount of data generated in PCEHR.
- How to handle and analyse the *big data* in real time to make sure there is no health hazard due to delay.
- How to ensure data security and privacy.

In this paper, we address the aforementioned issues of *big data* while proposing the framework for PCEHR system. We use the state-of-the-art technologies like Apache Kafka and Apache Storm for real time data capturing and processing. We use public key cryptographic technique to preserve privacy.

The rest of the paper is organized as follows: Sect. 2 describes the existing framework for eHR system. Section 3 discusses about the background of the related technologies. Section 4 describes the proposed framework. A qualitative analysis has been carried out at Sect. 5, and finally the paper ends with conclusion and future work in Sect. 6.

2 Existing eHealth System Frameworks

In recent years, many eHealth systems have been proposed. To design and develop these systems, researchers put emphasis on different issues such as privacy preserving, secure data transactions, high data availability through cloud and distributed approach, real time decision and storage. Some researchers perform surveys to measure the user acceptance and adoption capabilities. However the proposed systems fail to address the issues of *big data* circumstances. Below is a brief description of different approaches of existing eHealth system.

To ensure the maximum privacy, in [6] a PCEHR model has been proposed where a fully homomorphic encryption technique is proposed. Proposed end to end framework shows several entities like General Practitioners, Specialist doctors, Nurses, Pharmacist, diagnostic Labs, Private clinics, Hospitals, Family & Friends and how they access a global encrypted database server. To ensure maximum protection, an authentication server, an ACL (Access Control List) server and an authorization Server is used. Although it can ensure the privacy of data, the authors in [6] do not evaluate their concept in big data scenario. Hence to evaluate the system under *big data* environment, the challenges and research questions yet to be answered. A design of patient safety reporting system is introduced in [11]. XML parser and code generator is used to communicate with different database system (sources) and generate a report which helps personnel a secure and safety lifestyle. To enable privacy and security authors use data encryption techniques. Challenges in Big data are out of the scope of the paper. Privacy and security issues are also discussed on the paper [13]. A description of current available methods is discussed and some of the issues are mentioned briefly. However, the discussion was only limited to privacy and no further discussion on *big data* was included. In [9], the authors proposed a framework which illustrates a secure process and a recovery process to ensure the privacy. The scope of the paper was only the security and recovery and no *big data* environment is considered.

In [7], a platform call MyPHRMachine describes a way to reduce the impediments to data transfer. Authors claims that the proposed platform is low cost and can substitute by cheap software components. The platform is open sourced and trustable which is a cloud-based system where patients provide access their data to different third parties. It ensures some of the privacy, however the challenges of *big data* is out of the scope of the paper. The scope of paper [12] is to present a technique showing how to collect data from different hospitals. It uses a server client model with different gateways which collects data from multiple sources. This distributed model uses XML files to communicate with a local server and all the clients. Although the scope of the paper was collecting information, the entire picture for continuous (*big data*) data is missing. A simulation model of centralized and distributed data structure is carried out for health care data in [15]. The model uses Monte Carlo method which iteratively evaluate by a set of random numbers as inputs. The model examines on 10,000 patients input data. However, the scenario of continuous and unstructured data was out of the scope of the paper.

In [8], a health management system survey is carried out. No framework is proposed thus a *big data* and its challenges are out of the scope of the paper. An interview study on the benefit of electronic health record system is reported by [16]. A theoretical framework known as DeLone and McLean's Information Systems Success Model (D&M IS Success Model) is used to measure the adaptability of electronics health record system. They examine three health care models including 'RSL Care', 'Uniting Care Ageing South Easter Region' and 'Warrigal Care'. This theoretical framework (D&M IS Success) helps to understand an information system in terms of 'system quality', 'information quality', 'service quality', 'ease of use', 'user satisfaction' and 'net benefit'. However all the three abovementioned health care models doesn't support and implement continuous and unstructured data. Thus *big data* challenges are missing from this study.

Authors in [10], describes a personal health record system of Lombardy, Italy. The system provides a complete, integrated and contextualized patient history which helps patients by real time decision supports. It also supports storage system thus it can increase efficiency and real time emergency care. However security and privacy was missing and the big data challenges are out of the scope of the paper. New York Presbyterian System is described in [14]. The system helps patients to manage their health profile with list of medical reports and available medication they are going through. Patients can see the list of healthcare and care providers, insurer and can enroll them into any of the available system. However, the system allows patients to manage their profile and it uses typical database management system. Thus the concept and challenges in *big data* is out of the scope of the paper.

In the next section we perform some background study of state-of-the-art technologies which is used in our proposed framework in Sect. 4.

3 Technology Required for PCEHR

The previous section illustrates different eHealth systems, and identifies the stipulation of state-of-the art technology incorporation for a large scaled eHealth system such as PCEHR. These technologies would be able to process real time unstructured, continuous data set arriving from multiple heterogeneous sources. A brief description of these technologies are provided below.

3.1 Apache Kafka

Apache kafka is a fast, scalable, durable and distributed publish-subscribe messaging system. It can handle hundred of megabytes of read/write from thousands of clients in real time. When data is too big and continuous; data streams are partitioned and spread over a distributed machines (clusters). Data is persisted on disk and can be replicated within the cluster which prevents data loss. Each cluster is called “Brokers” which can handle terabytes of data without any impact on performance. Kafka can be actively use for real time processing where raw data can be consumed and then several data analysis activities such as aggregated, summarized, or transformed to another format is done for further consumption [17].

3.2 Apache Storm

Apache Storm is a distributed real time computation system. It is scalable, fault-tolerant and guarantees that data will be processed. Storm provides some set of general primitives to do real time computation. It creates topologies deployed in clusters. A topology is a graph of computation which contains processing logic and links among nodes which indicates how data is passed throughout the nodes. There are two types of nodes in storm cluster: ‘Master’ and ‘Worker’. Master nodes run on a daemon which is known as ‘Nimbus’ which is responsible for distributing the code around the clusters, assigning tasks to different machines, and monitoring the success and failure. Worker nodes runs a daemon called supervisor which listen to the work assigned to the individual machine,

start and stop worker processes when necessary based on what Nimbus assigned. Each workers execute a subset of topology (a running topology consists of many worker processed across many machine). The coordination between Nimbus and Supervisor is done through a Zookeeper cluster. A Zookeeper is a state-full cluster which keeps track of session/data into memory or local disk. On the other side both Nimbus and Supervisor is state-less. This means, even if Nimbus and Supervisor is failed, the entire data and session will not be destroyed. Figure 3 shows the component of storm cluster.

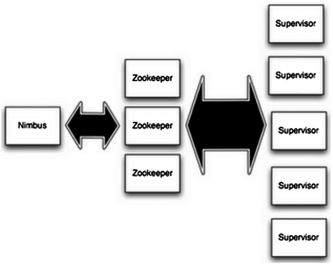

Fig. 3. Component of storm cluster [18]

Storm provides 'spouts' and 'bolts' for doing stream transformations. A spout is a source of stream where a bolt consumes any number of input streams and does some processing. For doing fairly complex computation requires multiple steps and thus multiple bolts. Bolt can do anything like, running a function, searching memory, aggregation even connecting with database. The network of spouts and bolts are a package which is known as a topology. Figure 4 shows a basic topology.

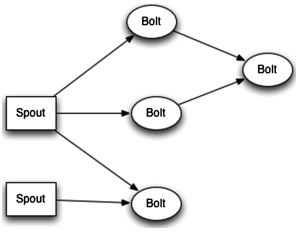

Fig. 4. A simple storm topology [18].

4 Proposed PCEHR Framework

The proposed model of PCEHR is depicted in Fig. 5. The model is divided into four modules consist of 'Patient Care', 'PCEHR Data Receiving', 'PCEHR Data Processing' and 'PCEHR System DB and Access System'. The proposed distributed system is an end to end solution showing how the whole PCEHR system works. This system is deployed into different locations. The first part of the system is 'Patient Care' which is deployed into patients home. This module is also deployed in a smart home or nursing

home where sensors and web applications are available to send health data. The next part is 'Data Receiving' section which is deployed in the PCEHR application server. The 'PCEHR Data processing' unit is also deployed in Application Server from where data will store in an encrypted database server. The rest of "Access System" is taken from [6] where operations operate through Authenticated server, ACL server and Authorization server. In the rest of the subsection, we have discussed each part of the entire system separately.

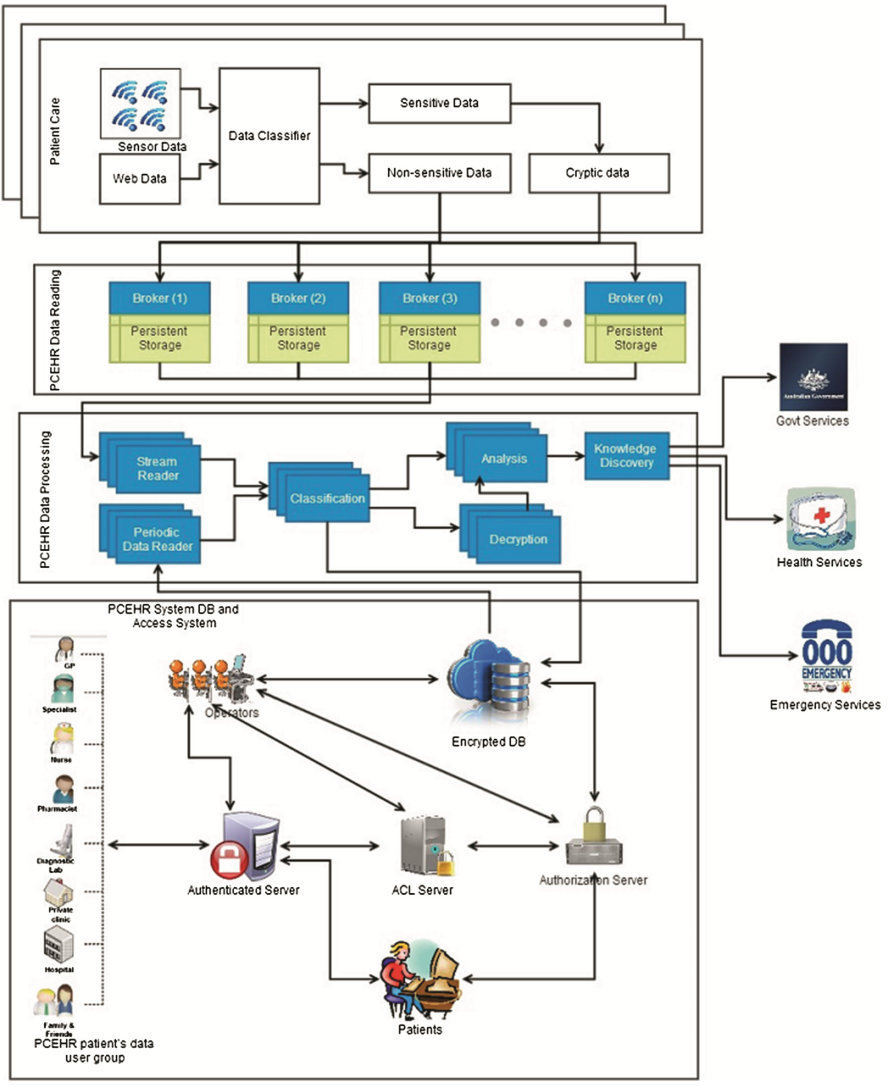

Fig. 5. Proposed architecture of the system.

4.1 Patient Care

This part of the system is deployed into individual's home or hospital or nursing home. Multiple sensors can generate huge data as well as doctors, nurses, pathologies, hospitals, laboratories or even nursing home agents can post complied data into the system. 'Patient Care' module consists of several sub-modules including sensor data, web data, data classifier, sensitive data, non-sensitive data and cryptic data. The next sub-section briefly describes all those sub-modules.

- ***Sensor Data***

This captures real time unstructured data from different sensor devices. A list of sensors can be deployed at patient's location which receives patient's continuous data such as blood pressure, sugar level etc. and post data to the internal processor. The list of sensors includes smart mobile application, sensor hardware or any type of device which can capture real time data for continuous monitoring. An internal processor is an integrated processor which is bind with sensors. As an example, smart phone devices have sensors and integrated processors. Sensor data are raw thus a preliminary processor required it to pre-process for post management activities. In real life, bar code reader, road cameras for vehicle identification are such examples of sensor device.

- ***Web Data***

This captured data are more structured and usually post by individuals or by third party. Doctors, diagnosis lab assistants or nurses posts data of a patient's current physical status. This type of data embeds describing medical reports, lab reports, and medical conditions. It includes several sensitive data including patients name, age, and address. This type of data also includes image type data i.e. scan copy physical/mental/dental reports. Thus, before posting the data to the PCEHR Server, a series of processing such as extraction, normalization, cleansing, transformation, joining is required. After performing the appropriate processing data can be standardize by an acceptable format (XML) which can be posted to 'Data Classifier'.

- ***Data Classifier***

It classifies data as sensitive or non sensitive. The purpose of the proposed framework is to perform proper privacy preserving practices thus this part of the system plays an important role in the entire architecture. Before the classifier works, patients can specify which data can be sensitive for them. If any patient doesn't perform how their data sensitivity classifier works then a default one is used. To protect the unauthorized access of the private data, this part indentifies sensitive data through user's defined rules. As an example, user can specify that s/he have their 'date of birth' is a sensitive data. Thus within a complete data set, 'Data Classifier' validates and tag the 'date of birth' as a sensitive data and send it to the appropriate handler. Other than that, non-sensitive data are sent to the appropriate handler.

- ***Sensitive Data***

Sensitive data can be defined as information that must be protected against unwanted disclosure. It may include personal information, dental record history, mental treatment history or report. In the stage of the system, the proposed framework applies different privacy/security algorithms to make sure data is secure to release. A wide verity of data encryption algorithm is available from the last decades. One of the state-of-the-art-technology in data encryption is homomorphic encryption technique. Since homomorphic encryption algorithm requires longer execution time thus we are not considering this technique in this proposed framework. The proposed system will apply the public key cryptography algorithm to ensure that sensitive data is protected against unauthorized access.

- ***Non-sensitive Data***

This module temporarily holds non-sensitive data. A patient has some of the non-sensitive information such as their family medical history, their life style including smoking, high risk sports, alcohol etc. Typically, non-sensitive information are the dataset a by which a patient can't be individually identified. This data can be post to any system for the purpose of govt. service, research activities. This type of data doesn't require any algorithms to protect from public access. This part of proposed framework holds the non-sensitive data and posts it to the 'PCEHR Data Receiving' module.

- ***Cryptic Data***

A modified data set which hold un-meaningful, non-understandable data. This is the modified sensitive data which can be post publicly. The proposed framework applies public key cryptography technique and stores the modified cipher data in this section. This type of cryptography technique depends on a piece of accessory information usually called 'key'. In this case the sensitive data is encrypted by a public key. The public key is open and anybody can use to encrypt the data. The data can be only decrypted by a private key. This is one of the most popular encrypted techniques which allow encrypting almost everything. Without the knowledge of the key it is near impossible to decrypt the cipher text into readable format.

4.2 PCEHR Data Receiving

A large number of patient's data are sent continuously. The propose framework uses Apache Kafka brokers to handle large data stream. Each broker consists of persistent storage, where data are store before processing. Kafka deploys as many as broker needed depending on the volume of the data. This module guarantees each data is ready for processing and kept in the memory as long as it is not processed.

4.3 PCEHR Data Processing

In this module of the propose framework Apache Storm has been used for real time data processing. One of the great characteristics of Apache Storm is it ensures no data failures. It is responsible of reading the stream data and sends it to classifier. Classifier classifies data and sends it to database. It also sends the encrypted data to decryption module to decrypt it. Both classification module and decryption module send the data to real time analysis module for analysing. After analysis, data is sent to knowledge discovery module. Classification is also responsible for data storage to database. Knowledge Discovery module sends the aggregated knowledge to different external sources like emergency server, health service and govt. service. Next sub section describes the internal sub-modules.

- ***Stream Reader***

This module pulls data from Apache Kafka brokers and sends it to classification module. This module contains both cipher and non sensitive data. 'Stream Reader' represents a simple program which pulls data from broker. This small module helps to maintain a strong, flexible and scalable architecture. In the typical system, data usually posted to web services regardless of how resourceful server is. If the data posting becomes more than expectable, cases were server rejecting data (data lost) or server crash. The proposed system used storm 'Spout' (described in previous section) which make the system flexible and a way to guarantee of data processing, failure prevent system.

- ***Periodic Data Reader***

This module pulls structured data from database. Data summarization, aggregation can help govt. and researchers to utilize the inside knowledge. Periodic data reader read data and sends it to classifier aiming to facilitate govt. or apply other health services to individuals. This is also a data reader module which read data from database. This structured data can be complied with an acceptable format so that data analysis process can be fastened. This part of propose framework uses data compression technique thus the amount of data will not be large enough to handle.

- ***Classification***

Classification classifies encrypted and non-encrypted data. Encrypted data is sent to the 'Decryption' module. Classification module is also responsible for classifying data. Some data may not be important to store and a data compression and summarization algorithm performs to compress a large a dataset into a tiny data set. As an example a patient's normal blood sugar for a range of time which is erased from the system if necessary. It will help reduce the overhead costs and complexity of the entire database server. However, classification also sends non-encrypted data to 'Analysis' module.

- ***Decryption***

This module is responsible to decrypt the sensitive data before analysis. A secret key can be stored to decrypt the data. After decryption, data is sent to 'Analysis' module of

the framework. In relation with the previous encryption module, data decryption is necessary prior to analysis. Some of the data is not necessary for analysis. A private key is stored in this section to decrypt the sensitive data. The proposed system used public key cryptography technique since it is low cost, highly available for any type of data and low time complexity.

- ***Analysis***

In this module, both sensitive and non-sensitive data is analyzed. Data is properly structured, classified, anatomized (if necessary) before discovering insight knowledge. A wide verity of data analysis is performed in this section depending on the needs which may include 'Frequency Distribution', 'Descriptive statistics', 'Mean comparison', 'Cross-tabulation', 'Correlations', 'Linear regressions' and 'Text analysis'. Analysis can act like a filter which sorts out huge pile of data (*big data*) before reaching to any conclusion. Data analysis can help to sort out further knowledge from data. Thus the analyzed data will be complied into any secure format and delivered to 'Knowledge discovery' section.

- ***Knowledge Discovery***

This module runs different machine learning algorithms or data mining algorithms to discover insight knowledge of a given data set. Based on the discovered knowledge, it can post the data to the associated third party service provider which is govt. services, patients health Services or emergency services.

4.4 PCEHR System DB and Access System

This part has been taken from [6] where authors used homomorpic encryption. Regardless of [6], we preferred to use public key cryptography technique. The rest of the model consists of several sub-modules which describe as follows:

- **PCEHR Patient's Data User Group**

It refers to a person or organizations or a group of persons who required to access patients data. They include general practitioner (GP), specialist doctor, pharmacist, health care provider, insurance, nurse, laboratory, hospital administrators, family members, and friends. The users can be categories in different roles with certain restrictions, such as GP might need to access history data whether laboratory doesn't required. They use patient's data to provide them better health services.

- ***Authentication Server, Access Control List (ACL) Server and Authorisation Server***

This server ensures that all the activity in PCEHR system is legitimate. Every users of PCEHR system is registered and whenever they require accessing PCEHR, they use their own username and password to login into the system. A large verity of algorithm can be associated with authentication server such as challenge response protocol, Kerberos, public key encryption to ensure high level of authenticity. Access Control List (ACL) server ensures a wide verity level of access list which

ensures which users will be accessing which part of the system. It can use different type relationship among subjects, objects and actions. As an example a mental health doctor doesn't have a view access to patient's dental health or an insurance company will not have a write access on patient's data. Thus this allows access to different object with its associated objects. To ensure maximum privacy, authorization server ensures the accessibility of patient's data. If patient provides permission to other users then this authorisation server will retrieve encrypted data which will only be decrypted by patient's private key. This way patient's data remains more secure and accessible only when patient wants to. A patient's profile can be divided into many sub profile including 'Mental profile', 'Sexual Profile', 'Physical Profile'. A patient may not interest to show his/her mental profile to a physical doctor. On the other hand patient may hide all of his/her profiles. Thus, if psychiatrist requires handling patient's mental health, they will ask for patient's permission. When patient provide proper access level permission, psychiatrist will be able to access patient's data though Authorisation server [6].

- ***Operators***

They are usually responsible for operating of PCEHR database system. They must respect the instruction and recommendation (if any) given by PCEHR Jurisdictional Advisory Committee and the PCEHR Independent Advisory Council (2013).

5 Analysis of the Proposed Framework

In this section we present a qualitative analysis to demonstrate how the proposed framework overcomes the *big data* challenges such as storing and processing data, handling and analysing data in real time, preserving privacy and maintaining security. Table 1 shows a comparison of different frameworks. It shows that our proposed model support and overcomes storing and processing, real time handling and analysis and privacy and security challenges. Table 2 shows a comparison of different framework which supports big data platform. From Table 2, it can be illustrated that only our proposed PCEHR system supports *big data* platforms. The following subsections describe how the proposed framework deal with *big data* challenges.

- ***Storing and Processing***

Classification in 'PCEHR Data Processing' is connected with the database server. As we describe previously classification will classify data about which data may need to store in database. For an example a patient's normal blood pressure or normal sugar level may not require to store. In such a case, classification module does not store the data. A data compression and summarization algorithm runs in a regular interval and a modified version of data is stored in database. This helps to increase data storing capability and fastens the search capabilities. Since a smaller version of data is storing regularly thus, data processing capabilities improves.

- ***Real Time Data Handling***

Real time data handling is done by Apache Storm which is integrated with 'PCEHR Data Processing' module. As describe previously, when data analysis is done, data is sent to the Knowledge discovery module. In this module different complex machine learning algorithms runs. Apache Storm guarantees each data will be handled in real time at least once. So our proposed framework supports real time data handling.

- ***Privacy and Security***

From the 'Patient Care' to 'Patient's Data User Group' in every point of our proposed framework ensures that patient's privacy is properly preserved. Sensitive data is encrypted before publish into the PCEHR system. This ensures users data privacy. And decryption happens only in the data processing module before sending to 'Knowledge Discovery' section. The storage data are also encrypted. The decrypted key is only kept securely in data processing unit.

Table 1. Comparison of different frameworks

eHealth systems	Storing and processing	Real time handling and analysis	Privacy and security
PCEHR model in [6]	No	No	Yes
Health care in [7]	No	No	Yes
Health care in [8]	No	No	No
Framework in [9]	No	No	Yes
Health record system in [10]	Yes	Yes	No
Patient safety reporting system in [11]	No	No	Yes
Medical data collection system in [12]	Yes	No	No
Electronic health record system in [13]	No	No	Yes
Personal heath record system in [14]	No	No	No
Health care system in [15]	No	No	No
Health record system in [16]	No	No	No
Our proposed model	Yes	Yes	Yes

Table 2. Comparison of different frameworks

eHealth system	Big data system	Database system
PCEHR model in [6]	No	Yes
MyPHRMachine in [7]	No	Yes
Health care in [8]	No	Yes
Framework in [9]	No	Yes
Health record system in [10]	No	Yes
Patient safety reporting system in [11]	No	Yes
Medical data collection System in [12]	No	Yes
Electronic health record system in [13]	No	Yes
Personal heath record system in [14]	No	Yes
Health care system in [15]	No	Yes
Health record system in [16]	No	Yes
Our proposed model	Yes	No

6 Conclusion and Further Works

In this paper we propose a framework for PCEHR system. Previous sections show that the PCEHR data is growing exponentially. Thus it is very important to consider *big data* scenario when developing an eHealth platform. Based on the motivation, this paper shows how structured and un-structured data are capture, classify sensitive and non-sensitive data, publish, process and gain knowledge to facilitate an individual. The qualitative analysis shows that using this model, *big data* challenges can be overcome. Further researches are being carried out by implementing different privacy preserving algorithms, implementing new and/or existing data mining algorithms and implementing knowledge discovery.

References

1. Philip Chen, C.L., Zhang, C.-Y.: Data-intensive applications, challenges, techniques and technologies: a survey on big data. Inf. Sci. **275**, 314–347 (2014)
2. Savitz, E.: Top 10 strategic technology trends for 2013 (2012). http://www.forbes.com/sites/ericsavitz/2012/10/23/gartner-top-10-strategic-technology-trends-for-2013/. Accessed 14 June 2014
3. Savitz, E.: 10 critical tech trends for the next five years (2012). http://www.forbes.com/sites/ericsavitz/2012/10/22/gartner-10-critical-tech-trends-for-the-next-five-years/. Accessed 14 June 2014

4. Laney, D.: 3D data management: controlling data volume, velocity, and variety. Appl. Deliv. Strat. Meta Group **949**, 1–4 (2001). http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
5. Zhang, X., Yang, C., Nepal, S., Li, C., Dou, W., Chen, J.: A MapReduce based approach of scalable multidimensional anonymization for big data privacy preservation on cloud. In: IEEE Third International Conference on Cloud and Green Computing, pp. 105–112 (2013)
6. Begum, M., Mamun, Q., Kaosar, M.: A privacy-preserving framework for personally controlled electronic health record (PCEHR) system. In: Australian eHealth Informatics and Security Conference, pp. 1–10 (2013)
7. Van Gorp, P., Comuzzi, M., Jahnen, A., Kaymak, U., Middleton, B.: An open platform for personal health record apps with platform-level privacy protection. Comput. Biol. Med. **51**, 12–23 (2014)
8. Ant Ozok, A., Wu, H., Garrido, M., Pronovost, P.J., Gurses, A.P.: Usability and perceived usefulness of personal health records for preventive health care: a case study focusing on patients' and primary care providers' perspectives. Appl. Ergon. **45**, 613–628 (2013)
9. Huang, L.-C., Chu, H.-C., Lien, C.-Y., Hsiao, C.-H., Kao, T.: Privacy preservation and information security protection for patients' portable electronic health records. Comput. Biol. Med. **39**, 743–750 (2009)
10. Barbarito, F., Pinciroli, F., Barone, A., Pizzo, F., Ranza, R., Mason, J., Mazzola, L., Bonacina, S., Marceglia, S.: Implementing the lifelong personal health record in a regionalised health information system: the case of Lombardy, Italy. Comput. Biol. Med. **59**, 1–11 (2013)
11. Lin, C.-C., Shih, C.-L., Liao, H.-H., Wung, C.H.Y.: Learning from Taiwan patient-safety reporting system. Int. J. Med. Inform. **81**, 834–841 (2012)
12. Jian, W.-S., Wen, H.-C., Scholl, J., Shabbir, S.A., Lee, P., Hsu, C.-Y., Li, Y.-C.: The Taiwanese method for providing patients data from multiple hospital EHR systems. J. Biomed. Inform. **44**, 326–332 (2010)
13. Ghazvini, A., Shukur, Z.: Security challenges and success factors of electronic healthcare system. Procedia Technol. **11**, 212–219 (2013)
14. Gordon, P., Camhi, E., Hesse, R., Odlum, M., Schnall, R., Rodriguez, M., Valdez, E.: Bakkenf, S: Processes and outcomes of developing a continuity of care document for use as a personal health record by people living with HIV/AIDS in New York City. Int. J. Med. Inform. **81**, e63–e73 (2012)
15. Lapsia, V., Lamb, K., Yasnoff, W.A.: Where should electronic records for patients be stored? Int. J. Med. Inform. **81**, 821–827 (2012)
16. Yu, P., Zhang, Y., Gong, Y., Zhang, J.: Unintended adverse consequences of introducing electronic health records in residential aged care homes. Int. J. Med. Inform. **82**, 772–788 (2013)
17. Apache Projects: Apache kafka (2014). http://kafka.apache.org. Accessed 20 May 2014
18. Apache Projects: Apache storm (2014). http://storm.incubator.apache.org/. Accessed 20 May 2014
19. IBM: Harness your data resources in healthcare (2010). http://www-01.ibm.com/software/data/bigdata/industry-healthcare.html. Accessed 20 May 2014
20. Austrian Institute of Technology: eHealth platform (2010). http://www.ait.ac.at/research-services/research-services-safety-security/health-information-systems/ehealth-platform/?L=. Accessed 20 May 2014

21. Department of Health: Australian Government: Personally controlled electronic health record system operator: annual report 2012–2013 (2014). http://www.health.gov.au/internet/main/publishing.nsf/Content/PCEHR-system-operator-annual-report2012-2013. Accessed 20 May 2014
22. Department of Health: Australian Government: The personally controlled eHealth record system (2014). http://www.nehta.gov.au/. Accessed 20 May 2014