Electronic Health Record Sharing and Access Controlling Blockchain Architecture using Data De-identification Method

Authored by

Munshi Rejwan Ala Muid, 160041057

Afrin Jubaida, 160041052

Hamim Hamid, 160041047

Supervised by

Dr. Muhammad Mahbub Alam

Professor

Dept. of Computer Science and Engineering

March, 2021

A thesis submitted in partial fulfillment of the requirements for the degree of B.Sc. Engineering from Department of Computer Science and Engineering,

Islamic University of Technology

Declaration of Authorship

This is to certify that the work presented in this thesis is the outcome of the analysis and simulations carried out by **Munshi Rejwan Ala Muid**, **Afrin Jubaida**, and **Hamim Hamid** under the supervision of **Dr. Muhammad Mahbub Alam**, Professor, Department of Computer Science and Engineering (CSE), Islamic University of Technology (IUT), Dhaka, Bangladesh. It is also declared that neither of this thesis nor any part of this thesis has been submitted anywhere else for any degree or diploma. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Authors

_____

Munshi Rejwan Ala Muid, ID:160041057

_____

Afrin Jubaida, ID:160041052

_____

Hamim Hamid, ID:160041047

Supervisor

_____

Dr. Mahbub Alam,

Professor,

Dept. of Computer Science and Engineering,

Islamic University of Technology (IUT)

Acknowledgement

At the outset, we express utmost gratitude to Almighty Allah for His blessings which allowed us to shape this research into reality and give it a form. This thesis owes its existence to a lot of people for their support, encouragement, and guidance. We would like to express our gratitude towards them.

We are very grateful to our supervisor **Dr. Muhammad Mahbub Alam**, Professor, Department of Computer Science and Engineering, Islamic University of Technology (IUT), for his supervision, knowledge and support, which has been invaluable for us.

Finally, we seize this opportunity to express our profound gratitude to our beloved parents for their love and continuous support both spiritually and mentally.

Abstract

The process of digitizing the health records of patients revolutionized the healthcare system as we know it today. Electronic Health Record (EHR)s, the digital records of patients, demand highest security among all data as healthcare organizations have more than 60% higher costs associated with data breaches than the cross-industry average. But they are also needed to be shared among researchers and medical personnel to improve medical service and research.

Main security concern for EHRs is that intruders can delete or tamper them, giving benefits to care-givers by hiding medical malpractices (e.g. misdiagnosis and delayed diagnosis) or care-receivers by creating false data helping them to claim insurance money.

In this paper, we aim to use blockchain beyond the field of cryptocurrency to build a system, such that even after access is provided to researchers or medical personnel, EHRs remain untampered and the identities of patients can't be revealed or traced back. However, this has to be done in such a way that the authenticity of EHRs remain established, and in the event of changing of care team, the new team will have access to complete and unchanged EHRs without hassle.

But we acknowledge the limited storage capacity of the existing blockchain system. We need a system that can relieve huge storage and computation burden from the blockchain system.

We will continue to use traditional cloud storage system to store information while saving the indexes of EHRs in a Blockchain to ensure integrity and increase scalability. We will ensure secure data sharing by applying random selection of data de-identification methods that guarantees integrity, security and trackability of user data without revealing user identity. Lastly, we will present a complete architecture of the proposed system with two levels of blockchain that ensures secure and efficient implementation of EHR.

Contents

Chapter 1

Introduction

1.1 Introduction

1.1.1 Overview

Electronic Medical Record (EMR) systems are thought to play a pivotal role in enhancing healthcare intelligence, quality, user experience, and costs. The EMR system could save billions of dollars each year in the long run. Sharing healthcare data would enable a smarter, better understanding of patterns and trends in public health and illnesses, resulting in improved quality services, better doctor recommendations [10], and service plans that make the most of limited national health service budgets for the health and wellbeing of the public [12]. We refer to patient data as healthcare data, and healthcare data systems as any system that collects, accesses, or stores patient data.

Cloud computing environments provide a good chance to successfully accommodate e-Health services in various situations [7]. Because of its scalability and mobility, this climate can provide various advantages, but there are certain obstacles that must be overcome [18]. A cloud-based Electronic Health Record (EHR) management system will offer two significant benefits: 1) the ability to share patient records with other clinical centers [7], and 2) the integration of all of a group of clinical centers' EHRs to aid medical staff in their work [25].

Despite the numerous advantages of cloud-based EHR management systems, security is a major concern. Abuse, leakage, loss, or theft of EHR in cloud-based management systems

are all possibilities. Intruders will, for example, delete or tamper with EHRs to tamper with therapies that give insurance companies benefits or hide medical malpractices (e.g. misdiagnosis and delayed diagnosis) [1]. Health insurance and electronic health records are inextricably linked. Dishonest health insurance companies may recruit hackers to remove or tamper with patients' electronic health records (EHRs) in order to prove the existence of pre-existing conditions [25]. Medical malpractice claims are common for a variety of reasons, including misdiagnosis and delayed diagnosis. Due to these concerns, patients are unable to prove medical malpractice in the majority of cases. Patients, on the other hand, alter medical records to obtain financial benefits despite having pre-existing medical conditions. Several cryptographic countermeasures are suggested to ensure the security of EHRs [1]. Unfortunately, because of the centralized features, security threats continue to exist. There are security concerns in a cloud-based healthcare system.

A healthcare blockchain is generally thought of as a distributed ledger that stores health records for sharing, exchanging, and other purposes among stakeholders[19]. Data for e-Health systems may come from a variety of places, including clinics, hospitals, and pathologies. All patient data is stored in a distributed ledger provided by a blockchain network in a blockchain-based EHR management system [7]. A transaction is a method of storing a group of associated data. Before being stored in the distributed ledger, each transaction is analyzed by a group of participants known as miners [16]. An illegal transaction that attempts to alter data in a distributed ledger would be rejected by blockchain networks. As a result, no one with access to a blockchain network can alter the data [16]. Smartcontracts, a key concept in blockchain, enable trustless features among various entities in the EHR management system. A smart contract is a computer program that includes a set of rules and agreements. These agreements and principles must be followed by all network participants. As a result, there is no need for a trusted third party to store data on the blockchain.

Large amounts of clinical research data are held by pharmaceutical firms, academic researchers, and government institutions such as the Food and Drug Administration and the National Institutes of Health [10]. The resulting research advances obtained from data pooling and analysis could improve public health, promote patient safety, and spur drug development if these data were shared more broadly within and across industries. By in-

creasing transparency in the clinical research process, data sharing can also increase public confidence in clinical trials and the conclusions drawn from them. However, much of this information is never shared. Clinical research data retention by investigators and organizations may result in missed opportunities in biomedical research. Despite the potential advantages of pooling and analyzing shared data, industry researchers face obstacles to data sharing, including concerns about data mining, erroneous secondary analyses, and unwarranted lawsuits, as well as a willingness to protect confidential commercial information [20]. Academic partners face important cultural obstacles to sharing data and engaging in longer-term collaborative endeavors, which stem from a desire to protect intellectual autonomy and a career advancement system based on publishing and citation requirements taking precedence. Some obstacles, such as the need to protect patient privacy, posed problems for both industries [**Kuo**]. In the future, there will be a variety of technological difficulties in analyzing potentially large and heterogeneous datasets [20]. Fortunately, recently discovered data-deidentification method can relieve the system from this difficulty.

Organizations would use de-identification to erase personal information from data they retain, use, archive, and share with other organizations. De-identification is a collection of techniques, algorithms, and techniques that can be used to de-identify various types of data with varying degrees of success. In general, as more aggressive de-identification strategies are used, privacy protection increases, but the dataset's usefulness decreases. De-identification is critical for government agencies, enterprises, and other institutions that want to share data with the public. The sharing of de-identified patient information under the framework created by the Health Insurance Portability and Accountability Act (HIPAA) Privacy Ruling, for example, allows for substantive medical research with social benefits.

As long as the data obtained from personal information has any usefulness, there is a chance that some information will be related back to the original people on which the data was based, however distant [20]. When de-identified data may be re-identified, de-privacy identification's safeguard is lost. Because the likelihood of re-identification can be difficult to quantify, the decision of how or if to de-identify data should be made in conjunction with the decision of how the de-identified data will be used, shared, or

published. Individual risks may persist in de-identified data. Allowing inferences about people in the data without re-identification, as well as effects on groups depicted in the data, are among the dangers [20].

According to the HIPAA Privacy Rule, if data has been de-identified, covered organizations are free to use or reveal it without restriction. The information is no longer called PHI, and it is no longer subject to the same rules and regulations that apply to PHI. It's important to remember that UMMS isn't a covered entity, so it can't release de-identified information [25].

1.1.2 Statistical Representation

According to a Ponemon Institute report, sponsored by IBM, total amount of healthcare records being compromised in the year 2019 was 41.2 million. Data breaches cost of healthcare was 6.5 million dollars in 2019, over 60 percent more than all other sectors together. Also, healthcare Industry have decided to spend 125 Billion dollars on internet security over the period 2020 to 2025. [1]

1.1.3 Problem Definition

We discover three barriers to blockchain adoption in cloud-based EHRs. To begin, blockchain adoption should remove a central authority's control over data repositories. To put it another way, the data should be decartelized to the greatest extent possible. As a result, tampering with data in the blockchain network becomes difficult. As a result, selecting the right blockchain network for EHR management systems is critical [25], [16]. Second, blockchain technology uses a platform that is distinct from conventional systems. As a result, creating a blockchain-based system necessitates starting from scratch with e-sign. Designing a system from the ground up takes time, money, and has an impact on current stakeholders. Also we need a security mechanism to protect the sensitive information that will be shared for research and medical purposes. In conclusion, we attempt to answer the following questions regarding the design of blockchain-based EHR management systems:

1. How can a blockchain network be integrated with a conventional cloud-based EHR

management system without affecting existing stakeholders?

2. How can blockchain networks be chosen in such a way that data control is protected?

3. How to provide privacy to the data and individuals the data belong to?

4. How to design the overall architecture that can provide all the facilities?

1.1.4 Contribution

We suggest a blockchain-based system architecture to implement a tamper-proof EHR management system in this paper. So far, there are three types of blockchain platforms: public blockchain, consortium blockchain, and private blockchain. Anyone can participate in the consensus process on the public blockchain. As a result, a specific organization entering the public blockchain has no control over the consensus activities, implying that control is decentralized. Bitcoin is an example of a public blockchain. In contrast, only a small number of pre-selected nodes will execute consensus tasks in the consortium blockchain. Each group of companies building a blockchain network nominates one or more nodes as consensus nodes, as an example of the consortium blockchain. Only certified nodes of an organization can execute consensus tasks in a private blockchain. As a result, the organization will control all of the nodes that are responsible for consensus. In this paper, we use 2 layer of the public blockchain network as our blockchain to create a tamper-proof EHR system. It's a design challenge to integrate the blockchain network with existing cloud-based EHR systems. A suitable design approach must be chosen that enables blockchain technology to be integrated while ensuring that existing stakeholders are not harmed. We'd like to use a multi-phase approach for combining blockchain with cloud-based EHR management systems in our system design. We create a prototype of a blockchain-based EHR management system using Hyperledger, a public blockchain network, to demonstrate the feasibility of integrating blockchain with conventional cloud-based EHR management systems. The following are the principal contributions of our work areas:

1. Using blockchain technology, a tamper-proof cloud-based EHR management system is proposed.

2. A efficient architecture for combining blockchain with existing cloud-based EHR management systems is proposed.

3. Introduces the concept of two-layer blockchain that facilitates privacy, security and scalability.

4. Each component in the proposed architecture is discussed in terms of its functionality.

5. A randomly selected de-identification method is proposed to provide identity protection and security during data sharing.

Chapter 2

Background Study

2.1 Related Terminologies

2.1.1 Blockchain

Blockchain is considered to be a chain of tamperproof, decentralized and chronological blocks. These blocks can contain various kind of information depending on the purpose of the blockchain [16]. Generally, blockchain can be divided into 4 categories: Public, private, hybrid and consortium.

Public or permissionless blockchain operates without any authority. Here anyone can join and apply to be a validator. These kind of blockchains work on Proof of Stake or Proof of Work algorithms [16]. Popular examples include economic applications like Ethereum and Bitcoin blockchain.

Private or permissioned blockchain are opposite of public one as every participants need approval of system administrator to gain membership or to be a validator. It was created to satisfy demands of business companies that needed security of blockchain yet needed to impose few control.

Blocks

Blocks are the record of hashed valid transactions that are encrypted into a Merkle tree. Each block contains the hash value of the previous block that makes the chain while preserving integrity. From each block we can iteratively track down the first block, called genesis block. Separate blocks may also be created at the same time, resulting in a

temporary fork. Any blockchain, in addition to a stable hash-based history, has a specific algorithm for ranking various versions of the history so that the one with the highest score will be chosen over the others. Orphan blocks are those that were not chosen for use in the sequence [12].

Smart Contracts

Smart contracts are scalable pieces of programs that run on each block of blockchain to establish few general rules for blockchain system without manual checking. Uses of AI, Reinforcement Learning, Advanced Statistical Algorithms are very common in smart contracts now a days; but their should be a trusted middle party or authority to impose or determine the rules for smart contract. The oldest piece of hardware similar to smart contract deployment is vending machines. The white paper on cryptocurrencies from 2014. The Bitcoin protocol is defined by Ethereum as a "weak" variant of the smart contract principle. Various cryptocurrencies have embraced scripting languages since Ethereum, allowing for more sophisticated smart contracts between untrustworthy parties.

Proof of work (PoW)

Proof of work is a cryptographic system where the verifier can keep track of how much computational(time, storage, unit) effort was put in for a specific task with verifier spending minimum possible resources [12]. It was invented to tackle security hazards like denial-of-service , spam etc. It is used by Bitcoin to solve the miner candidacy problem that is often adapted by other blockchain systems.

2.1.2 Electronic Health Records

Healthcare institutions having this huge responsibility of managing so many terabytes of data on their patients properly and efficiently, adopt EMR and EHR systems which are very efficient in providing easy access to their data and share them when needed [10]. So, it's important that they learn about the implementation of these systems properly. It helps the medical personnel to organize administrative and professional activities with the help of the information which is stored in the EHR system.

Electronic health record (EHR) is the digital collection of patient data that can be shared among different health care providers and researchers. EHRs include but are not limited to demographics, history, prescription and aversions, vaccination status, laboratory

test, radioscopy, vital signs, personal information and billing information.

The electronic health record (EHR) is an electronic record of patient health information generated by one/more sitting in any medical facility. EHR includes patient symptoms, progress report, suggested medicines, initial diagnosis, past records, vaccinations, lab reports etc. The EHRs mechanizes medical personnel's workflow. Many countries including USA, UK, Canada, Japan, China, Malaysia, Taiwan, New Zealand have taken steps on digitalized health information.

Several surveys have cast doubt on whether EHRs increase treatment quality. The New England Journal of Medicine conducted a report in 2011 that found evidence that procedures with EHR offered higher clinical treatment. EMRs have the potential to enhance patient planning in the future. According to a trade journal article, since someone accessing an EMR will see the patient's entire report, it eliminates guessing history, frequent specialist visits, smooths transitions between treatment environments, and can make for improved emergency care.

EHR adoption is hampered by the high cost of EHR and provider confusion about the benefit they will gain from adoption in terms of return on investment.Hospital executives and physicians who had introduced EHR recognized that any improvements in reliability were tempered by decreased effectiveness when the technology was applied, as well as the need to expand information technology personnel to support the infrastructure, according to a study conducted by ONC.

The use of an electronic medical record (EMR) will theoretically shorten the time it takes for patients to be found before they arrive at the hospital. According to a report conducted in the Annals of Internal Medicine, since the introduction of EMR, there has been a 65 percent reduction in time (from 130 to 46 hours).

Physicians are increasingly implementing mobile devices such as smartphones and tablets. According to a 2012 survey commissioned by Physicians Practice, 62.6 percent of respondents (1,369 physicians, hospital administrators, and other healthcare providers) say they use mobile devices in their work. When handheld devices grow more capable of syncing with electronic health record networks, doctors may be able to view medical information from afar. The majority of computers are extensions of desk-top EHR systems that connect and view files remotely using a range of applications. The benefits

of getting immediate access to medical information at any time and from any venue are apparent, but they come with a slew of security issues. Practices may require robust laws that regulate security protocols and patient privacy legislation as mobile systems become more common.

The willingness of the adopter to completely comprehend workflow and predict future clinical processes prior to adoption is critical to the effectiveness of eHealth initiatives. Failure to do so will result in expensive and time-consuming service interruptions.

Privacy concerns

In the European Union (EU), a new directly-binding instrument, the General Data Protection Legislation, was enacted in 2016 by the European Parliament and Council and will take effect in 2018 to secure the collection of personal data, particularly those for health-care purposes.

Threats to health care information can be categorized under three headings:

- Human threats, such as employees or hackers
- Natural and environmental threats, such as earthquakes, hurricanes and fires.
- Technology failures, such as a system crashing

Internal, social, deliberate, and accidental attacks are both possibilities. As a result, when considering ways to protect patients' health records, health information management experts will bear these risks in mind. In countries like Spain, security consciousness among health care practitioners has been discovered to be lacking. The Health Insurance Portability and Transparency Act (HIPAA) has created a mechanism to minimize the damage caused by these risks that is robust but not so narrow as to restrict the solutions open to healthcare practitioners with access to multiple technologies.

Legal issues

Failure or losses incurred during the implementation or use of an EHR device have been viewed as a legal hazard. Similarly, it's important to realize that adopting electronic health reports entails substantial legal threats.Small EHR machine vendors were especially worried about liability. Because of the geographic liability environment, certain smaller businesses could be required to leave markets. Larger EHR providers (or government-sponsored EHR providers) are more able to survive legal assaults.

While electronic recording of medical records enhances patient treatment, there is growing concern that such documentation can subject physicians to further malpractice litigation. Disabling physician notifications, choosing from dropdown menus, and using templates can enable physicians to skip a detailed analysis of past medical history and prescriptions, resulting in the omission of essential details.

Electronic time stamps are another possible problem. Many doctors are unaware that any time a medical record is changed, EHR devices generate an electronic time stamp. If a malpractice lawsuit goes to arbitration, the prosecution may seek a full list of all entries made in a patient's electronic record via the discovery process. Waiting for the end of the day to chart patient reports and making addendums to documents after the patient appointment may be troublesome, as it can result in incorrect patient details or suggest a potential intent to inappropriately change the patient's record.

In certain communities, hospitals aim to standardize EHR programs by offering local healthcare facilities cheaper copies of the hospital's applications [1]. This practice has been challenged as a violation of Stark laws, which forbid hospitals from supporting community healthcare providers first.

2.1.3 Cloud

Cloud computing is the on-demand accessibility of computer structure resources like storage capacity in form of google drive, iCloud etc . or computational power like google collab provides GPU. But none of these are owned or controlled by the user directly, rather these are leased/used on contractual basis [**Kuo**].

This system was invented to make use of the idle space of vast majority of computer users so that people don't have to buy an entire device to solve a singular purpose. But now it has turned into a multi-billion-dollar industry standing on servers dispersed over multiple sites from dominant servers. Servers are also classified as edge server, fog server, central server etc based on their placement and functionality [20].

Clouds are classified into enterprise clouds that are y be limited to a single organization or public cloud that are available to multiple organizations. Cloud computing permits companies faster running applications, with better manageableness and less preservation. It also allows IT experts to more efficiently utilize resources so that oscillating and impul-

sive demands can be met, or simply said- providing burst computing proficiency [25].

Service models of cloud can be categorized into 6 instances knows as Infrastructure as a service (IaaS), Platform as a service (PaaS), Software as a service (SaaS), Mobile "backend" as a service (MBaaS), Serverless computing, Function as a service (FaaS).

Private cloud

Private cloud refers to cloud infrastructure that is run exclusively for the benefit of a particular entity, whether it is managed internally or by a third party, and is hosted internally or externally. A private cloud initiative necessitates substantial involvement in order to virtualize the business climate, as well as a reevaluation of current resource decisions. It can boost profits, but every step of the process poses security concerns that must be resolved to avoid serious flaws. Self-contained data centers need a significant amount of money. They have a large physical footprint, necessitating space, hardware, and environmental controls allocations. These assets must be refreshed on a regular basis, which necessitates additional capital expenditures. They've been panned because consumers "still have to purchase, create, and maintain them" and therefore don't benefit from less hands-on management, basically "[lacking] the economic model that makes cloud computing such an interesting idea."

Public cloud

When cloud services are distributed over the public Internet, they are called "public," and they can be charged or free. There are few architectural differences between public and private cloud services, but when services (applications, storage, and other resources) are used by several users, security issues grow significantly. The majority of public cloud vendors offer direct-connection services, which allow customers to securely link their legacy data centers to cloud-based applications. Several considerations affect whether businesses and companies prefer a public cloud or on-premise approach, including solution capability, cost, integrational and operational dimensions, as well as safety and security.

Hybrid cloud

The term "hybrid cloud" refers to a mixture of a public cloud and a private environment, such as a private cloud or on-premises resources, that are different but connected to provide the advantages of multiple implementation models. The ability to connect collocation, managed, and/or delegated services with cloud resources is referred to as hybrid

cloud. A hybrid cloud infrastructure is a cloud computing service that integrates commercial, public, and group cloud services from a number of service providers. A hybrid cloud service spans separation and provider borders, making it impossible to categorize it as either private, public, or group cloud. It helps you to improve a cloud service's capacity or capability by aggregating, merging, or customizing it with another cloud service.

2.1.4 De-identification method

The method of de-identification is used to keep a person's personal identity hidden. To secure the identity of study subjects, data obtained during human subject research can be de-identified. To comply with HIPAA rules, which identify and stipulate patient privacy laws, biological data can be de-identified.

The method is also known as data anonymization when it is extended to metadata or general data regarding identity. Remove or cover personal identifiers, such as a person's name, and suppress or generalize quasi-identifiers, such as a person's date of birth, are popular techniques.Data re-identification is the method of using de-identified data to re-identify persons. Re-identifications cast doubt on the effectiveness of de-identification. One of the most common approaches to data privacy security is de-identification. Communications, graphics, biometrics, big data, data analytics, the internet, social networks, and video monitoring all use it.

Techniques

Masking personal identifiers and generalizing quasi-identifiers are two traditional de-identification techniques. Pseudonymization is the most common method for masking personal identities from data records, while k-anonymization is used to generalize quasi-identifiers.

seudonymization

The process of pseudonymization entails replacing actual names with a temporary identifier. It hides or obscures personal information in order to keep users invisible. And if the database will be changed, this approach allows you to follow the individual's record over time. However, if any unique combinations of attributes in a data record implicitly identify an object, it cannot be prevented.

k-anonymization

k-anonymization determines quasi-identifiers (QIs) as attributes that implicitly point to an individual's identity and deals with data by ensuring that at least k entities have the same combination of QI values. The handling of QI principles is governed by a set of guidelines. The k-anonymization, for example, replaces some original data in documents with new range values while leaving others untouched.

De-identification laws De-identification was considered to be "somewhat helpful as an added precaution" by the United States President's Council of Advisors on Science and Technology in May 2014, but not "a useful base for strategy" because "it is not resilient against near-term potential reidentification approaches." The HIPAA Privacy Rule defines mechanisms for safely using and disclosing health data without needing patient consent. Two HIPAA de-identification criteria – Safe Harbor and the Expert Determination Process – are at the core of these processes. Safe harbor depends on the elimination of unique patient identifiers (such as names), while the Expert Determination Approach necessitates expertise and familiarity with widely recognized statistical and scientific concepts and methods in order to make data non-identifiable.

Safe harbor

The safe harbor method uses a list approach to de-identification and has two requirements:

- The removal or generalization of 18 elements from the data.
- That the Covered Entity or Business Associate does not have actual knowledge that the residual information in the data could be used alone, or in combination with other information, to identify an individual. Safe Harbor is a highly prescriptive approach to de-identification. Under this method, all dates must be generalized to year and zip codes reduced to three digits. The same approach is used on the data regardless of the context. Even if the information is to be shared with a trusted researcher who wishes to analyze the data for seasonal variations in acute respiratory cases and, thus, requires the month of hospital admission, this information cannot be provided; only the year of admission would be retained.

Expert Determination

Expert Determination uses a risk-based approach to de-identification, determining the

probability that an individual will be detected from their protected health information using existing standards and best practices from the study. This approach necessitates that the details be made non-identifiable by a person with adequate expertise and experience with commonly accepted statistical and scientific concepts and methods. It calls for the following:

- That the risk is very small that the information could be used alone, or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information;
- Documents the methods and results of the analysis that justify such a determination.

2.2 Literature Review

We present 4 papers related to our work. First 3 presented here are related to modification of blockchain scalability to accommodate EHRs system's demand.

2.2.1 BPDS: A Blockchain based Privacy-Preserving Data Sharing for Electronic Medical Records [25]

Problem Statement:

- Electronic medical records (EMR) contain high privacy-sensitive information; thus, data leakage could hurt patients' reputation and finances.
- Majority of private clinics don't facilitate data sharing leading to further patient cost for treatment.
- A central architecture comes with issues like single-point-of-failure and random modification attacks.
- The interoperability between healthcare institutions still remains a severe challenge.

Proposed Solutions:

- The EMRs to be stored in cloud while the indexes of EMRs are to be stored in a consortium blockchain.

- Upgraded delegated proof of stake consensus
- Content extraction signature (CES) technique allows removal of sensitive data by patient
- Patients possess the full control over their specified EMRs
- Consortium blockchain allows implementation of desired architecture System

Architecture:

BPDS is a 3-layer design that has data acquirement layer, data storage layer and data sharing layer. The function of each layer is described as follows.

Data Acquirement Layer:EMRs are created and signed by data providers using a CES scheme. Valid extraction signatures of data receivers will be generated by removing sensitive information.

Data Storage Layer: Stores the EMRs and indexes. Components of data storage layer include: I. Cloud Storage which accommodates patients' encrypted EMRs, the extraction signature and data access records. It also generates url of EMR and timestamp as output. II. Consortium Blockchain Network which is used to reserve indexes of EMRs and predefined access permissions in the smart contracts.

Data Sharing Layer: In this layer, request for patients' EMRs will be accepted.

Workflow of the BPDS system:

- During patient-doctor appointment, medical personnel generate EHRs of patient.
- After receiving the EMRs, Patient sends the EMRs to the cloud and stores the indexes in consortium blockchain along names of authorized personnel (U).
- System Setup, Data Acquiring, Data Storing, Data Release and Data Sharing- these five stages will be followed then to achieve privacy preserving EMRs sharing.

2.2.2 A Novel Architecture for Tamper Proof Electronic Health Record Management System using Blockchain Wrapper [19]

Problem Statement:

- Storing electronic health information or EHR data in cloud comes with threat to security. Data can be tampered to give insurance companies or doctors or patient illegal benefits.
- Centralized characteristics of cloud-based systems always may have potential security threats.
- Replacing entire cloud-based system with blockchain architecture will create a tamperproof facility but it will be costly in terms of time and money.

Challenges:

- How to adapt blockchain with minimized cost?
- Choosing appropriate blockchain for decentralized architecture.

Solution:

- Integration of blockchain to pre-established cloud-based systems with the help of a novel architecture.
- The concept of blockchain handshaker which will provide the facilities of blockchain wrapper to support blockchain integration.
- A prototype is shown using Ethereum public blockchain to show the feasibility of proposed systems.

Proposed System Architecture:

User Application: Provides applications interfaces for users (doctors, nurses, system administrators, pathologists, etc.) and builds an initial transaction.

Blockchain Handshaker: Connects all other components together. It has three sub components:

- Transaction template manager (TTM): Comprises of predefined transaction templates.
- Transaction generator (TG): Builds transactions(Tc) with help of TTM from (TI).

- Transaction validator (TV): Receives initial transaction (T)I from user application, sends it to TG and receives blockchain transaction TC from TG. Next TV sends the TC to the blockchain network for validation. The blockchain network returns validated transaction T'I.

Public Blockchain Network: An open blockchain that validates the transaction with binary output.

Cloud: Host the traditional cloud-based EHR management system and works as the storage service. Receives TI' from TV and stores it.

Workflow of The Architecture:

- User app sends initial transaction TI (general health record) to BH for validation from blockchain.
- Using TI, BH generates blockchain transaction (Tc) and sends for validation.
- After validation, blockchain sends validation ack. to the BH and mines to add transaction data into the blockchain.
- BH sends validated T'I to the cloud and stores the data in cloud.

2.2.3 Scalable Architecture for sharing HER using the Hyperledger Blockchain [7]

Problem Statement:

- Though blockchain provides us immutability, anonymity, decentralization but scalability is a problem in terms of storage size.
- Brazil had 1.4 billion patient visits in 2018 while china had 7 billion. So, if we consider each visit to be one transaction, it will be a challenge for blockchain network to accommodate all of them.

Solution:

- Scalable model for sharing EHR records using multi-channel hyperledger blockchain.
- One blockchain to record patient visits and one blockchain to link patients to health institution.

Proposed System Architecture:

Four connected components work together.

Client component: An electronic device considered as real-world entity(patient/ physician/ hospital/ clinic). It sends/receives EHR and provides access permission of its records to other.

CA component: Creating public and private keys and delivers them client component for storing. It also validates the entity that wants to use the architecture. Three different CA's are there:

- Patient CA: Validates patients before delivering the keys. Patients are still anonymous. It can be deployed by ministry of health or by health institution.
- Professional and Institution CA: Validates medical personnel and medical institution respectively. It can be deployed by each state.

Storage Component: Provides external storage system for EHR.

Hyperledger component: Comprises of two different blockchains: • Global blockchain: Records all patients visits to a particular health institution. Each health institution must have a committing peer that stores the chain. This will record one associated regardless multiple visits. • Local blockchain: Records all EHRs associated to its patients. Each health institution must have a minimal infrastructure to run a hyperledger network.

Components Interaction:

Three common usage scenarios. For all the scenarios, the initials are assumed: Scenarios:

- A person (not registered in the system) visits physician m1 hospital h1.At first he registers himself in global Certification Authority obtaining identifier p1, public and private keys. Then he requests read only access to hospital' Patient CAh1. Then, m1 treats the patient, creates an EHR using institutions HIS, visit metadata stored in b1 (local blockchain). Finally p1 visited h1 is registered in bwc.
- p1 wants to perform any electrocardiogram at h2 with m1 Here m1 works in h1 and h2 and have access to both b1 and b2. At first, m1 requests to global blockchain, bwc for the institutions p1 visited and gets h1 as a result. Then h1 requests to b1 for

documents associated with p1 and gets link to ehr1. M1 gives p1 ehr2 and p1 sends ehr2 to its personal storage and physician register in b2 at h2. Lastly p1 request a read-only access to hospital patient CA CAh2.

- p1 visits h3 and helped by a nurse m2 who only works at hospital h3. First m2 requests to bwc for the institutions p1 visited and gets h1 and h2. Then m2 request the credential to p1 and access code in order to access the EHRs stored in hospital and heart clinic's HIS. This code is encrypted with p1's private key. The nurse using the code, obtains EHRs from h1 and h2. And decrypt with p1's public key. Later the nurse creates ehr3 and HIS register the metadata in b3. HIS register bwc that p1 visited h3. Patient request read-only access to clinic Patient CA CAh3. 4th Related paper gives us idea about ABS.

2.2.4 A Decentralizing Attribute-Based Signature for Healthcare Blockchain [20]

Problem Statement:

- Blockchain- centered information distribution applications requires preserving facility with dual capabilities.
- The users demand the verification of authenticity of the EHR data and the identification of the signer. But at the same time, the signer wants his real identification to be secret to stop tracing and inferring of his identification data.
- However, traditional blockchain systems, such as the one used for Bitcoin use pseudonyms as public keys. This system cannot satisfy the demand of privacy preserving veri[U+FB01]cation as described as real identity can be guessed from the series of transactions through inference attacks.

Challenges:

To employ the attribute-based signature (ABS) scheme in a healthcare blockchain, two important technical challenges need to be addressed:

- Attribute certificates being issued from different authority demands a centralized authority to supervise. So, how that can be actualized in blockchain is a question.

- Storage scalability issues of block-chain.

Proposed Solution:

- A decentralizing attribute-based signature (called DABS) technique is proposed; that provides with dual characteristics privacy without needing centralized agency.
- A blockchain-centered EHR storage system is proposed with on-chain and off-chain collaboration storage model to address storage scalability.
- Finally, official security study of the DABS verification protocol is supplied.

System Model:

1. On-Chain and Off-Chain Storage Model: The addresses of the EHR data are stored on the blockchain while EHR data itself is kept node associated with blocks after encryption.
2. Roles: There are mainly three roles in our protocol:
    - User: includes doctor, patient and researcher.
    - Authority Agency: They responsible for issuing the signature keys related with the attributes to the users.
    - Administrator: generates a global public parameter GP when the system is initialized and it assigns a global identity GID to each user entering the system. Administrator also manages EHR data.
3. Nodes: Nodes are divided into two categories: primary node and backup node.
    - Primary Node: They collect a group of transactions broadcasted on the network into one block, creating a new block.
    - Backup Node: They can create new transactions and publish them to the network. If they satisfy the access policy, they can verify the signature of the transaction.
4. Data Structure: A block includes a block header and a main block.

- The block header is made up of a version number, a previous block hash, a current block hash, timestamp, nonce and so on.
- The main block consists of a series of transactions. There are four kinds of data structure: SignedEHR, ProposalRecord, BlockHeader and MainBlock.

Workflow of the proposed protocol:

- Only doctors can create EHR data and put the address of EHR data on the blockchain.
- After an EHR data is created, doctor signs the data and broadcasts the address of it to the blockchain with his signature.
- The signed EHR data is stored in an off-chain database.
- Patients and other users in this system are verifiers. They can only access the data on the blockchain, and verify the signer's attributes. They cannot broadcast EHR-related data to the blockchain.

Chapter 3

Problem Definition

For the establishment of proper EHR system through blockchain, we need a complete architecture which will ensure:

- Storing of patient's EHR created by doctor in a confidential way, so that the patient's EHR is not leaked.
- Proper identity protection method for the patient. The patient's sensitive information remains hidden.
- A system is needed to track if shared data is misused. Or the shared EHR is again shared to some other unauthorized third parties other than the authorized one.

Chapter 4

Proposed Architecture

4.1 Overview

So a complete architecture is proposed which solve all the problems mentioned in our problem statement.

4.1.1 Figure

The next figure depicts the workflow and architecture of our proposed system:

Descriptions:

- Patient goes to doctor for any treatment.
- The doctor creates the EHR for the patient and sends the EHR to the patient after signing it digitally. This digital signature is for verifying or authenticating the doctor.
- After the patient decrypts the signed EHR and stores the raw EHR in the private cloud of herself.
- Now the patient stores the URL of the stored EHR, the hash of the EHR, the timestamp when the EHR was stored in B1.
- Suppose a third party (Doctor, Researcher, Company) requests the patient for her EHR. The request was made upon the national identification no. of the patient.
- The patient retrieves the requested EHR from her private cloud

- Now she passes the retrieved EHR for de-identification.

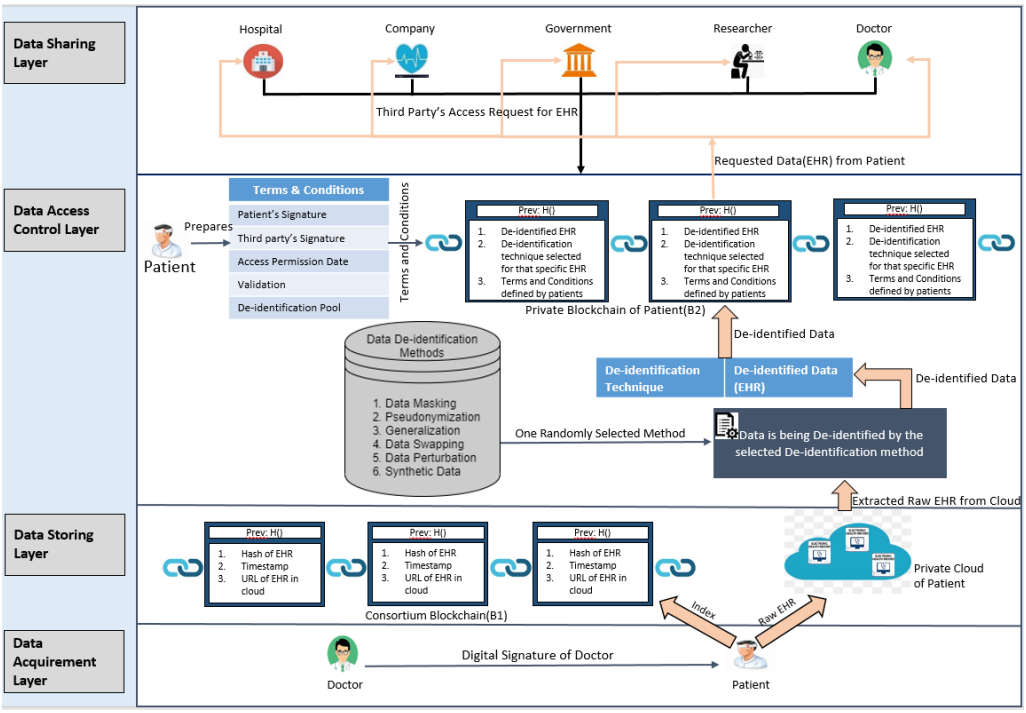


Figure 4.1: Proposed architecture

- One de-identification technique is randomly selected from the de-identification pool and applied to the retrieved EHR. So the EHR is now de-identified which means sensitive informations are hidden.

- The patient creates terms and condtions on sharing the de-identified EHR. The terms and conditions may include patient's signature, third party's signature, access permission date, validation and many things.

- Now the patient puts the de-identified EHR, de-identification technique that was applied on that particular EHR, the terms and conditions in B2.

- Now the requesting third party reads or accesses the de-identified EHR from B2.

4.1.2 Notations

Notations	Descriptions
$B_{1,2}$	Blockchain
$P_{1..n}$	Patient
$D_{1..n}$	Doctor
$T_{1..2}$	Third Party
H	Hash
$T_{1..2}$	no. of transactions
I	Index
$m_1..m_7$	Medical record, M

These are the notations that will be used throughout our whole book in case of necessity to make our writings more easier.

4.2 Architecture Explanation Details

From the figure previously shown, we will describe about the stake holders, functional properties, blockchain architecture of our proposed architecture along with the related terms and properties.

4.2.1 Stake holders

Stake holders are those with an interest or concern in our proposed architecture. So the main stake holders of our system are: 1) Patients 2) Doctors 3) Third parties (Hospitals, Government, Researchers, etc)

- **Patient:** Patients play an active and participative role as a real-world entity in the healthcare context as well as in our proposed architecture. Patient is mainly responsible for: (i) receiving electronic health record from the doctor; (ii) storing the electronic health record in the cloud; (iii) putting the index of the health record in the B_1 (iv) Receiving request from third parties(doctor, researcher, government, hospital); (v) Giving access of the electronic health record to the third parties upon some terms and policies and storing all terms policies along with the de-identified data to B_2. Besides one patient stores the de-identification technique that has been

applied to the EHR coming from the cloud in her local database for future checking. Patients reside in the core of our architecture which is data storing and access control layer.

- **Doctor:** Doctors are responsible for creating the electronic health record and sending the EHR to the corresponding patient after signing it digitally. Doctor may require the EHR in future for checking the EHR of the particular patient and also for any survey or research purpose. For this, doctors send access request for any particular electronic health record(EHR) stored in the cloud to that specific patient and that particular patient gives the records to the doctor in return along with some terms and policies for making the data invulnerable. So doctors will be active in both data acquirement and data sharing layer.

- **Third Parties:** Third parties may include hospitals, healthcare companies, government, researchers and doctors. Hospitals and doctors will need EHR from the patient to see the previous medical records of that particular patient. Healthcare companies, government and researchers will request to the patients for EHR for carrying out a research based on data. So third parties are placed in the data sharing layer of our proposed architecture. They will have their requested data from B_2 which has been provided by the patients.

4.2.2 Functional Properties

Our proposed system can be considered as four-layer architecture and the layers are data acquirement layer, data storage layer, data access control layer and data sharing layer. The detailed functionality of each layer is mentioned below:

- **Data acquirement layer:** In this layer, data providers or doctors will create a raw Electronic Health Record (EHR) for each registered patient. In the meantime, doctors generate their private/public key pair. Hash of the raw EHR will be created by using any hash algorithm and doctors will sign the hash of the EHR using their private key followed by an encryption mechanism. After that, doctors send this digitally signed EHR to their respective patients. Upon receiving EHR from doctors,

patient can verify the authenticity of EHR by using doctor's public key. At that time, patients will be the owners of their EHRs and will have complete control over them.

- **Data storage layer:** The function of this layer is to store the original EHR and its indexes. Components of data storage layer include:

  - **Cloud Storage:** Patients will store their raw EHR in their private cloud and from there they will get the location (url) of where the EHR has been stored along with a timestamp as output.
  - **Private Blockchain Network :** Each patient will have a private blockchain (B_1) which will be used to reserve the indexes of EHRs and to make the data sharing process effective. Indexes of EHR will contain the cloud storage location (url) where the data has already been stored, hash of the original EHR and timestamp obtained from the cloud.

- **Data access control layer:** The function of this layer is to ensure the permission of the access when any third party requests for EHR to the patients. Third party can request any patient for their EHR based on a globally identifiable number of patient which can be patient's NID or medical record number etc. When a patient will get any access request, he/she will first search for the cloud storage location of that data from his/her private blockchain. Then from that location he/she will retrieve the stored raw EHR and apply a de-identification technique, which will be chosen randomly, to the EHR as well as he/she will set some terms and conditions for accessing the data by any third party. A block will be created consisting of the de-identified data and applied de-identification technique with the selected tems and conditions in another private blockchain(B_2) of patient. Then the requested third party will have the permission to access that block if all terms and conditions meet. Here terms and conditions can be access permission date for third party, validation period of accessing the shared EHR, sign of patient as well as third party and the timestamp.

- **Data sharing layer:** In this layer, the researchers, medical workers, government

and healthcare institutions can request patients' EHRs and utilize them for making other health plans, getting better clinic treatment or carrying out medical research.

4.2.3 Blockchain Architecture

The proposed architecture is based on two separate blockchain. And both of the blockchains are private. The sole authority of the blockchain is the patients. Depending on the purpose of the blockchain, the contents inside the blockchains are different. All the blocks of a blockchain contain the basic contents of a blockchain like hash, previous hash, etc. But the transactions in each of the blockchains is different. A brief explanation of the blockchains is as below:

- **Blockchain1:** This blockchain resides in data storing layer along with the private cloud of the patients. This blockchain ensures the integrity of the stored EHR in cloud. The values in the transaction of each block are as below:
  - **Hash:** This field contains the hash[24] of the EHR that is stored in cloud. So if any attacker or intruder changes the EHR in the private cloud of the patient, the patient can chack it with the help of the hash in blockchain1. Even if a single character of the EHR is changed in the cloud, the corresponding hash will change entirely. So, there will be a mismatch between the hash stored in blockchain1 and the hash of the EHR in the cloud. This is how, the integrity of the EHR stored in cloud is ensured.
  - **Timestamp:** This timestamp indicates the time when the raw EHR was store in the private cloud of the patient after receiving it from the doctor.
  - **URL:** This is the location of the EHR in the private cloud[11]. This URL is stored for future purpose. In future, if any third party requests for EHR to a patient. The patient sees the URL of the EHR in blockchain1. After, patient goes to the corresponding URL and retrieves the requested EHR from the private cloud.

- **Blockchain2:** This is another private blockchain of the patient. This blockchain ensures that the patient's EHR is shared with the third parties with some given

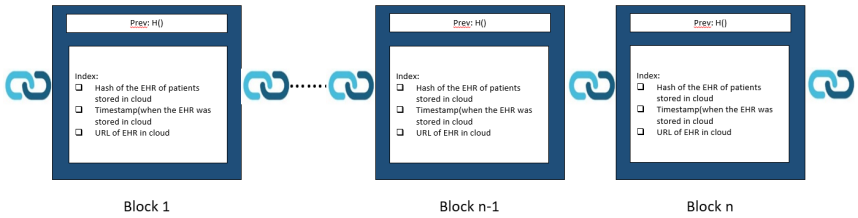


Figure 4.2: Private Blockchain of Patient containing the indexes of the EHR that are stored in the cloud

conditions. And read-only permission is given to the third parties. This blockchain contains the de-identified data, de-identification technique and some terms and conditions created by the patient for securing our data. The transaction of the blockchain2 contains:

– **De-identified EHR:** While extracting stored EHR from the cloud as per the request of the third party, a randomly selected de-identification technique is applied to the extracted EHR [17]. With the help of data de-identification, sensitive information or personal identifying information is hidden. So when the patient is sharing his de-identified EHR with the third party, the identity of the patient is hidden. This is one of the achievements of our architecture. The requested EHR is shared with third parties but identity of the patient is hidden. A example is shown below if the randomly chosen de-identification technique is "Masking".

Figure 4.3: Sensitive information like- name, age, etc are hidden using de-identification technique "Masking"

So when the patient is sharing his EHR, she is making sure that, no one can identify with the information in the EHR. So, the identity of the patient is hidden completely. One of the main advantages of using de-identification technique is that, data can be re-identified from the the de-identified one.

– **De-identification Technique:** A pool of de-identification technique is there from where one technique is selected randomly and applied to the extracted EHR from cloud. This storing of de-identification technique will help in future if we want to track the third party or parties who might have leaked the shared EHR. A short summary of how this is done is described below using a short example. Suppose this is a blockchain2 which has been shared to third parties A, B, C, D, E, F after giving reading permission to them. For the sake of understanding only the transaction in a block is shown.

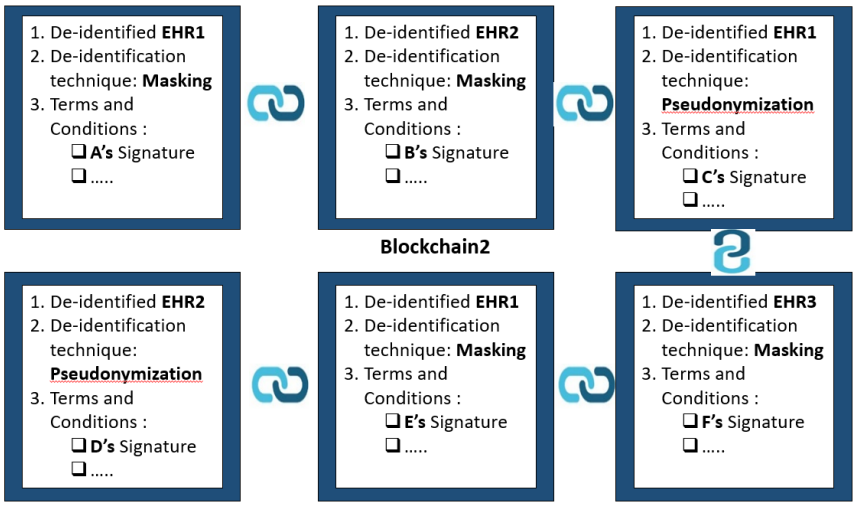


Figure 4.4: Sensitive information like- name, age, etc are hidden using de-identification technique "Masking"

Suppose, after few days, we find out that the patient's EHR1 is with third party X. But according to the blockchain2, the patient did not share data any EHR with third party X. Now the patient find out the de-identification technique that was applied to EHR1 when it was de-identified while sharing. And the patient sees, it was "Masking". Now patient group the blocks with EHR1 and de-identification technique-"Masking". And she comes to a conclusion that third party A and E might have leaked the data to third party X. Because A and E were the ones with whom she shared EHR1 using de-identification technique-"Masking". So either A or E is the culprit to share her confidential data with X. A graphical representation is shown in the next figure.

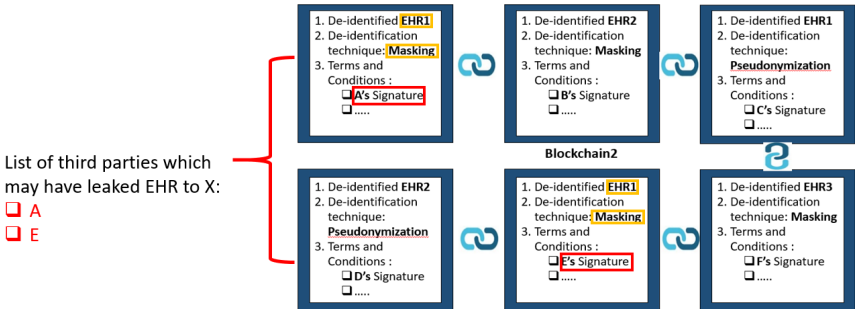


Figure 4.5: Private Blockchain of Patient containing the indexes of the EHR that are stored in the cloud

– **Terms and Conditions:** The terms and conditions is created by the particular patient which includes the third party's signature, the patient's signature, the validation of the contract or data sharing, access permission date. These terms and conditions are very important while we share our data with third party's otherwise a patient can not track to whom she has shared her EHR.

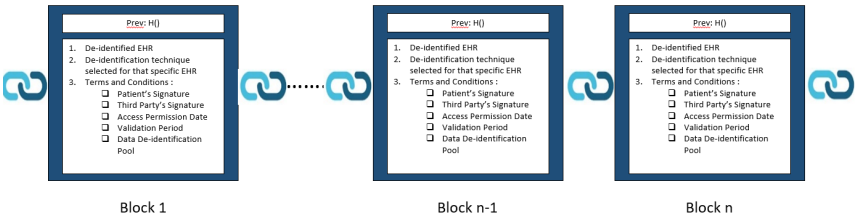


Figure 4.6: Private Blockchain of Patient containing the indexes of the EHR that are stored in the cloud

4.3 Use Case Scenario:

An effective approach of capturing the core functionalities of our system and evaluating the design architecture is to provide different use case diagrams. A use case diagram will show the substantial workflow of the model between different actors and help us to get the initial ideas of the whole system. Here separate use case diagrams are shown to give a better understanding of different activities by providing the direct graphical representation of core functionalities, organization and key principles of system users.

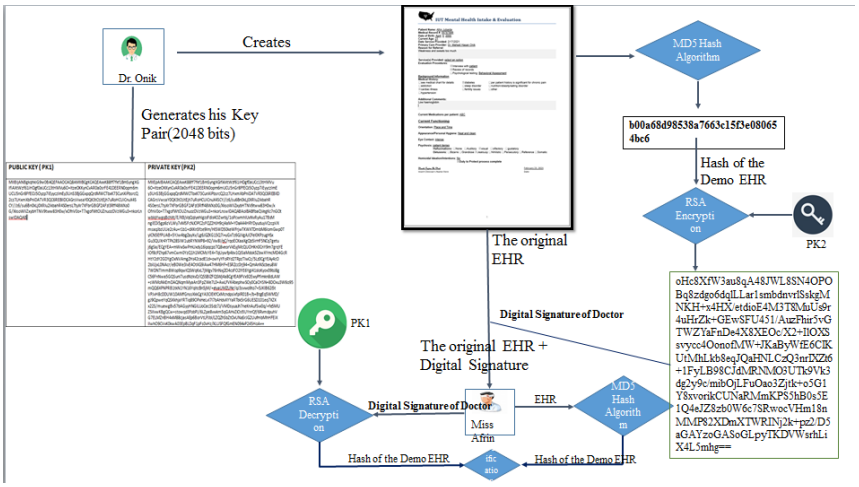


Figure 4.7: Creation and Singing of Raw EHR

The use case shown in figure is designed to represent all events related to doctor's collecting information from patient, to the formulation of a Electronic Health Record for that specific patient, the creation of his public/private key pair, encrypting the created document and finally provide the EHR to that patient and also the verification of the doctor.

- **Stakeholder Roles:** Each stakeholder performs different functions which are discussed as follows:
  - **Doctor/ Clinician:** First of all doctor collects patient's historical and administrative data, after collecting every necessary information he documents notes, prepares a raw EHR and at the same time generates his digital signature to sign the EHR and delivers that to patient. Doctor signs the EHR so that patient can verify the doctor to make sure the authenticity of him. Here to generate

public/private key pair, to sign EHR- different cryptographic algorithms are used.

– **Patient/Consumer:** After getting the signed EHR, first patient verifies the doctor using his public key, then reviews historical and administrative information for accuracy in his raw EHR.

Now we are representing the role of patients after getting his EHR from any hospital/doctor in the following scenerio:

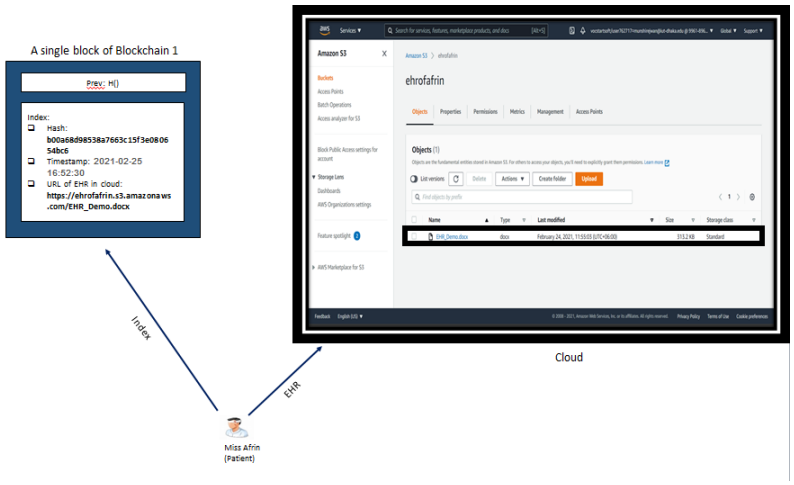
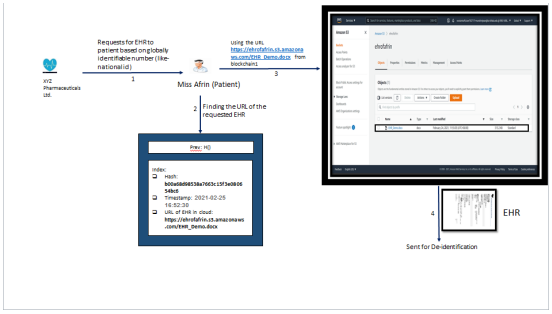


Figure 4.8: Storing Raw EHR in private cloud and indexes in private Blockchain

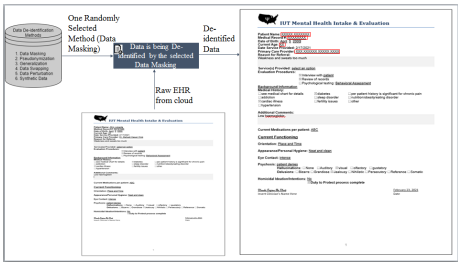
The use case shown in figure is designed to represent all events related to patient's keeping raw EHR to private cloud, to show the use of his private blockchain1.

- **Stakeholder Roles:** Here only stakeholder is the patient who performs the following activities.
  - **Patient/Consumer:** After getting the EHR from a verified clinician or hospital, patient keeps his raw EHR to his private cloud so that nobody have any access to his EHR without his permission, then he retrieves the URL of cloud where EHR is stored as well as the time when he keeps his raw EHR to cloud. He alsp creates the hash of his raw EHR and stores that hash, URL of cloud and timestamp in his private blockchain1 for integrity check in future. Here patient stores the hash of EHR in blockchain1 because even if any malicious third party temper his private blockchain1, he can not get the raw EHR to make any change and if any changes occur in EHR the hash of that changed EHR will not be the same as stored one. So in this way patient can check the integrity and authenticity of his stored EHR in blockchain1.

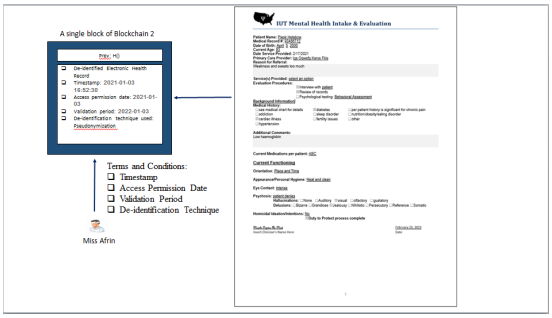
Use case senario of sending third parties access request, retrieving raw EHR and applying de-identification techniques is shown below:



(a) Access request of EHR from third party



(b) Retrieving raw EHR from cloud and apply de-identification technique randomly



(c) Store de-identified EHR and other terms and conditions in blockchain2

Figure 4.9: Grant access of EHR to a valid third party

The above use cases shown in figure is designed to include all events related to send access request of EHR from third party, to find URL from blockchain1, to retrieve requested EHR from cloud by patient, to apply the de-identification techniques on raw EHR, to specify some terms and conditions, to keep all these terms and conditions to private blockchain2 and finally to grant access for third party.

- **Stakeholder Roles:** Here the stakeholder is only the patient who performs the following activities:
  - **Third Parties(Hospital/Company/Researcher/Doctor):** Collects patient's identification that may be NID / medical record, requests for accessing that patient's EHR, provides his digital signature.
  - **Patient/Consumer:** Finds URL of requested EHR from private blockchain1,

retrieves raw EHR from cloud by using that URL, selects any de-identification tachnique randomly and applies on EHR, specifies some terms and conditions for third party, provides his digital signature, keeps all these terms and conditions to his private blockchain2 and grants access of EHR to the specific requested party.

Chapter 5

Discussion

5.1 Thought Analysis

5.1.1 Purpose of Blockchain1:

Information that contain in an EHR are exceedingly individual and delicate. Health Insurance Portability and Accountability Act (HIPAA) [5] says that, a patient must have the complete control over his health data and he has the proper right to set rules and limits on who can get to and retrieve health data agreeing to his choice. On the off chance that we consider the progressing circumstance we may see that when a patient must exchange his clinical information from one healing center to another, he is ordinarily required to sign paper-based assent that indicates what sort of information will be shared and the data around the beneficiary. EHR data sharing is mostly still a repetitive manual process through fax or mail and frequently takes days or indeed months for the records to gotten to be accessible. This is basically due to the need of precise framework support for secure and dependable EHR data sharing, which may also bring about major delays for patient care.

Ecosystems for health information exchange (HIE) aim to ensure that patient information from EHR are safely, proficiently, and precisely shared across the country. Be that as it may, HIEs have constrained selection, and numerous territorial systems are still disconnected. Besides, the current framework needs standard design, coming about in a

disappointment to guarantee legitimate security and get to control for patients once data are shared.

HIEs are generally outlined as a single, completely trusted substance that's exclusively dependable for overseeing and putting away EHR data from numerous participating hospitals. Whereas a centralized framework may be less demanding to manage, it endures from a single point of failure and may demonstrate to be a performance bottleneck for real-world deployment. In expansion, a centralized authority with get to to sensitive health data has demonstrated to have more security and privacy concerns from end users.

Based on the success of the Bitcoin cryptocurrency ,blockchain technologies [4] have recently risen with tremendous momentum. Blockchain employments a distributed ledger to supply a shared, immutable, and transparent history of all the activities that have happened to all the members of the network. It empowers a new era of transactional applications that set up believe, accountability, and straightforwardness. Blockchain, in specific permissioned blockchain innovation, makes it possible for a user to have complete control of information and protection without a central point of control; hence, it is profoundly cost-effective and proficient for building applications for sharing EHR data [6].

Blockchain technology has the potential to transform health care by placing the patient at the center of the health system and increasing the security, privacy, and interoperability of health data [14] [26]. In our system, we try to set two private blockchains which are operated by patient.That means we want to provide a blockchain based architecture for EHR data sharing, for ensuring authentication and integration.

The purpose of using our first blockchain is to ensure the authentication of data that are stored in patient's private cloud. Patient will store indexes in his private blockchain1. And the indexes contain the Hash of the EHR, URL and timestamp. Here patient will store the raw EHR in his private cloud only and Hash of that raw EHR will be stored in the blockchain so that any malicious attacker can not change the content of the EHR even if he tempers the blockchain1 somehow. As patient only stores the Hash in his private

blockchain1, when he wants to share his EHR with any third party he needs the whole raw EHR. To finds that specific EHR in a short time, he stores the cloud location(URL) as well as the time of storing EHR in cloud((timestamp) in his private blockchain1. So if any malicious attacker changes the content of EHR then the hash will be changed as well and that hash will not be the same as stored in patient's private blockchain1. In this way patient can easily authenticate the EHR data.

5.1.2 Purpose of Blockchain2:

The state of health care records is as of now detached since still there's no common models and benchmarks that would permit the secure exchange of touchy data among partners. Patients get to consents to the current EHRs are exceptionally restricted and patients are regularly incapable to share their information with analysts or suppliers effectively. In spite of all the progresses in pharmaceutical, different EHR frameworks don't communicate successfully. Each health care institution gives administrations, tracks, and upgrades the patient's clinical data set each time a restorative benefit is given. This data incorporates individual information, such as the patient's gender and date of birth, as well as data on the particular benefit given, such as the strategy performed, the care plan. Those information's are usually stored in a database within the organization or within a defined network of health care stakeholders. This stream of data beginning from the quiet through the health care organization each time a benefit is performed ought to not halt at the person organizational level or at the health care database as it were. Instead, that information representing each patient interaction should be directed into a nationwide blockchain transaction layer [9].

Thus, data stored on the blockchain [13] may be all around accessible to a particular person through the blockchain private key. The private key empowers patients to share their data with diverse health care organizations more seamlessly. Health care information's are sensible information that must be kept in secret. Thus, each health organization's EHR system must execute privacy policies in order to ensure that only the patient and the healthcare Blockchain Technology applied to Electronic Health Records can have access to personal health records [2].

The purpose of using the second private blockchain of patient in our system is to ensure the authenticity of patient himself and also third parties with whom patient shares his EHR data. Patient gives the read only permission to third parties. This blockchain contains the de-identified EHR, the de-identification technique selected for each EHR to de-identify, patient's digital signature, third party's digital signature and some terms and conditions defined by patient to access the shared EHR. The terms and conditions that set up by patient can be the access permission date that means from which date patient grants access to his EHR to the third parties and the validation period which means upto the designated time third parties will have permission to access the EHR. We use this second blockchain so that any malicious attacker can not easily get the deidentified EHR without tempering the blockchain and also to keep track the third parties by storing their signatures as well.

5.1.3 Why we use De-identification Technique:

De-identification and anonymization are techniques that are utilized to remove patient identifiers in electronic health record information. The use of these techniques in multi-center investigate studies is fundamental in significance, given the need to share electronic health record information over different environments and institutions whereas shielding patient protection.

Electronic health records (EHR) contain a huge amount of organized information and free content. Investigating and sharing clinical information can progress healthcare and encourage the advancement of medical program. In any case, uncovering confidential data is against moral standards and laws.

One approach to encourage the revelation of data for the purposes of genomic research or any research, and to ease a few of the issues, is to de-identify information before disclosure to analysts or at the most punctual opportunity a short time later . Numerous investigate ethics boards will waive the consent necessity in case the primary 'use' of the data is to de-identify it. De-identified patient data can be utilized to improve care, assess the costs

of care, and bolster open health activities. Researchers have been doing just this for a long time.

Within the medical informatics region, a few de-identification algorithms have been developed. There are different de-identification techniques that include pseudonymization, k-anonymization, masking and so on.Few research [3] [15] [8] [21] [22] [23] on de-identifying health records have been done in recent years. These research mainly focus on de-identifying datasets extricated from EHR database table, and none has displayed a de-identification algorithm for a full EHR database, guaranteeing satisfactory levels of secrecy, medical rightness and readability. Moreover, till now no research has been found where de-identified EHR has been stored on blockchain. We store patients' de-identified EHR in blockchain to give an extra level of privacy and integrity that means we want to make sure that no malicious attacker can change our de-identified EHR or access those EHR easily. Besides this, in our system patient choose any of the de-identification technique randomly so that even if his de-identified EHR will be available to an unknown user, he can't make any guess in an immediate manner. Along with the de-identified EHR, which technique has been used for each specific EHR to de-identify it will also be stored in patients' second private blockchain so that he can keep a track of third party access easily as well as the EHR can be simply re-identified when necessary.

5.1.4 The reason why most of the tasks are given to the patients:

- In the present world the patients do not feel safe to keep their medical information with the hospitals.
- The doctors/hospitals may misuse the EHR of the patients.
- In this present world where everything is digitalized, it is not that difficult for the patients to handle their own EHR.
- As both of the blockchains are private, the patients handle their own EHR, cloud and blockchains.

5.2 Implementation Scope/Challenges:

It is possible to bring all the tasks in the proposed architecture into one single system. Two separate blockchains can be implemented together. The de-identification pool can be created by which one de-identification technique will come randomly and that technique will be applied to the EHR that is to be shared. And all these tasks are to be integrated on single software or system.

5.3 Limitations and Future Works:

In the proposed architecture the patients are handling most of the tasks. Also the whole architecture is not implemented yet in a single platform. Two blockchains have been used in the whole architecture which is time consuming and cosltly to maintain. Integrating and implementing the whole architecture in one single platform is left for future.

Chapter 6

Conclusion

Storing EHR and sharing it with third parties for research and any other purpose is a challenging work. Because we need to ensure the integrity of our stored data. At the same time we need to keep our shared EHR confidential. So a blockchain based architecture has been proposed by us which can solve this issues. Data de-identification technique has been used to hide the sensitive information about a patient.

References

[1] Nahid AlThqafi, Hessah AlSalamah, and Ahmad Daraiseh. "Achieving Patient-Centered Fine-Grained Access Control in Hospital Information Systems - Using Business Process Management Systems". In: *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5: HEALTHINF, (BIOSTEC 2016)*. INSTICC. SciTePress, 2016, pp. 39–48. ISBN: 978-989-758-170-0. DOI: 10.5220/0005630200390048.

[2] S Attili et al. "Blockchain: the chain of trust and its potential to transform healthcare–our point of view". In: *ONC/NIST Use of Blockchain for Healthcare and Research Workshop. Gaithersburg, MD: ONC/NIST*. 2016.

[3] Jules J Berman. "Concept-match medical data scrubbing: how pathology text can be used in research". In: *Archives of pathology & laboratory medicine* 127.6 (2003), pp. 680–686.

[4] Whitfield Diffie and Martin Hellman. "New directions in cryptography". In: *IEEE transactions on Information Theory* 22.6 (1976), pp. 644–654.

[5] Alevtina Dubovitskaya et al. "ACTION-EHR: Patient-Centric Blockchain-Based Electronic Health Record Data Management for Cancer Care". In: *Journal of medical Internet research* 22.8 (2020), e13598.

[6] Ariel Ekblaw et al. "A Case Study for Blockchain in Healthcare:"MedRec" prototype for electronic health records and medical research data". In: *Proceedings of IEEE open & big data conference*. Vol. 13. 2016, p. 13.

[7] Andressa Fernandes et al. “Scalable Architecture for sharing EHR using the Hyperledger Blockchain”. In: Mar. 2020, pp. 130–138. DOI: 10.1109/ICSA-C50368.2020.00032.

[8] Dilip Gupta, Melissa Saul, and John Gilbertson. “Evaluation of a deidentification (De-Id) software engine to share pathology reports and clinical documents for research”. In: *American journal of clinical pathology* 121.2 (2004), pp. 176–186.

[9] Marko Hölbl et al. “A systematic review of the use of blockchain in healthcare”. In: *Symmetry* 10.10 (2018), p. 470.

[10] Chu Ya Huang et al. “A personalized medication management platform (PMMP) to improve medication adherence: A randomized control trial”. English. In: *Computer Methods and Programs in Biomedicine* 140 (Mar. 2017), pp. 275–281. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2016.12.012.

[11] Yashpalsinh Jadeja and Kirit Modi. “Cloud computing-concepts, architecture and challenges”. In: *2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET)*. IEEE. 2012, pp. 877–880.

[12] Neesha Jothi, Nur’Aini Abdul Rashid, and Wahidah Husain. “Data Mining in Healthcare – A Review”. In: *Procedia Computer Science* 72 (Dec. 2015), pp. 306–313. DOI: 10.1016/j.procs.2015.12.145.

[13] Tsung-Ting Kuo, Hyeon-Eui Kim, and Lucila Ohno-Machado. “Blockchain distributed ledger technologies for biomedical and health care applications”. In: *Journal of the American Medical Informatics Association* 24.6 (2017), pp. 1211–1220.

[14] Xueping Liang et al. “Integrating blockchain for data sharing and collaboration in mobile healthcare applications”. In: *2017 IEEE 28th annual international symposium on personal, indoor, and mobile radio communications (PIMRC)*. IEEE. 2017, pp. 1–5.

[15] Stephane M Meystre et al. “Automatic de-identification of textual documents in the electronic health record: a review of recent research”. In: *BMC medical research methodology* 10.1 (2010), pp. 1–16.

[16] Satoshi Nakamoto. "Bitcoin: A Peer-to-Peer Electronic Cash System". In: *Cryptography Mailing list at https://metzdowd.com* (Mar. 2009).

[17] Ishna Neamatullah et al. "Automated de-identification of free-text medical records". In: *BMC medical informatics and decision making* 8.1 (2008), pp. 1–17.

[18] "Opportunities and Challenges of Cloud Computing to Improve Health Care Services". In: (). DOI: 10.2196/jmir.1867.

[19] Mohammad Saidur Rahman et al. "A Novel Architecture for Tamper Proof Electronic Health Record Management System Using Blockchain Wrapper". In: *Proceedings of the 2019 ACM International Symposium on Blockchain and Secure Critical Infrastructure*. BSCI '19. Auckland, New Zealand: Association for Computing Machinery, 2019, pp. 97–105. ISBN: 9781450367868. DOI: 10.1145/3327960.3332392. URL: https://doi.org/10.1145/3327960.3332392.

[20] Y. Sun et al. "A Decentralizing Attribute-Based Signature for Healthcare Blockchain". In: *2018 27th International Conference on Computer Communication and Networks (ICCCN)*. 2018, pp. 1–9. DOI: 10.1109/ICCCN.2018.8487349.

[21] György Szarvas, Richárd Farkas, and Róbert Busa-Fekete. "State-of-the-art anonymization of medical records using an iterative machine learning framework". In: *Journal of the American Medical Informatics Association* 14.5 (2007), pp. 574–580.

[22] Özlem Uzuner, Yuan Luo, and Peter Szolovits. "Evaluating the state-of-the-art in automatic de-identification". In: *Journal of the American Medical Informatics Association* 14.5 (2007), pp. 550–563.

[23] Sumithra Velupillai et al. "Developing a standard for de-identifying electronic patient records written in Swedish: precision, recall and F-measure in a manual and computerized annotation trial". In: *International journal of medical informatics* 78.12 (2009), e19–e26.

[24] Hongjun Wu. "The hash function JH". In: *Submission to NIST (round 3)* 6 (2011).

[25] Qi Xia et al. "BBDS: Blockchain-Based Data Sharing for Electronic Medical Records in Cloud Environments". In: *Inf.* 8 (2017), p. 44.

[26] Peng Zhang et al. “FHIRChain: applying blockchain to securely and scalably share clinical data”. In: *Computational and structural biotechnology journal* 16 (2018), pp. 267–278.