BLOCKCHAIN-BASED ELECTRONIC MEDICAL RECORDS (EMR) MANAGEMENT WITH OCR ASSISTED SUMMARIZATION

A PROJECT REPORT

Submitted by

VATSAL GUPTA [17104060]

Under the guidance of

Mr. Vikas Hassija

(Assistant Professor, Department of Computer Science & IT)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

Sector 62, Noida, Uttar Pradesh 201309

DECEMBER 2020

ABSTRACT

Blockchain have been an interesting research area for a long time and the benefits it provides have been used by a number of various industries. Similarly, the healthcare sector stands to benefit immensely from the blockchain technology due to security, privacy, confidentiality and decentralization. Nevertheless, the Electronic Health Record (EHR) systems face problems regarding data security, integrity and management. This project discusses how the blockchain technology can be used to transform the EHR systems and could be a solution of these issues. We present a framework that could be used for the implementation of blockchain technology in healthcare sector for EHR. The aim of the proposed framework is firstly to implement blockchain technology for EHR and secondly to provide secure storage of electronic records by defining granular access rules for the users of the proposed framework. Moreover, this framework also discusses the scalability problem faced by the blockchain technology in general via use of off-chain storage of the records. This framework provides the EHR system with the benefits of having a scalable, secure and integral blockchain-based solution.

PREFACE

Blockchain is one of the emerging technologies and rightfully so. It ports a centralized system into a decentralized system which is secure and cannot be messed with. In this project, I plan to use this beautiful piece of technology to implement an innovative solution in the field of healthcare. Recently, due to the COVID-19 pandemic, we have seen that the healthcare sector struggles in many aspects. Through this project, I plan to address a few of them. The main objective of this project is to build a decentralized application, that keeps track of patients' records, to make the healthcare system more efficient and simple. The platform itself is very simple. All one has to do is, login, add their name and documents, and specify his/her trusted doctors and pharmacists. The platform is quite simple and intuitive to use.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my guide, Mr. Vikas Hassija, for his valuable guidance, consistent encouragement, timely help, and providing me with an excellent atmosphere for doing my project. During the entire duration of my dissertation work, despite his busy schedule, he has extended cheerful and cordial support to me for completing this project work. I would also like to convey my sincere regards to all other faculty members of the department of CSE & IT, JIIT, who have bestowed their great effort and guidance at appropriate times. My thanks and appreciations also go to people who have willingly helped in developing this project.

Contents

List of Tables

List of Figures

ABBREVIATIONS

EHR	Electronic Health Record
EHR	Electronic Medical Record
IPFS	InterPlanetary File System
P2P	Peer-to-Peer

Chapter 1

INTRODUCTION

1.1 Problem Statement

Over decades, medical facilities have evolved elegantly. Still most of us are the witness of the fact that whenever we see a doctor, we need to put forward our medical file in front of him/her. Our file contains our previous prescriptions, medical reports, X-Rays, MRIs etc. It is a tedious task to keep record of all these.

Furthermore, health data contained in legacy systems is isolated and difficult to share with others because of varying formats and standards. In short, the current healthcare data landscape is fragmented and ill-suited to the instantaneous needs of modern users. At present, Electronic Medical Record (EHR) are stored on centralized databases in which medical data remains largely non-portable. Centralization increases the security risk footprint and requires trust in a single authority. Moreover, centralized databases cannot ensure security and data integrity, regardless of de-identification and controlled access requirements. Centralized health databases are legally a requirement and necessity in most countries worldwide and therefore require an added layer of technology to improve their portability and security.

1.2 Research Motivation

Before the advent of modern technology, healthcare sector used paper based system to store the medical records, i.e., using handwritten mechanism. This paper-based medical record system was inefficient, insecure, unorganized and was not temper-proof. It also faced the issue of data- duplication and redundancy as all the institutions that patient visited had various copies of patient's medical records.

The healthcare sector faced a trend shift towards EHR systems that were designed to combine paper-based and electronic medical records (EMR). These systems were used to store

clinical notes and laboratory results in its multiple components [1]. They were proposed to enhance the safety aspect of the patients by preventing errors and increasing information access [2]. The goal of EHR systems was to solve the problems faced by the paper-based healthcare records and to provide an efficient system that would transform the state of healthcare sector [3].

The EHR systems have been implemented in a number of hospitals around the world due the benefits it provides, mainly the improvement in security and its cost-effectiveness. They are considered a vital part of healthcare sector as it provides much functionality to the healthcare [4]. These functionalities are electronic storage of medical records, patients' appointment management, billing and accounts, and lab tests. They are available in many of the EHR system being used in the healthcare sector. The basic focus is to provide secure, temper-proof, and shareable medical records across different platforms. Despite the fact that notion behind usage of EHR systems in the hospitals or healthcare was to improve the quality of healthcare, these systems faced certain problems and didn't meet the expectations associated with them [3]. A study was conducted in Finland to find the experiences of nursing staff with the EHR, it was concluded that EHR systems faced the problems related to them being unreliable and having a poor state of user-friendliness [5].

1.3 Solution Identified: Blockchain

Blockchain technology was introduced by Nakamoto [6], for his popular work of digital currency or crypto-currency, i.e., bitcoin. Nakamoto used blockchain technology to solve the double spending problem of bitcoin but soon this novel technology was being used in many other applications. Blockchain is a chain of blocks that are connected together and are continuously growing by storing transactions on the blocks. This platform uses a decentralized approach that allows the information to be distributed and that each piece of distributed information or commonly known as data have shared ownership. Blockchains holds batches of transactions that are hashed thus providing them security and they are managed by peer-to-peer networks. A blockchain has certain benefits such as security, anonymity, and integrity of data with no third party intervention. These benefits make it a reasonable choice to store patient's medical records on it, because the innovation of technology in the healthcare industry has made the security of

patient's medical data a top priority. A number of researchers have also identified that using blockchain technology in healthcare would be a feasible solution [7, 8, 9].

The advantages of having a secure, immutable and decentralized Electronic Health Record (EHR) database:

- Single version of the truth verified by the consensus of the participating hospitals
- Easy to share selective or all EHRs as consented by the patient
- Full medical history of a patient at one single point
- Easy verification of medical prescription
- Redacted EHRs for research purposes
- Increased transparency
- No insurance fraud

Chapter 2

LEARNING

2.1 Blockchain

A blockchain is a continuously growing list of records, called blocks, which are linked and secured using cryptography. Each block typically contains a crypto- graphic hash of the previous block, a timestamp and transaction data. By design, a blockchain is inherently resistant to modification of the data. It is “an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way”. For use as a distributed ledger, a blockchain is typically managed by a peer-to-peer network collectively adhering to a protocol for inter-node communication and validating new blocks. Once recorded, the data in any given block cannot be altered retroactively without the alteration of all subsequent blocks, which requires collusion of the network majority.

2.2 Importance of Electronic Health Records

Blockchain technology has a large potential to transform business operating models in the long term. Blockchain distributed ledger technology is more a foundational technology—with the potential to create new foundations for global economic and social systems—than a disruptive technology, which typically “attack a traditional business model with a lower-cost solution and overtake incumbent firms quickly”. The use of blockchain promises to bring significant efficiencies to global supply chains, financial transactions, asset ledgers and decentralized social networking.

2.3 Difference between Public and Private Blockchain

The sole distinction between public and private blockchain is related to who is allowed to participate in the network, execute the consensus protocol and maintain the shared ledger. A public

blockchain network is completely open and anyone can join and participate in the network. The network typically has an incentivizing mechanism to encourage more participants to join the network. Bitcoin is one of the largest public blockchain networks in production today. One of the drawbacks of a public blockchain is the substantial amount of computational power that is necessary to maintain a distributed ledger at a large scale. More specifically, to achieve consensus, each node in a network must solve a complex, resource-intensive cryptographic problem called a proof of work to ensure all are in sync.

2.4 Importance of EHRs

EHRs and the ability to exchange health in- formation electronically can help you provide higher quality and safer care for patients while creating tangible enhancements for your organization. EHRs help providing accurate, up-to-date, and complete information about patients at the point of care. It helps securely sharing electronic information with patients and other clinicians. It promotes legible, complete documentation and accurate, streamlined coding and billing. It enhances privacy and security of patient data. It reduces costs through decreased paperwork, improved safety, reduced duplication of testing, and improved health.

Chapter 3

RESEARCH METHODOLOGY & DESIGN

3.1 Preliminaries

This section formally describes the preliminaries used in proposed framework. It describes the software platform used for development of this framework and its advantages. Ethereum and IPFS being the most prominent and important for implementation of this framework are also discussed in the following section.

3.1.1 Ethereum

Ethereum is a distributed blockchain network that uses the idea of blockchain that was previously used in the popular crypto currency Bitcoin [6]. Ethereum was formally introduced in year 2015 and the idea behind Ethereum was to create a trustless smart contract platform that would be open-source and would also hold the feature of programmable blockchain. This technology also shares the peer-to-peer networking that makes it distributed. This platform also makes use of its own crypto currency known as Ethers [10]. This crypto currency can be used for sharing it between accounts connected on Ethereum blockchain [11]. Ethereum also provides the programmers a language in which they can customize their own blockchain, this language is known as Solidity. It was developed for smart contracts that are the main feature of Ethereum.

3.1.2 Information Transaction

In Ethereum, transaction is the way external entity would interact with Ethereum. It can be used by external user to update the state of the record or information stored on the Ethereum blockchain network. An Ethereum transaction contains following elements [12]:

- **From** – message sender, having a 20-bytes address.

- **To** – message recipient, also having a 20-bytes address.
- **Value** - the fund amount (wei) transferred from sender to recipient
- **Data (optional)** – contains the message that is being sent to the recipient
- **Gas** – For every transaction on the Ethereum blockchain the sender needs to pay some fees for performing that operation this fee is known as Gas. Every transaction contains the gas limit and gas price in it.
- **Gas Price** - that fee the transaction sender is willing to pay for gas
- **Gas Limit** - maximum gas that could be paid for this transaction

3.1.3 Smart Contracts

Smart contract are known as the piece of code that is used to perform any task on the blockchain. This piece of code is executed when the users send the transactions. They run on the blockchain directly thus making themselves secure from any kind of tampering and alterations. Smart contract commonly use solidity language and they can be used to program any kind of operation that a programmer wants to do on the blockchain. After programming the required operations the programmers can compile them by using EVM bytecode that would be explained in next section. And after compiling them it could be executed and deployed on the Ethereum blockchain [13]. The programming language of JavaScript and Python are encapsulated with the Solidity language provided by Ethereum to write code in smart contracts.

3.1.4 Interplanatery File System (IPFS)

IPFS is a protocol that uses peer-to-peer network for data storage. It provides secure data storage as data stored on IPFS is protected from any alteration. It uses a cryptographic identifier that protects the data from alteration as any attempt to make change on the data stored on IPFS could only be done by changing the identifier. All the data files stored on IPFS contains a hash value that is generated cryptographically. It is unique and is used for identification of stored data file on the IPFS [14]. This secure storage strategy of IPFS protocol makes it a favorable choice for storing critical and sensitive data. The cryptographic hash that is generated could be stored on the decentralized application to reduce the exhaustive computational operations over the blockchain. InterPlanetary File System (IPFS) protocol works using a Peer-to-Peer (P2P)

network, this network contains a data structure known as IPFS object that contains data and link in it. Data is unstructured binary data and link consists of an array. The IPFS protocol works in the following way [15]:

- Files stored on IPFS are assigned a unique cryptographic hash
- Duplicate files are not allowed to exist on the IPFS network
- A node on the network stores content and index information of the node

3.1.5 Ganache: Truffle Suite

Ganache is a personal blockchain for rapid Ethereum and Corda distributed application development. Ganache can be used across the entire development cycle; enabling blockchain developers to develop, deploy, and test their dApps in a safe and deterministic environment. Ganache UI is desktop application supporting both Ethereum and Corda technology. In addition, an Ethereum version of ganache is available as a command-line tool: ganache-cli (formerly known as the TestRPC). All versions of Ganache are available for Windows, Mac, and Linux [16].

3.2 Workflow

The workflow would be as follows:

1. Anyone can register on the Blockchain network as a doctor or a patient.
2. The patient can upload his/her medical records and give his/her trusted doctors permission to view them.
3. After logging in, the doctor can view these documents.
4. The doctor and the patient can view the summary of the reports by running an OCR (PyTesseract) module.

3.3 Architecture

System design is the most important and vital part of any framework as it is used for the development of the system from its theory. This section includes the modules, architecture

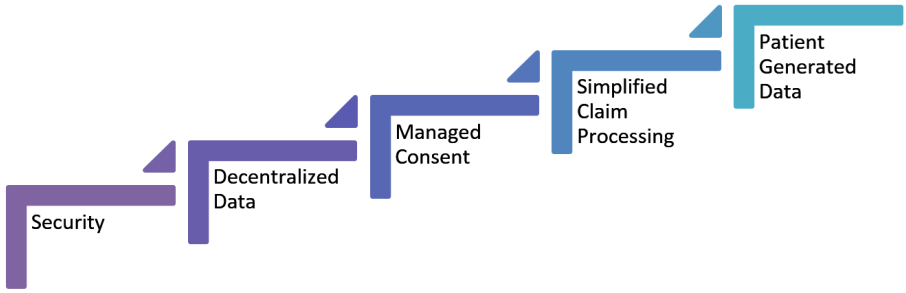

Figure 3.1: Workflow

and various elements that are combined together to form the whole system's framework. As defined earlier the purpose behind this proposed framework is to create such a decentralized system that is temper-proof, secure and confidential blockchain-based system for electronic health records. The system architecture of the proposed framework is visible below in figure 3.2.

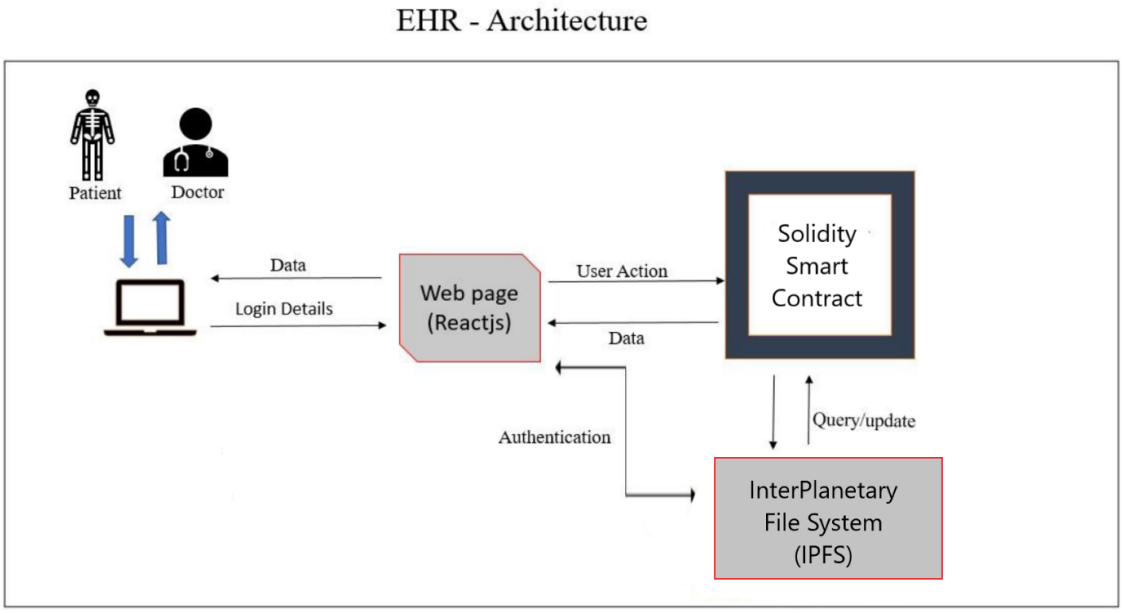

Figure 3.2: Architecture

3.4 Tech Stack Used

- ReactJS, CSS and Javascript for Front-End
- Python, Flask for OCR and hosting API

- Solidity for Smart Contracts
- Metamask for simulation
- Ganache (Truffle Suite) for Test Nodes and simulation
- IPFS for Database (storing electronic records)
- Infura API and Web3 for provider

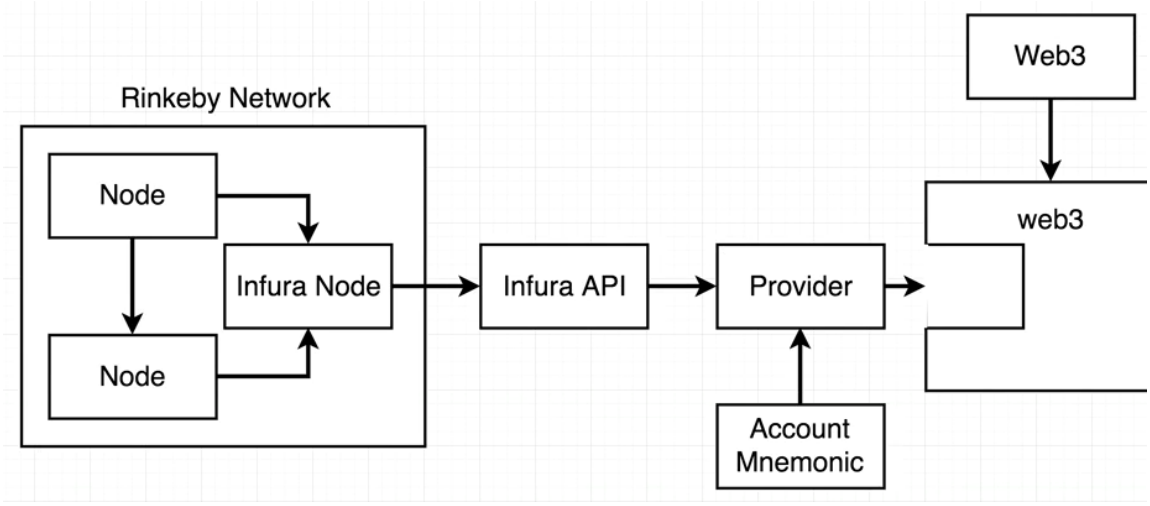

Figure 3.3: Contract Deployment Flow

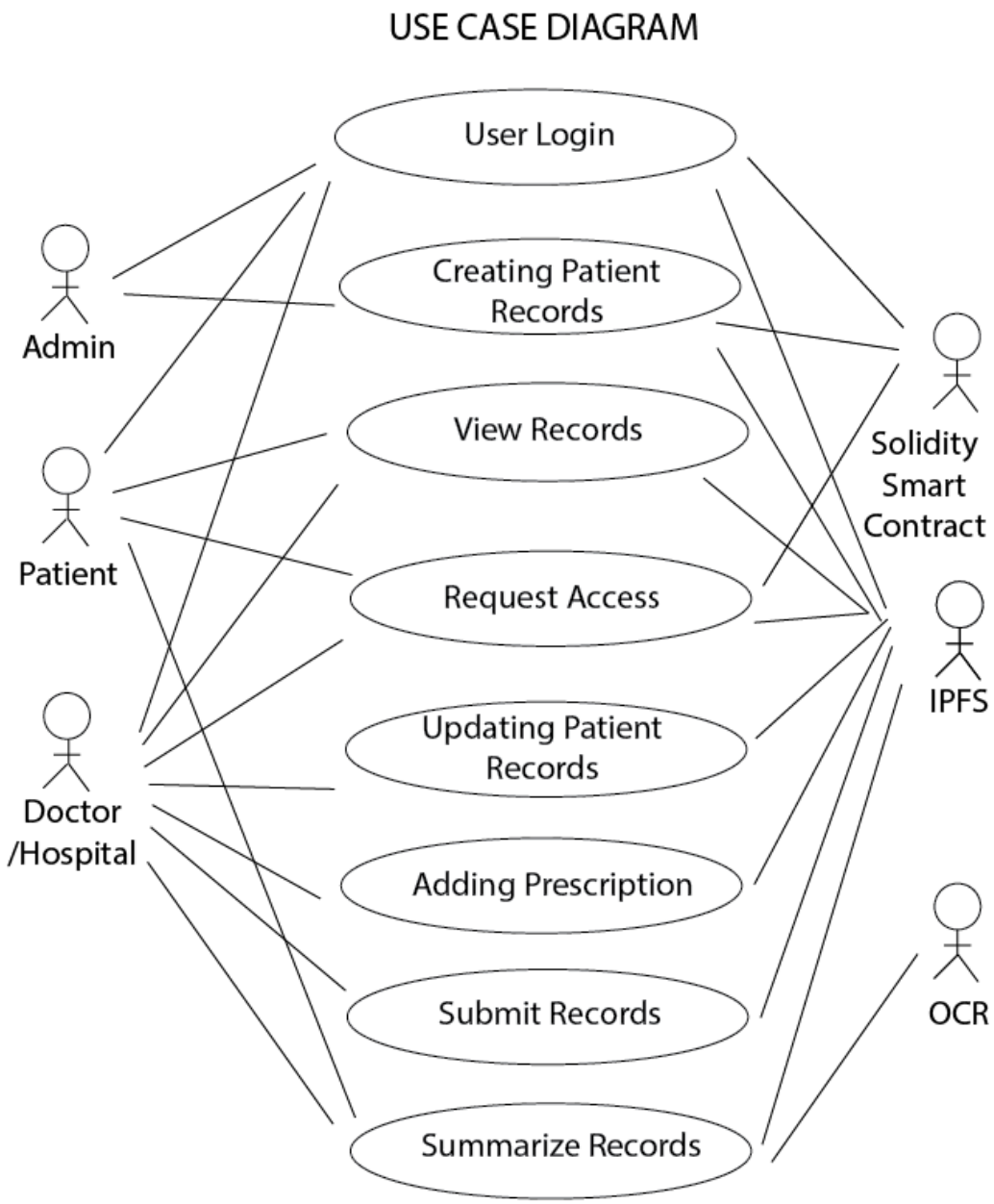

Figure 3.4: Use Case Diagram

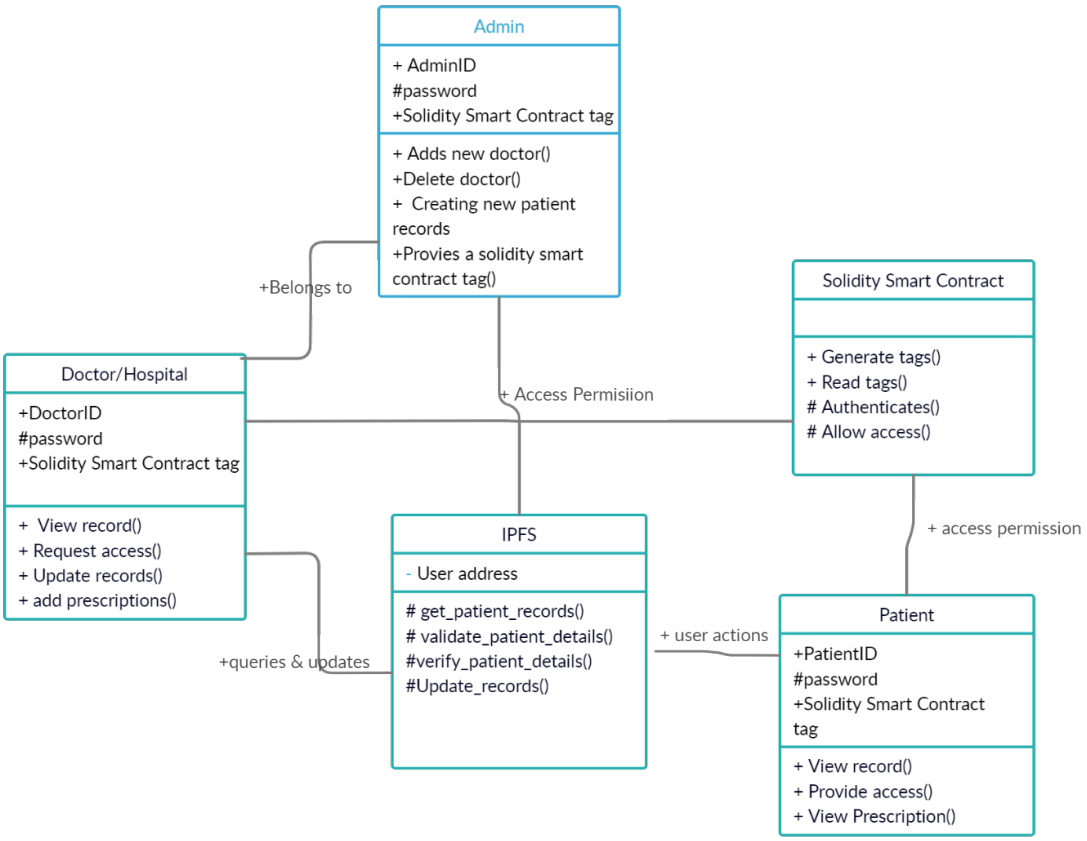

Figure 3.5: ER Diagram

Chapter 4

CONCLUSION AND FUTURE DIRECTION

4.1 Conclusion

In this paper we discussed how blockchain technology can be useful for healthcare sector and how can it be used for electronic health records. Despite the advancement in healthcare sector and technological innovation in EHR systems they still faced some issues that were addressed by this novel technology, i.e., blockchain. Our proposed framework is a combination of secure record storage along with the granular access rules for those records. It creates such a system that is easier for the users to use and understand. Also, the framework proposes measures to ensure the system tackles the problem of data storage as it utilizes the off-chain storage mechanism of IPFS. And the role-based access also benefits the system as the medical records are only available to the trusted and related individuals. This also solves the problem of information asymmetry of EHR system.

For the future, we plan to implement the payment module in the existing framework. For this we need to have certain considerations as we need to decide how much a patient would pay for consultation by the doctor on this decentralized system functioning on the blockchain. We would also need to define certain policies and rules that comply with the principles of the healthcare sector.

Bibliography

[1] R. Akkaoui, X. Hei, and W. Cheng, "Edgemedichain: A hybrid edge blockchain-based framework for health data exchange," *IEEE Access*, vol. 8, pp. 113467–113486, 2020.

[2] T. GÓrski and J. Bednarski, "Transformation of the uml deployment model into a distributed ledger network configuration," in *2020 IEEE 15th International Conference of System of Systems Engineering (SoSE)*, pp. 255–260, IEEE, 2020.

[3] T. Górski and J. Bednarski, "Applying model-driven engineering to distributed ledger deployment," *IEEE Access*, vol. 8, pp. 118245–118261, 2020.

[4] B. Shala, U. Trick, A. Lehmann, B. Ghita, and S. Shiaeles, "Blockchain and trust for secure, end-user-based and decentralized iot service provision," *IEEE Access*, vol. 8, pp. 119961–119979, 2020.

[5] A. Hafid, A. S. Hafid, and M. Samih, "Scaling blockchains: A comprehensive survey," *IEEE Access*, vol. 8, pp. 125244–125262, 2020.

[6] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," tech. rep., Manubot, 2019.

[7] W. J. Gordon and C. Catalini, "Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability," *Computational and structural biotechnology journal*, vol. 16, pp. 224–230, 2018.

[8] A. Boonstra, A. Versluis, and J. F. Vos, "Implementing electronic health records in hospitals: a systematic literature review," *BMC health services research*, vol. 14, no. 1, p. 370, 2014.

[9] T. D. Gunter and N. P. Terry, "The emergence of national electronic health record architectures in the united states and australia: models, costs, and questions," *Journal of medical Internet research*, vol. 7, no. 1, p. e3, 2005.

[10] S. Gupta and M. Sadoghi, "Blockchain transaction processing.," 2019.

[11] U. W. Chohan, "Cryptocurrencies: A brief thematic review," *Available at SSRN 3024330*, 2017.

[12] G. Wood *et al.*, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, no. 2014, pp. 1–32, 2014.

[13] D. Vujičić, D. Jagodić, and S. Ranđić, "Blockchain technology, bitcoin, and ethereum: A brief overview," in *2018 17th international symposium infoteh-jahorina (infoteh)*, pp. 1–6, IEEE, 2018.

[14] T. Dey, S. Jaiswal, S. Sunderkrishnan, and N. Katre, "Healthsense: A medical use case of internet of things and blockchain," in *2017 International conference on intelligent sustainable systems (ICISS)*, pp. 486–491, IEEE, 2017.

[15] “Ipfs documentation,” Aug 2020.

[16] T. Suite, “Ganache: Overview: Documentation.”