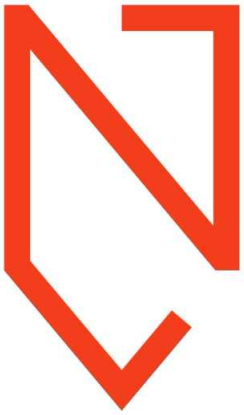


Noroff University College

An evaluation of Blockchain Technology to Provide Data Integrity for Electronic Medical Records

Submitted in partial fulfilment
of the requirements of the degree of

Bachelor in Cyber Security

of Noroff University College

Erlend Halsnes

Kristiansand, Norway
May 2022

Declaration

I declare that the work presented for assessment in this submission is my own, that it has not previously been presented for another assessment, and that work completed by others has been appropriately acknowledged.

Name: Erlend Halsnes **Date**: May 24, 2022

Abstract

One of the biggest problems of healthcare today is that organisations hold multiple fragmented health records about patients, but lack secure enough platforms to safely store and process these records. As medical records are a critical asset used to diagnose and treat patients, the information which they hold, such as information about patients physical and mental health history, are seen as highly sensitive and confidential. This project aims to look for the possibilities of emerging disruptive technologies to improve data security within the healthcare industry, and specifically how blockchain technology could be used to secure data integrity of electronic medical records (EMRs). Future EMR management systems may benefit from taking advantage of the same security mechanisms of blockchain technology which for the past decade has secured the data integrity of various distributed financial ledgers across the world. Blockchain technology may have vast potential in verifying the critical data integrity of EMRs and provide patients the ability to self-audit their medical records if appropriate solutions are developed merging the two technologies. Blockchain-based alternatives to EMR storage and access control may provide significant security improvements within this field, and may be the key technology for effectively securing and storing EMRs securely, globally. This project demonstrates a working solution which used blockchain-technology in order to verify the integrity of an EMR.

Keywords: *Blockchain, Cloud Security, Decentralization, Distributed Ledger Technology, Electronic Health Records, Data Integrity*

Acknowledgements

For guidance and supervision for this bachelor project, I would like to thank my supervisor Prof. Fabricio Bortoluzzi at Noroff University College, Department of Kristiansand, Norway.

A thank you also goes to my closest friends for support in reading and feedback throughout the period.

Contents

List of Figures

List of Tables

Listings

1

Introduction

1.1 Problem Statement

The emergence of cloud computing has proven to be largely disruptive to how users interact with data online. Rather than being kept locally on individual users' hard drives, user data is increasingly being kept in massive online data centers. enabling users to retrieve their data stored in the cloud on demand, from any location, using any device they choose. There are a lot of benefits that come with using this method of data storage, such as improved scalability and availability of the data (Filippi 2013).

However, the shift toward more centralized storage solutions, in which thousands of users store their data in a small number of very large data centers, raises significant ethical, security, and privacy related issues.

The storage of data becomes more centralized when more of it is kept in the same physical location. This makes it possible to have a single point of failure, or at the very least, fewer points of failure overall. The users' personal and business data may not be under their control, which calls into question the data's privacy. As a result, users must have faith that cloud storage providers will appropriately manage and secure user privacy. These problems, when taken together, reduce user autonomy, which means that users are forced to rely on the infrastructure and services provided by third parties.

According to researchers at Swinburne University of Technology, some educational institutions have already begun designing several different blockchain-based solutions for decentralized storage and management of data. These blockchain-based solutions can be created through the application of

blockchain technology to create decentralized storage solutions in which the users themselves are in control of their data, even though the data itself is stored in a decentralized cloud network (Chowdhury et al. 2018).

1.2 Research Objectives

This research aims to explore how blockchain technology can be used for the purpose of storing EMRs, and to which degree the technology can improve on traditional storage and sharing methods in terms of security. To be more specific, the purpose of this research is to investigate the ways in which blockchain-based cloud EMR management systems could potentially provide increased security and integrity. The literature review will discuss existing secondary literature concerning the concepts and potential implementations of such a system, the status of the research, and attempt to explain how these systems would function and be created.

It is difficult to share electronic medical records (EMRs) because the data can be compromised in a number of different ways, including through theft, data loss, unauthorized access, data leaks, and data tampering.

In order to implement an application that demonstrates how blockchains can be used to secure the data integrity of EMRs, a literature review has been conducted. The purpose of this literature review is to provide an overview of the related research that has been carried out on relevant topics. This literature review was also carried out in order to facilitate the implementation of an application. It will delve into the motivation behind why a better solution is needed, how it can be created, what its potential may be, and explain how Blockchain's inherent natural features are a good fit for this application. Specifically, it explains how Blockchain's ability to store immutable data is a good fit for this application.

Key concepts and technologies will also be looked at, including relevant policies and regulations related to private data storage and to EMRs specifically, as well as blockchain technology in general, decentralization, traditional EMR management systems, threat landscape and distributed ledger technology. In addition, specific iterations of already existing systems and software, such as Bitcoin, Ethereum, Hyperledger Fabric, and the Interplanetary File system, will be explained. In addition, the context for selecting an appropriate set of technologies from which to combine in the development of the project's EMR data integrity security demo will be presented.

1.3 Scope and Limits

The scope of this research is limited to look only at concepts which are related to Electronic Health Records, Blockchain Technology, and any technologies and concepts which could be useful in merging the two technologies, as well as concepts which lay a foundation of understanding the core concepts. As an illustrative example, the sections that describe the related research on smart contracts require a foundational knowledge of blockchains, and the sections that describe the related research on blockchains require a foundational knowledge of decentralization and distributed ledgers. Similarly, the sections that describe the related research on smart contracts require a foundational knowledge of distributed ledgers. The ideas that are going to be covered in this paper have been selected with

great care with the intention that each new idea will build on the one that came before it. As a result, the reader will be able to comprehend the complete contents of the paper by adhering to the structure of the document.

Furthermore, this project is scoped specifically to address the way of which blockchain technology can be used in securing the *integrity* of data, and specifically *not* to address *confidentiality* or *availability*, regardless of whether the security related concepts discussed herein affect the confidentiality and/or availability of health record data. This narrowed down, integrity-focused scope extends to the project artifact demo in the sense that the demo will not be built with data availability or confidentiality as a priority, and will instead be intended to solely focus on demonstrating the security mechanisms that secure integrity.

In determining the scope of the project, practical constraints were also taken into consideration. The main limiting factor which affected the scope was the available time and resources allocated to the research project. This research has been conducted by one individual bachelor's student, with limited skills and experience of application development. Because of this, the scope of the project artefact was altered in the sense that it would be created as a minimal viable product, with no front end, and many design-decisions would be made with the goal of facilitating development as much as possible. In addition, there was no available financial budget allocated, and as a result, all of the available resources that were used, including referenced papers and third-party software that was used in the project demo, were free and open source. This also meant that the project demo would not be run on a live blockchain main net, which would incur financial costs in the form of transaction fees. Instead, the demo would be built to run on either live decentralized test nets or a local simulated test network.

1.4 Document Structure

This structure for the final report document consists of three main sections: front matter, document body, and back matter. The Front matter is the first major section to be presented, and it is comprised of the following sub-sections: the Declaration, the Abstract, the Acknowledgements, and the Contents.

The Document Body is the second main section, and it is broken down into the following five subsections:

- Introduction: a description of the problem statement, research objectives, scope, and limits, as well as a description of the structure of the document.
- Literature Review - Which served two purposes: To provide a review of the related literature and research, and explanations of the particular technical concepts used in the project that readers of the paper may not be familiar with.
- Design and Implementation - Describes the steps involved in creating the "artifact" of the project, which in this instance will be the computer program that demonstrates how blockchain technology can be used to keep the data integrity of an electronic medical record.

- Results - Presenting the results of the research, to critically evaluate the findings in relation to the results from the literature review.

- Conclusion - Summarize the entire work, and draw a close to arguments made within it, including a summary of research, an evaluation of the degree of which the research objectives have been met, the research contribution and a discussion around areas of further exploration that could build on this research.

The last main section - the Back matter, includes references and relevant appendixes, such as earlier deliveries related to the bachelor project.

2

Literature Review

This section introduces the essential concepts related to EMR, Cloud Security and Blockchain Technology. It describes the major motivating factors related to security challenges of secure storage, various promising technologies and an introduction to blockchain as well as an overview of the proposed use case of implementing of blockchain to solve the discussed challenges.

2.1 Traditional EMR management systems

The Healthcare industry is still undergoing a massive global digitization in regards to the migration towards electronic medical records. A series of recommendations, policies, legislation and regulations have been put in place to facilitate and accelerate this adoption, such as the Health Insurance Portability and Accountability Act (HIPAA) of 1996 (Centers for Medicare & Medicaid Services 1996), the Health Information Technology for Economic and Clinical Health Act (HITECH) of 2009 and the American Recovery and Reinvestment Act (ARRA) of 2009 (Cappers and Scheer 2009), all of which were created to motivate the implementation of EMR's.

In a recent study "Security and privacy of electronic health records: Concerns and challenges", researchers reviewed the security and privacy of current EMR security systems. They found that EMR's contributed to medical sharing between authorized healthcare providers, resulting in improved overall quality of healthcare delivered to patients, but also warned of the challenges of implementing secure storage and access control systems for the data. They highly recommended that current systems be improved upon, and that an efficient encryption scheme which can easily be applied by both patients and health professionals is developed and applied to EMRs (Keshta and Odeh 2021).

A systematic and comprehensive review from the Victoria University, Melbourne, looked at the strengths and weaknesses of existing implementations of EMR cloud storage solutions in regards to the security and privacy preserving mechanisms of these systems. The review identified three main categories of technological challenges:

1. Trust: A major challenge of EMR cloud storage is that of trust. Data owners are not able to control whether or not their data is being protected and controlled appropriately, and as such a great deal of trust of the cloud service provider is needed. This, paired with the inherent single-point of failure issue of cloud, is one big obstacle to be addressed.

2. Access Control: The researchers point out that access control schemes limit a cloud storage solutions flexibility and scalability due to high complexity of access control schemes. Additionally, it is pointed to the great risk that cloud service providers themselves are in control of and may abuse their access to bypass access control, allowing unauthorized internal personnel or third third parties to see, steal of alter sensitive data.

3. Encryption: Existing encryption schemes such as ABE face challenges of searchability, side channel flexibility and multi-authorization attributes. More lightweight encryption design schemes are needed in order to better protect against privacy leaks.

(Chenthara et al. 2019)

Cloud Computing and Storage

Cloud computing can mean different things to different people. According to researchers at the University of Putra, "Cloud is described as the chain of servers and connections to give a computing benefit for storing the user data." (Daryabar, Dehghantanha, and Choo 2017).

The National Institute of Standards and Technology (NIST) defines it as:

"A model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or cloud provider interaction" (Mell, Grance, et al. 2011).

While cloud computing is experiencing global adoption at a rapid pace, the world is simultaneously facing a significant increase in the amount of data produced by both people and devices, thus increasing the demand for scalability and data security. A systematic literature review by researchers at Delhi Technological University proposed that blockchain technologies provides significant inputs in regards to finding an ideal approach to scalable and secure cloud data storage and processing (Sharma, Jindal, and Borah 2020).

2.2 Why EMR's critically need improved security

This section will attempt to provide insights into the major motivating factors for building better security solutions for EMRs. Both regulatory incentives and mandates, and the increasing risk of industry data breaches.

GDPRs effect on the health care industry

General Data Protection Regulation (GDPR) which was implemented 25th of May 2018, aimed to enhance individual's control and rights of personal data. It applies to any individual within the EU, but extends to apply to all service providers, including government bodies which deals with data from an individual within the EU. The GDPR makes specific references to the storage, processing and sharing of medical and patient data specifically, stating that a data subject has the legal right to apply for access to health information about the data subject. The GDPR also states that any entity which stores sensitive data, which the GDPR specifically deems EMRs as, must take every reasonable step in protecting the sensitive data from breaches. Failure to comply, may impose businesses fines up to 20 million euros, or up to 4% of the business worldwide turnover (*General Data Protection Regulation* 2016).

Researchers at the Tokyo Institute of Technology discussed in a study of Blockchain Based Patient Consent Models for EMRs that blockchain immutability and append-only functionality provides a fundamental challenge of incompatibility with the GDPR. While the GDPR mandates that patients or data subjects have a right to request their personal data be erased, which naturally becomes an issue due if their data was to be stored directly on-chain, due to the inability to erase data from the blockchain. To get around this issue, the researchers proposed a hybrid solution, where the EMR data itself is not stored on a blockchain, but is rather encrypted and stored in a traditional non-blockchain database. In their model, each transaction combines a randomised number with the patient's unique ID number into a unique hash, in order to pseudonymize the patient data. That way, the EMR is stored off-chain, and each EMR's transaction log, or log of changes, is stored on the blockchain, to maintain data integrity. This also allows for EMRs to be erased, enabling GDPR compliance. (Tith et al. 2020)

Threat landscape

In an article on healthcare data breaches published in May 2020, researchers analysed the causes and consequences of data breaches from 2005-2019 in healthcare from numerous globally reputable sources including the HIPAA Journal, PRC Database, The Office for Civil Rights Department of Health and Human Services Report and IMB sponsored Ponemon Institute Reports. The Analysis provides detailed insights into the severity of the risks that EMR systems and thus patients private data face on a global scale. The findings showed that more than 10 billion confidential records were exposed across industries from 2005 to 2009, whereas 43,38% or 3912 of these were from the healthcare sector alone. The study also showed that malicious intentional hacks were the main cause of healthcare data breaches, accounting for 64% of breaches from 2005-2019, and that the portion of healthcare breaches attributed to hacks increased to 92% when only looking at breaches from 2015-2019, suggesting that hacking as a data breach threat is increasing dramatically. The number of individuals globally affected by these data breaches are upwards of 255 million, according to the study (Seh et al. 2020).

"Despite the numerous advantages of EHRs, the digital health data of patients is at huge risk today. As chronicled in our study, data breach trends have also undergone a massive transformation. The comprehensive analysis undertaken in this study reveals that the healthcare industry is the focus of many cyber invaders" (Seh et al. 2020).

The Value Proposition of improving security

In addition to the potential global savings of mitigating costly data security breaches discussed in the previous section, EMR implementation in health care is thought to improve the quality of health care by increasing effectiveness of care and treatment. The cost savings of properly implemented EMR systems are estimated $81 and $162 billion USD annually (MacKinnon and Wasserman 2009).

2.3 Blockchain technology

The interest in Blockchain can be explained by the technology's inherent ability to provide security, anonymity and data integrity without the need for a trusted third party intermediary to control transactions, and this creates promising research areas, particularly in terms of technical challenges and limitations (Yli-Huumo et al. 2016).

"Many sectors, like finance, medicine, manufacturing, and education, use blockchain applications to profit from the unique bundle of characteristics of this technology. Blockchain technology (BT) promises benefits in trustability, collaboration, organization, identification, credibility and transparency"(Leible et al. 2019).

Because of it's immutable append-only function, and a public record of transactions, blockchain technology can provide users with complete transparency over each step a system makes. The result is the creation of an environment which eliminates the need for a trusted authority, as malicious behaviour on the system is technically too difficult to accomplish (Leible et al. 2019).

Distributed Ledger Technology

Distributed ledger technology is an umbrella term used to describe the mechanism of a distributed ledger, produced, stored and audited by independent systems of users or "nodes." The nodes record, share and synchronise transactions by each of them running the same consensus algorithm. The consensus algorithms ensure that the network of nodes all agree on what data to write into the next block in the blockchain. For a transaction to be agreed upon as a valid transaction, and thus applicable to be added to the next block, more than half of the network needs to agree that the transaction is valid. This is what is called "reaching consensus". The most common types of consensus algorithms are called proof of work and proof of stake (Nakamoto 2009; Buterin 2013; Leible et al. 2019).

Both proof of work and proof of stake algorithms are mechanisms which allocate voting power in the consensus protocol. The weight of a user's (or node's) vote is proportional to what quantity of economic resources the node bring to the network. In the case of proof of work, economic resources are computer power or more specifically hash power, and the amount of hash power is proven by running the hashing algorithms. In the case of proof of stake, economic resources are proven by providing digital proof that the user or node has access to a given amount of crypto currency, such as Ethereum, by running the proof of stake algorithm (Nakamoto 2009; Buterin 2013; Leible et al. 2019).

This crucial concept solves the problem of a network attack, because in order to attack the network by a 51% attack, an attacker would need to prove to bring more proof of economic resources provided to the network than the rest of the network, either by having more than 50% of hashpower in the case

of proof of work, or by proving ownership of more than 50% of total staked ethereum, in the case of ethereum 2.0's proof of stake (Nakamoto 2009; Buterin 2013; Leible et al. 2019).

Bitcoin

The first blockchain network to gain massive publicity and adoption was the Bitcoin network. It's mysterious creator by the pseudonym Satoshi Nakamoto, published the Bitcoin: A Peer-to-Peer Electronic Cash System whitepaper in 2009. Whomever Nakamoto is, (s)he envisioned an impenetrable trustless cash-system built on a decentralized open blockchain, open to everyone with an internet connection. Bitcoin runs on a Proof-of-Work algorithm, meaning the nodes of the network, referred to as "miners", spend computing power in order to secure the network and validate transactions (Nakamoto 2009).

The bitcoin network today is able to process 3,1 transactions per second (TPS) at the time of writing, and this value has ranged from 2-4 since 2017 (Blockchain.com 2022).

Ethereum and Smart Contracts

The second largest blockchain network to date, Ethereum, was first proposed by its founder, Vitalik Buterin in 2013, by his publication of the Ethereum White Paper. Ethereum is fundamentally different from Bitcoin though they both share many similarities and are both open source blockchain technologies featuring distributed ledgers, digital currency capabilities and immutability. Ethereum differs from Bitcoin by it's Smart Contract capabilities, allowing uses to deploy immutable and executable applications or programmes to it, making Ethereum a platform for decentralized applications. This makes ethereum not only a distributed blockchain or ledger, but also a decentralized publicly available computer. Additionally, Ethereum aims to switch it's consensus algorithm from proof of work to proof of stake (Buterin 2013; Xu, Chen, and Kou 2019).

Smart contracts are the key functionality which enables decentralized applications (DAPPS) to make use of blockchain functionality. They allow businesses to set up automated contracts on blockchain, removing the need for a third party validator for contract settlement (Buterin 2013; Leible et al. 2019; Xu, Chen, and Kou 2019).

Permissioned or Permissionless blockchains

The are different types of blockchains, and this section will briefly explain the two most common types: Permissioned and permissionless.

An important characteristic which can determine if a blockchain is appropriate for a particular use case or application, is whether the blockchain is permissioned or persmissionless. Permissionless blockchains such as Bitcoin and Ethereum allows anyone access to both read the blockchain's contents and write to it. This access does not require identification, and write access is given to any node which can submit a valid transaction to the network, for a transaction fee. Permissioned blockchains, on the other hand are closed blockchains, which have stricter access control. A permissioned blockchain may limit read and write access to only a set of authorized nodes. The utility of each type of blockchains are as such, different (Wüst and Gervais 2018).

The permissions of a blockchain play a huge role in the blockchain network's decentralization capability. Typically, permissionless blockchains have much greater degrees of decentralization, due to their open access. On the contrary, permissioned blockchains such as Hyperledger are often more centralised, but often offer superior throughput and confidentiality (Wüst and Gervais 2018).

Hyperledger Fabric

"The Hyperledger Project (www.hyperledger.org) is a collaborative effort to create an enterprise-grade, open-source distributed ledger framework and code base. It aims to advance blockchain technology by identifying and realizing a cross-industry open standard platform for distributed ledgers, which can transform the way business transactions are conducted globally"(Cachin 2016).

Like it's permissionless counterparts, Bitcoin and Ethereum, Hyperledger Fabric is a distributed ledger protocol run by decentralized nodes. The Hyperledger protocol distinguishes between validating nodes and non-validating nodes, where validating nodes are responsible for the consensus algorithm, checking transactions and maintaining the distributed ledger. The non-validating nodes functions as proxies which connects clients who are issuing transactions to a validating nodes (Cachin 2016).

Hyperledger's infrastructural and customizable framework could play a prominent role in the future by allowing the creation of a variety of new applications (Leible et al. 2019).

Immutability and integrity

Blockchains achieve data Integrity, meaning the data is accurate, complete, and unchanged from it's original state, by fundamentally only allowing certain actions to take place on the blockchain. Unlike a traditional database or ledger, where data can be added, altered or changed by authorized users, a blockchain only allows data to be added. The main reason for this is that any new block added to a blockchain needs to contain the hash of the previous block. Thus, there are no conditions regarded as valid in the consensus protocols where data already stored on the blockchain can be changed or removed (Hofmann et al. 2017).

Information stored on a blockchain is however not always immutable and append only, as information can also be stored inside smart contract variables, which enables a wide range of information storing functionality. With time and proper implementation, complex data storage applications could be developed utilizing this feature (Buterin 2013).

Blockchain and file storage

The future presents numerous promising opportunities in regards to blockchain technology's potential to innovate and disrupt the cloud storage arena. Blockchain-based cloud storage solutions enable users to exclusively access and control their own data, without needing to trust the third party to secure data integrity, while maintaining the availability and of access and up time of cloud storage (Sharma, Jindal, and Borah 2020).

InterPlanetary File System (IPFS)

Fueled by blockchain's hype over the recent years, demand has also risen for solutions that provide decentralized file storage. Storing large files directly on the blockchain is often very impractical and expensive due to the block size limits and transaction fees of blockchains such as ethereum and bitcoin. One provider of decentralized file storage is IPFS, which has become increasingly popular within the blockchain space due to it's interoperability with many platforms and also its decentralized nature. IPFS is a globally distributed, open source, peer-to-peer file system which allows users to upload any file to the network of distributed peers or nodes, and allows multiple nodes to host the encrypted files, ensuring availability as long as any of the network's nodes still have the file (Nizamuddin et al. 2019).

Web3 storage

Web3.Storage combines the benefits of decentralized storage technologies with the frictionless experience that modern development workflows demand. Web3.storage lets users easily access free decentralized cloud storage. All that is needed is a free API token to use their simple client library or the HTTP API directly. All data is available via a content ID on the public IPFS network, making it compatible with decentralized web tools and applications. The data is saved on the Filecoin network, which has a unique economic mechanism and over 15 terabytes of storage, allowing Web3.Storage to be provided for free (Labs 2021).

2.4 Blockchain technology's proposed use case within EMR storage and access control

Researchers and authors of "Secure Electronic Medical Records Storage and Sharing Using Blockchain Technology" Usman and Qamar (2020) suggests that a blockchain-based EMR system is the key solution for dealing with the sharing of EMRs, explaining that the untamperable nature of blockchains is a good fit for maintaining the integrity of such sensitive data.

This assessment is backed up by Westphal and Seitz (2021), saying: "Blockchain solutions offer efficient approaches for trustworthy data management, especially in the medical field when storing and processing sensitive patient data".

Furthermore, Usman and Qamar (2020) explains that a merge of traditional cloud storage and blockchain-based access management are already being developed and implemented.

Complete and accurate patient data are valuable assets for both patients and doctors Chenthara et al. (2019). Accessible, safe storage which protects the privacy of medical data are as such important issues.

Usman and Qamar (2020) have been successful in developing a blockchain-based EMR management system built on Hyperledger. This system safeguards medical data and makes it possible to distribute it in a secure manner. Patients will not only have access to, but also complete control over, all of their medical information thanks to this system, which gives patients the ability to manage which physicians are granted read and write access to their electronic medical records (EMRs) blockchain.

2.5 Related projects

Another project that has shown great promise is one called “Blockchain Solution to Healthcare Record System using Hyperledger Fabric,” which was developed by students at Frankfurt University of Applied Sciences (Kshitij Yelpale and Kamath 2020). The Blockchain Solution to Healthcare Record System Using Hyperledger Fabric makes use of the consortium blockchain network provided by Hyperledger fabric in order to make it possible for organizations, or hospitals in this particular scenario, to work together in hosting the blockchain. Each organization has the capacity to manage the network, including the addition of physicians, patients, and privileges. This decentralization, despite being relatively minor in comparison to other public blockchains that are more distributed and larger in scale, does provide the system with the increased security that comes from not being centralized. This is beneficial in the event of attacks that would otherwise compromise a system that had a single point of failure. The consortium network would continue to function even if one of the hospitals were taken offline as a result of a natural disaster, a power outage, or a cyber attack; the amount of data that would be lost would be limited or completely mitigated.

Patients using this Hyperledger based EMR management system could view all records in their own EMR, from start to beginning, providing full transparency of the patients transactions and doctor interaction, which is exactly what provides these systems with increased integrity and security. The immutability of transactions, in conjunction with the ability to know who wrote what data at what time into which patient’s records, provides users with the peace of mind that their data has not been tampered with, deleted, or altered. Furthermore, in the event that incorrect data has been appended to their records, the details surrounding this transaction would be available and could then be addressed.

2.6 Literature Review Conclusion

Based on the findings of this literature review, the healthcare industry is the most susceptible to cyber incidents on a global scale, making it one of the most vulnerable sectors overall. It appears that intentional manually executed malicious hacks are the most common type of attack that results in data breaches in the health care industry, and electronic medical records are the primary target of these attacks.

The unique characteristic qualities of blockchain technology appear to have enormous potential for disrupting and innovating the way electronic medical records (EMRs) are secured, and the technology might be the key to making EMRs more resilient. There have already been numerous iterations of blockchain-based EMR systems developed and tested, with encouraging results showing that these systems have the potential to improve the availability, integrity and confidentiality of EMRs. The combination of blockchain technology and EMR systems presents significant challenges in terms of regulatory compliance, particularly with regard to immutability and the right of data subjects to have their data erased.

For the purposes of this project, which is to develop an application that makes use of blockchain technology in order to ensure the data integrity of EMRs, both permissioned and permissionless blockchains would be eligible options to use. Additional functionality could be implemented in a project of a larger scale if there was more time and resources available; in this case, a permissioned

blockchain solution, such as Hyperledger Fabric, would most likely be the most suitable option. However , given the scope of this project, the Ethereum blockchain will suffice and will prove to be a preferable option. This is because the Ethereum platform has a larger developer user base, which means that more development resources are readily available.

3

Implementation

3.1 Design process

This project's goal was to develop a software application that would demonstrate the application of blockchain technology to the task of securing the data integrity of files, particularly medical records. When conducting research on the development of blockchain technology, a number of open source projects that used blockchains in relation to storage and file sharing as well as larger projects featuring full fledged blockchain-based EMR management systems were examined. The "Blockchain Solution to Healthcare Record System using Hyperledger Fabric" (Kshitij Yelpale and Kamath 2020) and the "Securing Information Exchange with Blockchain" project's "Dshare" application (Kedia 2019) are two examples of these types of projects. It would also be essential to get obtain a certain level of proficiency in the programming language Solidity, which is used for coding smart contracts on Ethereum.

As a test subject for the demonstration program, a text file containing an anonymized medical record called 'emr.txt' (3.1) was created. Due to the fact that "EMR" is, in essence, an umbrella term that describes a wide variety of medical files and documents, these can take on a variety of forms. It was decided to use a straightforward text file for the purposes of this project; however, the program is capable of reading and writing any kind of file, and the reader is free to test the program using any file of their choosing.

```
1 Example Randomville State Jane Doe Medical Center
2
3 Outpatient Visit Summary
4
5 Name: Smith, John
```

```
Gender: Male
Date of Birth: 05/29/1968
Race: White (Caucasian)
Etinicity: Not Hispanic or Latino
Language: English
Attending:
PCP:

Visit Information:
Reason(s) for visit:
Aortic Valve Regurgitation,
Abdominal Pain
Astma
Back Pain

Location:
3178 Lyndon Street Pensylvania 18067

Vital Signs/Measurements:
Temperature:
Heart Rate: 60 bpm
Blood Preassure: 120mmHg / 80mmHg
Height:
Weight:
BMI: 24.22 kg/m2

Health Issues/Problems:
Abnormal presence of protein in Urine
Asthma due to seasonal Allergies mainly grass
Bursitis of left hip
Chest Pain

MRN: 123456789
```

Listing 3.1: Sample EMR file

3.1.1 Choice of platform

Initially, this project attempted to deploy and modify an open source project called “Blockchain Solution to Healthcare Record System Using Hyperledger Fabric”, (Kshitij Yelpale and Kamath 2020) which would simulate a network of hospitals, where each hospital would serve as nodes in the Hyperledger Blockchain Network. This would include systems for access control, data sharing, and file encryption, as well as a user-friendly front end that allows users to connect mobile or in-browser ethereum wallets. This would enable users to provide a health practitioner with time-limited access to a patient’s files or records. This approach was later abandoned because the difficulty of putting it into practice was grossly underestimated by the author. During a series of supervisor meetings, the scope of this bachelor’s project would subsequently be narrowed down to create a minimalist program that demonstrated how blockchains can be used to secure data integrity of files - and thus medical records.

Table 3.1 outlines the main differences between Hyperledger and Ethereum as blockchain platforms. As the figure describes, these are two quite different approaches, and their different characteristics plays a key role in application design for each type.

	Hyperledger	Ethereum
Blockchain Type	Permissioned	Permissionless
Degree of decentralization	Low	High
Patient-data storage	Third party/IPFS	Third party/IPFS
Patient data stored on-chain	Possible but impractical	Entirely impractical
Consensus-algorithm	Pluggable	Proof of Work. Planned switch to proof of stake $\sim$august 2022.
Speed/scalability/tps	3 000-20 000	$\sim$13 but might increase to $>$100 000 when switched to proof of stake
Cryptocurrency	None	Ether (Eth)
Transaction fee in $	0 $	$\sim$5 $
Smart Contract language	Go and Java/Javascript	Solidity

Table 3.1: Comparison of Hyperledger Fabric and Ethereum platforms.

Within the new scope of prioritising ease of development over long-term viability the decision to switch from Hyperledger as a blockchain platform to Ethereum was made. The fact that Ethereum is the platform for developing decentralized applications that has seen the most widespread adoption meant that educational resources would be more plentiful, which sped up the process of learning enough to be able to build the demonstration program.

3.1.2 Backend

When it comes to the “backend” of Ethereum development, there are several different options that are viable. As a result of the high transaction fees associated with using the public Ethereum network, a number of free-to-use “Test” networks have emerged, and blockchain developers make extensive use of them. Test networks such as Ropsten, Kavan, and Rinkeby are included in this group. These test networks are comparable to the public Ethereum network in that they are immutable blockchain networks that are decentralized and public. Although they feature the same security mechanisms as the “main net,” the ether tokens on the test networks have no value and are distributed to users through a variety of “faucets” for free. This makes it possible for developers to test decentralized applications and smart contracts without running the risk of unexpected transaction costs being incurred as a result of bugs in their code.

In addition to live test networks, there are open source solutions available that can simulate a blockchain network locally on a machine. This can be done in place of live test networks. Ganache and Hardhat are the names of two popular applications that are used for this purpose. Both are development tools that enable a developer to create a locally hosted Ethereum blockchain and test applications without ever uploading anything to a distributed network. As a blockchain backend for this project, the Hardhat Ethereum development environment was selected because of its comprehensive documentation pages and functionalities such as web3 integration and effective troubleshooting.

3.2 Implementation approach

A “pseudo code” version of the program was created, and the program itself was broken down into the nine distinct steps that must be taken in order for the program to accomplish its objective. The following is a rundown of each step:

1. Initialize local simulated blockchain test network, compile smart contracts and deploy smart contracts to the local blockchain.
2. Request a filename from the user/operator and generate an MD5 hash value from the file.
3. Write the file's MD5 Hash value to a smart contract variable on the blockchain.
4. Upload the file to the IPFS decentralized cloud storage and save related CID value.
5. Prepare scripts to query the blockchain for the MD5 hash value.
6. Download the file from IPFS using the CID value provided upon uploading.
7. Generate the MD5 hash from the downloaded file and save it.
8. Call the smart contract that stores the MD5 hash value to read the value, and save it locally.
9. Compare the two hash values, to verify the file's integrity after it has been uploaded and downloaded. Tell the user/operator if the file's integrity has been verified.

After a structure for the program had been established, the process of coding each step could begin. Python was selected as the language to be used for programming because it was the language that the author had the most experience with. As a result, the author was spared the necessity of learning an additional programming language. Python is used to program the primary structure, but numerous other programming languages and protocols, such as Golang, Java, and Solidity, are also being utilized.

3.2.1 Smart Contracts

The creation of a smart contract that contained a string variable allowed for the successful archiving of data on the blockchain. The contract would necessarily require the availability of functionality for reading and writing to the variable. A straightforward "call," which is a request made to the network to read the contents of a public variable, is all that is required to read data from an Ethereum smart contract. It is also possible to make variables private, or more specifically, to make them so that only certain Ethereum wallet addresses are able to read their contents. A private variable was selected to be used for the purpose of this demonstration.

Unlike the read or "call" operations, writing to or modifying a smart contract variable requires sending a transaction to the smart contract. This means that writing or editing variables has a financial cost, known as a transaction fee, which is paid by the issuer of the write request. The fee often referred to as "gas" is paid in "ether," which is the native currency of the Ethereum network.

Because writing to the blockchain requires a transaction, each modification to the smart contract variable will be recorded in a version log that is open to public inspection after the transaction has been completed. Users are able to check which wallets interacted with a contract and at what time using websites such as `etherscan.io`, which allow anyone to view the complete history of transactions. This feature is one that makes it possible for decentralized applications to be easily audited. It also has the potential to facilitate complex applications for file integrity verification and version control, as well as EHR management systems that could enable patients to view an immutable log of every health practitioner that interacted with their personal health-record-contract.

```
// SPDX-License-Identifier: MIT
pragma solidity ^0.8.0;
import "hardhat/console.sol";

contract Hashstorage {
    string private filehash;

   constructor(string memory _filehash) {
       console.log("", _filehash);
        filehash = _filehash;
    }

    function getfilehash() public view returns (string memory) {
        return filehash;
    }

    function setfilehash(string memory _filehash) public {
        console.log("Setting filehash value from '%s' to '%s'", filehash, _filehash);
        filehash = _filehash;
    }
}
```

Listing 3.2: Hashstorage solidity smart contract

Listing 3.2 shows the source code of this project's solidity smart contract Hashstorage. The contract has been declared compatible with the compilers for solidity version 0.8.0, and it imports the built-in console.log integration that is included with hardhat. This makes it possible to print messages and contract variables as outputs. The fact that the filehash string is marked as "private" indicates that functions that are not part of the Hashstorage contract are unable to modify the value of the contract variable. The contract includes two functions called getfilehash() and setfilehash(). These functions retrieve and set the value of the filehash variable, respectively.

3.2.2 IPFS/Web3 storage

Integration of Web3.storage IPFS is a vital part of this project and is one of its primary components. Web3.storage is being used for the purposes of this project in order to upload files to the IPFS network of decentralized cloud storage nodes utilizing a web3.storage API. NodePackageManager is responsible for the installation of the web3.storage client. This is accomplished by running the "npm install web3.storage" command, which installs the library along with all of its dependencies and files. It is necessary to provide a web3.storage API Token in order to make use of the web3.storage JavaScript client library. Tokens can be obtained by registering for a free account on web3.storage.com, which is currently the only way to do so. It is necessary to include the API token, which is unique to the account, in the file upload command, as demonstrated in listing 3.3. By default, the author of this project has provided his own web3 API token, which can be optionally switched out by editing the value depicted in listing 3.3

```
#STEP 4: Uploads file to ipfs using web3.storage.

file_name = input('Enter filename to upload to IPFS: ')
stream = os.popen('node web3-storage/put-files.js --token=<token> '+file_name)
output = stream.read()
```

Listing 3.3: Web3.storage IPFS integration

When files are uploaded to IPFS, each file is assigned a CID value, which stands for “Content Identifier.” This value is what is used to identify files that are hosted on the IPFS network. CID values are file hashes, which means that they do not represent a transaction or location. Instead, they are similar to a file’s fingerprint and serve as a unique identifier for the file. Hashing a file on IPFS makes use of the SHA-256 algorithm by default. SHA-256 is a cryptographic hashing algorithm. For the purposes of this project, the CID value will be saved in a text file that will be given the name cidstorage.txt, and it will be utilized when re-downloading files from IPFS.

3.2.3 Hardhat

Hardhat offers the opportunity to interact with smart contracts in a direct manner by means of a command line interface as well as by executing scripts written in JavaScript. “deploy.js,” “GetHashStorage.js,” and “SetHashStorage.js” are the names of the three scripts that were developed, all of which are utilized by hardhat in the process of interacting with smart contracts and deploying them.

3.2.4 Step-by-step explanation of the program

In this section, a concise explanation is provided for each of the actions that the program performs in order to demonstrate the blockchain-integrated file integrity check.

1. Performs a “cleaning” of the hardhat environment, followed by the compilation of smart contracts and the deployment of those contracts to a locally hosted test network of the Ethereum blockchain. Additionally, it saves the blockchain Transaction ID for each contract in their own individual text files so that it can be accessed at a later time.

2. The operator is prompted for a filename before an MD5 hash of the file is generated. After that, the hash is recorded in the uploadhash.txt file so that it can be accessed at a later time. It then executes the ’editHashstorage.py’ program, which reads the Transaction IDs from the ’hashstorageTXID.txt’ file and the file hash from the ’uploadhash.txt’ file. It then writes these values into two javascript files named ’Sethashstorage.js’ and ’GetHashStorage.js’ respectively. Later on, when it comes time to write the file hash to the smart contract that is hosted on the blockchain, these two JavaScript files will be used.

3. In this step, a hardhat command is executed with the SetHashStorage.js file in order to interact with the smart contract and write the file hash to the smart contract called “Hashstorage.” The integrity of the hash is safeguarded by the blockchain network from the moment it is recorded in the storage variable of the smart contract. In the future, the smart contract will be queried, also known as “called,” in order to retrieve the original unmodified hash value.

4. This function uses the web3.storage module to upload the file to the Interplanetary File System (IPFS). This results in the generation of a ’CID’ value, which is then stored in the text file ’cidstorage.txt’ for later use. The ’CID’ will be required at a later point in time in order to successfully download the file from the ’IPFS’ network.

5. During this stage, the “editHashstorage.py” program is executed, which then modifies the variables found in the “sethashstorage.js” file.

6. Using the integration between ipfs and web3.storage, the file is downloaded from the IPFS network. The CID value is read from the cidstorage.txt file, and then the command “ipfs get cid-value¿ / file-name¿” is executed, which downloads the file from IPFS that is associated with the CID value.

7. In this step, an MD5 hash is generated from the file that was downloaded, and the hash is written into a file called “downloadhash.txt.”

8. During this stage, the hardhat command npx hardhat run —network localhost is executed with the GetHashStorage.js script prefix. This step calls the HashStorage smart contract and retrieves the hash value from the blockchain. After that, the hash value that was read from the blockchain is inserted into the file called “uploadhash.txt.”

9. This final step takes the hash value that was just read from the smart contract on the blockchain and compares it to the hash value that was generated from the file that was downloaded using IPFS. “uploadhash.txt” is the file that contains the hash value. It determines whether or not the two hashes are identical, and if they are, it displays the message “Data Integrity Check Success” in the terminal. This notifies the user that the file has not been altered from the time it was uploaded until the time it was downloaded.

3.2.5 Steps to reproduce

The following paragraphs will describe the steps that need to be taken in order to replicate and run this program on another computer.

Prerequisites and Dependencies

Although it may be compatible with a wide variety of other operating system versions and distributions, it is recommended that this program be installed on an Ubuntu 20.04.4 operating system. Both Ubuntu 20.04.4 and Kali 2022.1 have been utilized in the testing for the following guide. It is expected that the following tools will be installed on the system, each in the appropriate version.

Node Version Manager (NVM) version 0.39.1, Node version 16.4.2, Node package manager (NPM) version 7.18.1, Git version 2.25.1 and Python version 3.8.10 or 3.10.4.

```
1 $ nvm install 16.4.2
2 $ nvm use 16.4.2
3 $ nvm alias default 16.4.2
4 $ node --version && npm --version
```

Step by step guide to reproduce

Download the project repository from github.

```
1 $ git clone https://github.com/lite-cyber/blockchain-demo.git
```

Proceed by moving into the ’blockchain-demo’ project directory.

Initialize project with NPM init

```
$ npm install hardhat
$ npm init -y
```

The next step is to install the nomiclabs hardhat-ethers dependencylibrary, then bootstrap the project by executing npx hardhat and beginning the process of running a hardhat node.

```
$ npm install @nomiclabs/hardhat-ethers ethers
$ npx hardhat
$ npx hardhat node
```

This should now show an output of 20 generated ethereum key-pairs. (Never, under any circumstances, should you transfer actual currency to these addresses, as they are default development accounts.) Keep the terminal window open, then open a second terminal and navigate to the blockchain-demo directory.

In the new terminal, check to see that the current versions of Node and NPM are still 16.4.2 and 7.18.1. After that, install the web3.storage IPFS dependency using the commands below.

```
$ node --version && npm --version
$ wget https://dist.ipfs.io/go-ipfs/v0.12.2/go-ipfs_v0.12.2_linux-amd64.tar.gz
$ tar -xvzf go-ipfs_v0.12.2_linux-amd64.tar.gz
$ cd go-ipfs
$ sudo bash install.sh
$ ipfs --version
// should show IPFS version 0.12.2

$ ipfs init
$ ipfs daemon
```

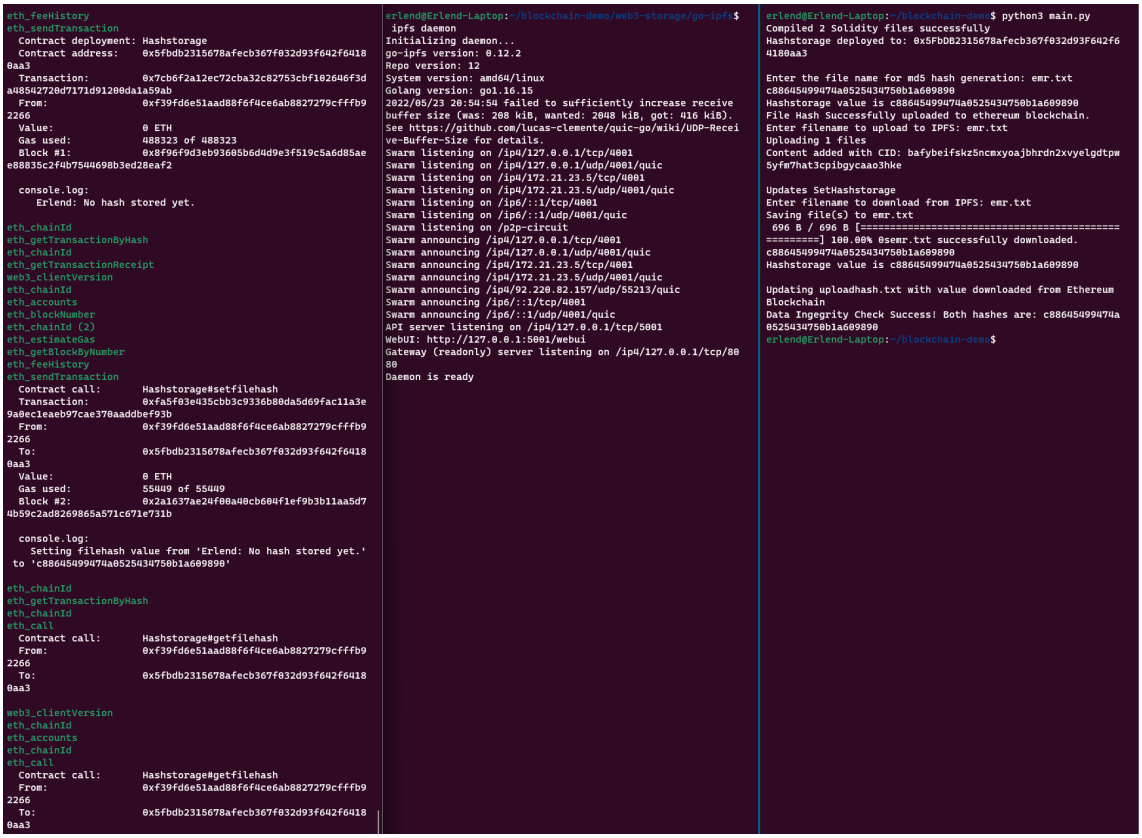
At this point, the IPFS daemon should be operational. Make sure that the final output states that the daemon is ready. Keep the currently active ipfs daemon terminal open, then launch a new terminal and navigate to the 'blockchain-demo' project directory.

First things first, let's check the versions of node and npm in this new terminal. After that, cd into the web3-storage directory and then run npm install. Returning to the blockchain-demo directory, execute the main.py Python file with Python 3. Make sure that the same filename (including any file extensions) is given in response to all three of the prompts.

```
$ cd web3-storage
$ npm install
$ cd ..
$ python3 main.py
```

The final output should say "Data Integrity Check Success! Both hashes are "hash" as demonstrated in figure 3.1. This program should be able to upload any file. IPFS does not have a maximal file size limit, but there is a maximal total storage limit tied to an account which is 1 Terabyte for free accounts.

As depicted in figure 3.1, the artefact demo program requires three separate processes in order to function. The first terminal runs the Hardhat blockchain node which needs to run in the background for the main program to function. The second terminal is for running the IPFS daemon, which allows the



(a) sample1

Figure 3.1: All three terminals with their final outputs

web3.storage integration to upload files through it and onto the decentralized IPFS storage network. The third terminal is where the main python file is eventually executed, and where the operator or user is prompted to write which file the program will hash, upload, download and perform an integrity check on.

The first terminal, depicted on the left in figure 3.1 shows four separate interactions with the blockchain. The first interaction is the contract deployment of the contract 'Hashstorage', which in figure 3.1 is deployed and given the address '0x5fb...0aa3'. The output reveals that the contract was sent in the transaction '0x7cb...59ab', from the ethereum address '0xf39...2266' costing 488323 'Gas', and that it was included in Block number 1, which has the block ID of '0x8f96...eaf2'. (Hashes have been shortened for the purpose of this explanation.)

The next interaction of the blockchain is a 'Contract call' of the function 'setfilehash' of the contract 'Hashstorage'. Because this interaction calls a function that writes to the blockchain, it requires a transaction to made, and thus the transaction id '0xfa...f93b' is included. The call is sent from the same address that initially deployed the contract, '0xf39...2266', and it it sent to the address '0x5fb...0aa3' which is the contract address of the Hashstorage contract. The transaction cost 55449 'Gas' it is included in block 2 which is given the block ID of '0x2a...731b'. Lastly the console.log output reveals that the filehash value is changed from 'no value' to now contain a value of 'c886...9890' - which is the MD5 hash value of the emr.txt file chosen by the operator in the third terminal.

The next two calls depicted in the first terminal on the left of figure 3.1 are calls to the 'getfilehash'

function of the 'Hashstorage' contract. Because these calls do not write any data, but merely reads from the contract, there is no transaction ID, gas cost or block number associated with the calls, only addresses of the sender and receiver of the call. The operator of the program is able to read the emr.txt file's hash value from the blockchain entirely for free.

4

Results

4.1 Research results

The research presented in the Literature Review in chapter 2 indicates not only that there is a demand for improved data security within the health sector, but also that EMRs in particular are a popular target for malicious hackers and that EMRs today are prone to large scale data-leaks as a result of centralized storage solutions. This demand for improved data security is in addition to the fact that there is a demand for improved data security within the health sector. The study's findings indicated that the unique characteristics of decentralized blockchain technology can be used in conjunction with decentralized Peer to Peer storage solutions like the Interplanetary File System to achieve decentralized storage of patient records utilizing blockchains to ensure the records' integrity. This can be accomplished by combining the two types of storage solutions. According to the findings of the research, these kinds of solutions have already been proposed and tested, and the results have shown that they show promise in terms of data security and scalability. Trust, access control, and encryption were found to be the three primary challenges that are presented by traditional EMR management systems in Chapter 2 Section 2.1. Section 2.3 of Chapter 2, discussed the way that public blockchain networks like Bitcoin and Ethereum use consensus algorithms to solve the problem of trust. This makes these networks inherently trustless systems and eliminates the requirement for third party intermediaries to validate transactions or data changes.

Public and private key encryption is already a inherent and integral mechanism used in most blockchains. In order to transact on the various blockchain networks, peers use their key pairs in order to send and receive currency, and also to interact with smart contracts. As such, the identity of users in the blockchain environment is essentially already encrypted in that a ”user“ of a blockchain network is

identified by their public wallet address, and not by their personal ID. This mechanism paired with decentralized file storage opens up the possibility for applications to be created where files are encrypted with user's wallet's public keys, before they are stored on the decentralised storage networks. In this example, the data stored will only be accessible or 'readable' by a user which can provide the corresponding key which was used to encrypt the file. Additionally, smart contracts allow for complex programs to be developed in the future where users may choose which other users can also access their encrypted data, essentially creating blockchain-based access control.

4.2 Implementation results

The program that was developed and discussed in chapter 3 showed how these technologies can work together to not only store data securely on decentralized storage-platforms, but also how the Ethereum protocol, through its smart contract capabilities, can be used to verify the data integrity of files that are stored on decentralized protocols. This was demonstrated by the fact that the program was able to demonstrate how these technologies can work together to store data securely on decentralized storage-platforms. The program demonstrated that such a solution could be developed without cost, using only open source tools. The program that was demonstrated in chapter 3 demonstrated that the technologies that are bringing millions of nodes together in creating a massive network dedicated to achieving consensus around the data integrity of the blockchain can be tapped into by the public, to use the Ethereum decentralized computer to secure not only the integrity of the Ethereum ledger, but also the integrity of an individual's files.

In spite of the fact that the demo project was hosted locally, it was still able to demonstrate the concepts of how an arbitrary application can tap into the network of resources that are available through the use of decentralized blockchain technologies. This is consistent with what the research from chapter 3 indicated. The literature review and the program demonstration suggest that there may be a vast amount of potential that has not yet been taken advantage of in the further development of applications that make use of these immutability technologies in the creation of more transparent, more secure, and more immutable decentralized methods of data storage.

The implementation demonstrated that a medical record stored on the IPFS Network would benefit from increased availability due to the decentralized storage, without risking data tampering, because the file itself is re-downloaded based on a hash-value (CID) as opposed to a file location of a traditional database. This prevents the data from being altered. The implementation, in an essence, demonstrated that a medical record kept on the IPFS Network would have improved accessibility. It is possible for the hash value to serve as an immutable fingerprint that is used for integrity verification when it is read from the blockchain prior to being compared to the hash value of the file that was downloaded. This helps to ensure that the data is accurate. The hash of the file is stored on the blockchain, which allows the Ethereum protocol to ensure that the data in the file has not been tampered with. Through the utilization of a smart contract variable, it is possible to render the hash value of the file to be immutable.

5

Conclusion

5.1 Introduction

The primary objective of the research was to investigate the potential applications of blockchain technology in the medical field, with the end goal of improving the data integrity of medical records. The Literature Review discovered that a number of researchers had the opinion that the combination of these technologies could be beneficial. It was demonstrated how this could be accomplished in the implementation chapter of this report, with a minimal viable product demonstrating a blockchain-based data integrity verification solution. This was done to show how this could be accomplished. This chapter will attempt to provide a summary of the findings from both the research on the relevant literature and the results of the implementation in order to arrive at some conclusions regarding the research objectives and to provide some recommendations for additional research.

5.2 Summary of Research

The bachelor's project was provided with information that was both necessary and absolutely essential thanks to the literature review. The remainder of the project became simpler and easier to comprehend as a result of the creation of a solid knowledge base. In the process of developing the project artifact, the literature review acted as a foundational knowledge base that was helpful in determining design decisions.

Initially it was proposed to develop a complex and feature-packed hyperledger-based EMR-management system for the implementation part of the research. It was presented how such a system would work,

and an effort was made to build on previous iterations of such an artefact. Several attempts of deploying existing complex hyperledger-projects which would be used in testing the various security-mechanisms of a hyperledger-network were made, but difficulty of deploying such a solution eventually proved to be outside of a feasible time frame to achieve a working artefact. This lead to the proposal of a much simpler, ethereum based artefact, which would then go on to be developed from scratch.

A collection of the technologies outlined in the literature review was utilized in the production of the artifact based on Ethereum that is discussed in chapter 3. It was successful in demonstrating a way of using blockchain technology for verifying the integrity of data while also making use of a solution for decentralized data storage. - removing, at least in theory, the possibility of a failure at a single point in both the data storage and the hash value storage.

5.3 Research Objectives

The purpose of this study was to investigate how blockchain technology could be applied to the task of storing EMRs, as well as the extent to which this technology can enhance the level of security offered by more conventional methods of storing and sharing data. In particular, the purpose of this study was to investigate whether or not the proposed blockchain-based cloud-EMR management systems could improve the security and integrity of medical records.

On the basis of these findings, one can reach the conclusion that blockchain technology has the potential to be utilized to enhance and protect the data integrity of files, and consequently EMRs. According to the findings of the research, a variety of tools are currently available that make it easier to develop solutions that combine blockchain technology with peer-to-peer decentralized storage solutions and that are also interoperable with one another. This indicates that the development of EMR storage solutions that make use of decentralized blockchain networks that are freely accessible to the public is possible, and that these solutions can make use of immutable blockchain ledgers to store hashes of files that can be used to verify the integrity of files that are stored on decentralized peer-to-peer storage protocols. This is significant because it means that the development of EMR storage solutions is feasible.

In addition, the study came to the conclusion that these problems could be solved by utilizing open source protocols, such as Hyperledger Fabric, the Ethereum network, Hardhat, Web3.storage, and the IPFS protocol, which are all available at no cost to developers. According to the findings of this research, open and public blockchain platforms like Ethereum have an advantage over private consortium blockchains like Hyperledger in the sense that they are significantly more decentralized and, as a result, more resistant to attacks than the latter. According to the findings of the research, the level of network security increases in direct proportion to the amount of financial resources that are collectively contributed to the blockchain network by its users. This makes the network more resistant to disruption.

5.4 Research Contribution

The results of this study make multiple important contributions to the existing body of scholarly research. First, the 'blockchain-demo' artifact that was developed as part of this research demonstrated

decentralized data storage and integrity verification using web3.storage, IPFS, Ethereum, and Hardhat. This was accomplished. Using hashes that were stored on smart contracts on a simulated ethereum network, it was demonstrated that blockchain technology could be used to verify the authenticity of a medical record that was kept on a peer-to-peer storage protocol. This was accomplished by using a simulated version of the ethereum network. The hash that would be used to verify the data integrity of the EMR file, if modified to deploy to the public Ethereum main net, would be secured by over a thousand Terra-hashes per second of pooled computing resources protecting the integrity of the blockchain. This would be the case if the EMR file was modified to deploy to the public Ethereum main net.

The insights that were gained from this research could be of use to researchers as well as developers. This study makes a contribution to our understanding of how blockchains are secured, which mechanisms they use for maintaining the data integrity of the decentralized ledger, how decentralization of networks increases network security by removing single points of failure from the network, and how developers can make use of these security mechanisms when creating new solutions for scalable and secure data storage.

The findings will be of interest to developers who are looking for suitable suites of combined development tools that will work together, as the project demonstrates interoperability of several different technologies and to some extent explains how they work together and why they were chosen over others. Last but not least, the findings are of particular interest to individuals or organizations that are looking for free and open source solutions for doing blockchain development, whether they are researchers, students, or professional developers. This is because the findings show that there are a number of different solutions available.

5.5 Future Work

Researchers have a wide variety of opportunities to build upon the work that has been done here. Four specific areas, namely Mainnet or Test Net integrations, encryption, access control, and file sharing, will be the focus of the conversation regarding areas for improvement.

The objective of future research should be to broaden the capabilities of the local simulated Hardhat network so that it can provide support for Ethereum live test networks like Ropsten and Rinkeby. The researchers should, at some point in the future, also develop a version of the application or add an option to use the live Ethereum main net to run the application. This would make it possible for the data stored on a blockchain to be *truly* decentralized, which would increase the security of the data that is stored. Researchers would then be required to exercise extreme caution before uploading any sensitive information to the live network, as it stands a chance of being saved indefinitely.

A further area of research that builds on this work would be to encrypt files before they are uploaded to IPFS. This would be an expansion of the previous work. If real data sets are going to be used, this step is absolutely necessary for the protection of the EMR data. This encryption could make use of Ethereum wallet key pairs, and it could be made compatible with other popular wallets like Metamask. Researchers ought to make an effort to discover the method that is the most appropriate for encrypting data that is integrated with blockchains.

Building on the previous example, once the integration of data encryption using Ethereum wallet key pairs has been developed, researchers or other interested parties could develop solutions for file sharing and access control. These solutions would involve encrypting files with the public key of another user so that they could be downloaded by that user, who would then use his or her private key to decrypt the file. It is possible that the same user could be granted access to retrieve the file hash that is stored on the blockchain smart contract. This would allow the user to call the smart contract and verify the shared file's integrity using the blockchain hash storage smart contract.

Lastly it is recommended that future research be directed toward the development of a user-friendly front end for the application so that users can interact with it without having to make use of a command line interface. It's possible that the front-end could be modified to include interoperability with the suggested improvements from earlier.

If all of these enhancements were pursued, a finished product that is suitable for use in the healthcare industry might not be too far off. This product might also have the potential to be tested by medical professionals on actual patients as part of a larger research project, if the appropriate steps were taken.

Bibliography

Blockchain.com (2022). *Transactions-per-Second*. URL: https://blockchain.com/charts/transactions-per-second.

Buterin, Vitalik (2013). 'Ethereum White Paper: A Next Generation Smart Contract & Decentralized Application Platform'. In: URL: https://github.com/ethereum/wiki/wiki/White-Paper.

Cachin, Christian (2016). 'Architecture of the Hyperledger Blockchain Fabric'. In.

Cappers, Peter and Rich Scheer (Mar. 2009). *American Recovery and Reinvestment Act of 2009: Final Report on Customer Acceptance, Retention, and Response to Time-Based Rates from Consumer Behavior Studies*. Tech. rep. DOI: 10.2172/1424221. URL: https://doi.org/10.2172/1424221.

Centers for Medicare & Medicaid Services (1996). *The Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. Online at http://www.cms.hhs.gov/hipaa/.

Chenthara, Shekha et al. (2019). 'Security and Privacy-Preserving Challenges of e-Health Solutions in Cloud Computing'. In: *IEEE Access* 7, pp. 74361–74382. DOI: 10.1109/access.2019.2919982.

Chowdhury, Mohammad Jabed Morshed et al. (2018). 'Blockchain versus database: A critical analysis'. In: *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. Ieee, pp. 1348–1353.

Daryabar, Farid, Ali Dehghantanha, and Kim-Kwang Raymond Choo (2017). 'Cloud storage forensics: MEGA as a case study'. In: *Australian Journal of Forensic Sciences* 49.3, pp. 344–357. DOI: 10.1080/00450618.2016.1153714. eprint: https://doi.org/10.1080/00450618.2016.1153714. URL: https://doi.org/10.1080/00450618.2016.1153714.

Filippi, Primavera De (2013). 'Flawed cloud architectures and the rise of decentral alternatives'. eng. In: *Internet Policy Review* 2.4, pp. 1–10. ISSN: 2197-6775. DOI: 10.14763/2013.4.212. URL: http://hdl.handle.net/10419/213976.

General Data Protection Regulation (2016). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%5C%3A32021R0876.

Hofmann, Frank et al. (2017). 'The immutability concept of blockchains and benefits of early standardization'. In: *2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K)*. Ieee, pp. 1–8.

Kedia, Harsh (Aug. 2019). *hKedia/dShare: First release of dShare built with Ethereum*. Version v1.0.0. DOI: 10.5281/zenodo.3359852. URL: https://doi.org/10.5281/zenodo.3359852.

Keshta, Ismail and Ammar Odeh (July 2021). 'Security and privacy of electronic health records: Concerns and challenges'. In: 22.2, pp. 177–183. DOI: 10.1016/j.eij.2020.07.003. URL: https://doi.org/10.1016/j.eij.2020.07.003.

Kshitij Yelpale, Jathin Sreenivas and Varsha Vasudev Kamath (2020). 'Blockchain Solution to Healthcare Record System using Hyperledger Fabric'. In: URL: https://github.com/kshitijyelpale/blockchain-hyperledger-fabric-electronic-patient-records/blob/main/docs/Final%5C%20report.pdf.

Labs, Protocol (2021). *Web3.storage Documentation*. URL: https://web3.storage/docs/ (visited on 05/23/2022).

Leible, Stephan et al. (2019). 'A Review on Blockchain Technology and Blockchain Projects Fostering Open Science'. In: *Frontiers in Blockchain* 2, p. 16. ISSN: 2624-7852. DOI: 10.3389/fbloc.2019.00016. URL: https://www.frontiersin.org/article/10.3389/fbloc.2019.00016.

MacKinnon, William and Michael Wasserman (2009). 'Implementing electronic medical record systems'. In: *IT professional* 11.6, pp. 50–53.

Mell, Peter, Tim Grance, et al. (2011). 'The NIST definition of cloud computing'. In.

Nakamoto, Satoshi (2009). *Bitcoin: A peer-to-peer electronic cash system*. URL: http://www.bitcoin.org/bitcoin.pdf.

Nizamuddin, N. et al. (2019). 'Decentralized document version control using ethereum blockchain and IPFS'. In: *Computers Electrical Engineering* 76, pp. 183–197. ISSN: 0045-7906. DOI: https://doi.org/10.1016/j.compeleceng.2019.03.014. URL: https://www.sciencedirect.com/science/article/pii/S0045790618333093.

Seh, Adil Hussain et al. (May 2020). 'Healthcare Data Breaches: Insights and Implications'. In: 8.2, p. 133. DOI: 10.3390/healthcare8020133. URL: https://doi.org/10.3390/healthcare8020133.

Sharma, Pratima, Rajni Jindal, and Malaya Dutta Borah (Aug. 2020). 'Blockchain Technology for Cloud Storage: A Systematic Literature Review'. In: *ACM Comput. Surv.* 53.4. ISSN: 0360-0300. DOI: 10.1145/3403954. URL: https://doi.org/10.1145/3403954.

Tith, Dara et al. (Oct. 2020). 'Patient Consent Management by a Purpose-Based Consent Model for Electronic Health Record Based on Blockchain Technology'. In: 26.4, pp. 265–273. DOI: 10.4258/hir.2020.26.4.265. URL: https://doi.org/10.4258%5C%2Fhir.2020.26.4.265.

Usman, Muhammad and Usman Qamar (2020). 'Secure Electronic Medical Records Storage and Sharing Using Blockchain Technology'. In: *Procedia Computer Science* 174. 2019 International Conference on Identification, Information and Knowledge in the Internet of Things, pp. 321–327. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2020.06.093. URL: https://www.sciencedirect.com/science/article/pii/S1877050920316136.

Westphal, Erik and Hermann Seitz (2021). 'Digital and Decentralized Management of Patient Data in Healthcare Using Blockchain Implementations'. In: *Frontiers in Blockchain* 4, p. 36. ISSN: 2624-7852. DOI: 10.3389/fbloc.2021.732112. URL: https://www.frontiersin.org/article/10.3389/fbloc.2021.732112.

Wüst, Karl and Arthur Gervais (2018). 'Do you need a blockchain?' In: *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*. Ieee, pp. 45–54.

Xu, Min, Xingtong Chen, and Gang Kou (July 2019). 'A systematic review of blockchain'. In: 5.1. DOI: 10.1186/s40854-019-0147-z. URL: https://doi.org/10.1186/s40854-019-0147-z.

Yli-Huumo, Jesse et al. (Oct. 2016). 'Where Is Current Research on Blockchain Technology?—A Systematic Review'. In: *Plos One* 11.10, pp. 1–27. DOI: 10.1371/journal.pone.0163477. URL: https://doi.org/10.1371/journal.pone.0163477.

A

Artefact Source code

https://github.com/lite-cyber/blockchain-demo

B

Short Paper

The use of blockchain technology in securing data integrity of electronic medical records

1st Erlend Halsnes
Cyber Security
Noroff University College
Kristiansand, Norway
erlend.halsnes@stud.noroff.no
Supervisor: Fabricio Bortoluzzi

***Abstract*—One of the biggest problems of healthcare today is that organisations hold multiple fragmented health records about patients, but lack secure enough platforms to safely store and process these records. As medical records are a critical asset used to diagnose and treat patients, the information which they hold, such as information about patients physical and mental health history, are seen as highly sensitive and confidential. This project aims to look for the possibilities of emerging disruptive technologies to improve data security within the healthcare industry, and specifically how blockchain technology could be used to secure data integrity of electronic medical records (EMRs). Blockchain-based alternatives to EMR storage and access control may provide significant security improvements within this field, and may be the key technology for effectively securing and storing EMRs securely, globally. This project aims to identify, configure and demonstrate a working blockchain-based EMR solution, and test its ability to secure the integrity of EMRs.**

***Index Terms*—Blockchain, Cloud Security, Decentralization, Distributed Ledger Technology, Electronic Medical Records**

I. Introduction

This research aims to explore how blockchain technology can be used for the purpose of storing EMRs, and to which degree the technology can improve on traditional storage and sharing methods in terms of security. More specifically this research aims to look at how proposed blockchain-based cloud EMR management systems could provide improved security of integrity. The literature review will discuss existing secondary literature regarding the concepts and potential implementations of such a system, what the status of the research is, and attempt to explain how these system would operate and be created.

Sharing of EMRs is challenging because the data is prone to several risks such as theft, data loss, unauthorized access, data leaks, data tampering and more.

In order to implement a functioning demonstration of a blockchain-based storage and access control for EMRs, the following section will aim to provide an overview of the related research conducted on relevant topics. It will dive into the motivation of why a better solution is needed, how it can be created, what it's potential may be, and explain how Blockchain's inherent "natural" features are a good fit for this application.

Key concepts and technologies will also be looked at, including relevant policies and regulations related to private data storage and to EMRs specifically, as well as blockchain technology in general, decentralization, traditional EMR management systems, threat landscape and distributed ledger technology. Additionally, specific iterations of existing systems and software such as Bitcoin, Ethereum, Hyperledger Fabric, the Interplanetary File system will be explained for context, as well as provide the background for choosing an appropriate set of technologies of which to combine in development of the project's EMR management system demonstration.

II. Related Work

This section introduces the essential concepts related to EMR, Cloud Security and Blockchain Technology, briefly describing the major motivating factors related to security challenges of secure storage, various promising technologies and an introduction to blockchain, as well as an overview of the proposed use case and implementation of blockchain to solve the discussed challenges.

A. Traditional EMR management systems

The Healthcare industry is still undergoing a massive global digitalization in regards to the migration towards electronic medical records. A series of recommendations, policies, legislation and regulations have been put in place to facilitate and accelerate this adoption, such as the Health Insurance Portability and Accountability Act (HIPAA) of 1996 [1, p. 1], the Health Information Technology for Economic and Clinical Health Act (HITECH) of 2009 and the American Recovery and Reinvestment Act (ARRA) of 2009 [2], all of which were created to motivate the implementation of EMRs.

In a recent study "Security and privacy of electronic health records: Concerns and challenges", researchers reviewed the security and privacy of current EMR security systems. They found that EMR's contributed to medical sharing between authorized healthcare providers, resulting in improved overall quality of healthcare delivered to patients, but also warned of the challenges of implementing secure storage and access control systems for the data. They highly recommended that current systems be improved upon, and that an efficient encryption scheme which can easily be applied by both patients and health professionals is developed and applied to EMRs [3].

A systematic and comprehensive review from the Victoria University, Melbourne, looked at the strengths and weaknesses of existing implementations of EMR cloud storage solutions in regards to the security and privacy preserving mechanisms of these systems. The review identified three main categories of technological challenges:

1) Trust: A major challenge of EMR cloud storage is that of trust. Data owners are not able to control whether or not their data is being protected and controlled appropriately, and as such a great deal of trust of the cloud service provider is needed. This, paired with the inherent single-point of failure issue of cloud, is one big obstacle to be addressed.
2) Access Control: The researchers point out that access control schemes limit a cloud storage solutions flexibility and scalability due to high complexity of access control schemes. Additionally, it is pointed to the great risk that cloud service providers themselves are in control of and may abuse their access to bypass access control, allowing unauthorized internal personnel or third third parties to see, steal of alter sensitive data.
3) Encryption: Existing encryption schemes such as ABE face challenges of searchability, side channel flexibility and multi-authorization attributes. More lightweight encryption design schemes are needed in order to better protect against privacy leaks.

[4]

B. Cloud Computing and Storage

Cloud computing can mean different things to different people. According to researchers at the University of Putra, "Cloud is described as the chain of servers and connections to give a computing benefit for storing the user data." [5].

The National Institute of Standards and Technology (NIST) defines it as:

"A model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers,storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or cloud provider interaction" [6].

While cloud computing is experiencing global adoption at a rapid pace, the world is simultaneously facing a significant increase in the amount of data produced by both people and devices, thus increasing the demand for scalability and data security. A systematic literature review by researchers at Delhi Technological University proposed that blockchain technologies provides significant inputs in regards to finding an ideal approach to scalable and secure cloud data storage and processing [7].

C. Why EMR's critically need improved security

This section will attempt to provide insights into the major motivating factors for building better security solutions for EMRs. Both regulatory incentives and mandates, and the increasing risk of industry data breaches.

1) GDPRs effect on the health care industry: General Data Protection Regulation (GDPR) which was implemented 25th of May 2018, aimed to enhance individual's control and rights of personal data. It applies to any individual within the EU, but extends to apply to all service providers, including government bodies which deals with data from an individual within the EU. The GDPR makes specific references to the storage, processing and sharing of medical and patient data specifically, stating that a data subject has the legal right to apply for access to health information about the data subject. The GDPR also states that any entity which stores sensitive data, which the GDPR specifically deems EMRs as, must take every reasonable step in protecting the sensitive data from breaches. Failure to comply, may impose businesses fines up to 20 million euros, or up to 4% of the business worldwide turnover [8].

Researchers at the Tokyo Institute of Technology discussed in a study of Blockchain Based Patient Consent Models for EMRs that blockchain immutability and append-only functionality provides a fundamental challenge of incompatibility with the GDPR. While the GDPR mandates that patients or data subjects have a right to request their personal data be erased, which naturally becomes an issue due if their data was to be stored directly on-chain, due to the inability to erase data from the blockchain. To get around this issue, the researchers proposed a hybrid solution, where the EMR data itself is not stored on a blockchain, but is rather encrypted and stored in a traditional non-blockchain database. In their model, each transaction combines a randomised number with the patient's unique ID number into a unique hash, in order to pseudonymize the patient data. That way, the EMR is stored off-chain, and each EMR's transaction log, or log of changed, is stored on the blockchain, to maintain data integrity. This also allows for EMRs to be erased, enabling GDPR compliance. [9]

2) Threat landscape: In an article on healthcare data breaches published in May 2020, researchers analysed the causes and consequences of data breaches from 2005-2019 in healthcare from numerous globally reputable sources including the HIPAA Journal, PRC Database, The Office for Civil Rights Department of Health and Human Services Report and IMB sponsored Ponemon Institute Reports. The Analysis provides detailed insights into the severity of the risks that EMR systems and thus patients private data face on a global scale. The findings showed that more than 10 billion confidential records were exposed across industries from 2005 to 2009, whereas 43,38% or 3912 of these were from the healthcare sector alone. The study also showed that malicious intentional hacks were the main cause of healthcare data breaches, accounting for 64% of breaches from 2005-2019, and that the portion of healthcare breaches attributed to hacks increased to 92% when only looking at breaches from 2015-2019, suggesting that hacking as a data breach threat is increasing dramatically. The number of individuals globally affected by these data breaches are upwards of 255 million, according to the study[10].

"Despite the numerous advantages of EHRs, the digital

health data of patients is at huge risk today. As chronicled in our study, data breach trends have also undergone a massive transformation. The comprehensive analysis undertaken in this study reveals that the healthcare industry is the focus of many cyber invaders" [10].

3) The Value Proposition of improving security: In addition to the potential global savings of mitigating costly data security breaches discussed in the previous section, EMR implementation in health care is thought to improve the quality of health care by increasing effectiveness of care and treatment. The cost savings of properly implemented EMR systems are estimated $81 and $162 billion USD annually [11].

D. Blockchain technology

The interest in Blockchain can be explained by the technology's inherent ability to provide security, anonymity and data integrity without the need for a trusted third party intermediary to control transactions, and this creates promising research areas, particularly in terms of technical challenges and limitations [12].

"Many sectors, like finance, medicine, manufacturing, and education, use blockchain applications to profit from the unique bundle of characteristics of this technology. Blockchain technology (BT) promises benefits in trustability, collaboration, organization, identification, credibility, and transparency"[13].

Because of it's immutable append-only function, and a public record of transactions, blockchain technology can provide users with complete transparency over each step a system makes. The result is the creation of an environment which eliminates the need for a trusted authority, as malicious behaviour on the system is technically too difficult to accomplish [13].

.

1) Distributed Ledger Technology: Distributed ledger technology is an umbrella term used to describe the mechanism of a distributed ledger, produced, stored and audited by independent systems of users or "nodes." The nodes record, share and synchronise transactions by each of them running the same consensus algorithm. The consensus algorithms ensure that the network of nodes all agree on what data to write into the next block in the blockchain. For a transaction to be agreed upon as a valid transaction, and thus applicable to be added to the next block, more than half of the network needs to agree that the transaction is valid. This is what is called "reaching consensus". The most common types of consensus algorithms are called proof of work and proof of stake [13]–[15].

Both proof of work and proof of stake algorithms are mechanisms which allocate voting power in the consensus protocol. The weight of a user's (or node's) vote is proportional to what quantity of economic resources the node bring to the network. In the case of proof of work, economic resources are computer power or more specifically hash power, and the amount of hash power is proven by running the hashing algorithms. In the case of proof of stake, economic resources are proven by providing digital proof that the user or node has access to a given amount of crypto currency, such as Ethereum, by running the proof of stake algorithm [13]–[15].

This crucial concept solves the problem of a network attack, because in order to attack the network by a 51% attack, an attacker would need to prove to bring more proof of economic resources provided to the network than the rest of the network, either by having more than 50% of hashpower in the case of proof of work, or by proving ownership of more than 50% of total staked Ethereum, in the case of Ethereum 2.0's proof of stake [13]–[15].

2) Bitcoin: The first blockchain network to gain massive publicity and adoption was the Bitcoin network. It's mysterious creator by the pseudonym Satoshi Nakamoto, published the Bitcoin: A Peer-to-Peer Electronic Cash System whitepaper in 2009. Whomever Nakamoto is, (s)he envisioned an impenetrable trustless cash-system built on a decentralised open blockchain, open to everyone with an internet connection. Bitcoin runs on a Proof-of-Work algorithm, meaning the nodes of the network, referred to as "miners", spend computing power in order to secure the network and validate transactions [14].

The bitcoin network today is able to process 3,1 transactions per second (TPS) at the time of writing, and this value has ranged from 2-4 since 2017 [16].

3) Ethereum and Smart Contracts: The second largest blockchain network to date, Ethereum, was first proposed by its founder, Vitalik Buterin in 2013, by his publication of the Ethereum White Paper. Ethereum is fundamentally different from Bitcoin, although they both share many similarities are are both open source blockchain technologies featuring distributed ledgers, digital currency capabilities and immutability. Ethereum differs from Bitcoin by it's Smart Contract capabilities, allowing anyone user to deploy immutable and executable applications or programmes on it, making Ethereum a platform for decentralised applications. Additionally, Ethereum aims to switch it's consensus algorithm from proof of work to proof of stake [15], [17].

Smart contracts are the key functionality which enables decentralised applications (DAPPS) to make use of blockchain functionality. They allow businesses to set up automated contracts on blockchain, removing the need for a third party validator for contract settlement [13], [15], [17].

4) Permissioned or Permissionless blockchains: The are different types of blockchains, and in this section will briefly explain the two most common types: Permissioned and permissionless.

An important characteristic which can determine if a blockchain is appropriate for a particular use case or application, is whether the blockchain is permissioned or persmissionless. Permissionless blockchains such as Bitcoin and Ethereum allows anyone access to both read the blockchain's contents and and write to it. This access does not require identification, and write access is given to any node which can submit a valid transaction to the network, for a transaction fee. Permissioned blockchains, on the other hand are closed blockchains, which

have stricter access control. A permissioned blockchain may limit read and write access to only a set of authorized nodes. The utility of each type of blockchains are as such, different [18].

The permissions of a blockchain play a huge role in the blockchain network's decentralisation capability. Typically, permissionless blockchains have much greater degrees of decentralisation, due to their open access. On the contrary, permissioned blockchains such as Hyperledger are often more centralised, but often offer superior throughput and confidentiality [18].

5) Hyperledger Fabric: "The Hyperledger Project (www.hyperledger.org) is a collaborative effort to create an enterprise-grade, open-source distributed ledger framework and code base. It aims to advance blockchain technology by identifying and realizing a cross-industry open standard platform for distributed ledgers, which can transform the way business transactions are conducted globally"[19].

Like it's permissionless counterparts, Bitcoin and Ethereum, Hyperledger Fabric is a distributed ledger protocol run by decentralised nodes. The Hyperledger protocol distinguishes between validating nodes and non-validating nodes, where validating nodes are responsible for the consensus algorithm, checking transactions and maintaining the distributed ledger. The non-validating nodes functions as proxies which connects clients who are issuing transactions to a validating nodes [19].

Hyperledger's infrastructural and customizable framework could play a prominent role in the future by allowing the creation of a variety of new applications [13].

6) Immutability and integrity: Blockchains achieve data Integrity, meaning the data is accurate, complete, and unchanged from it's original state, by fundamentally only allowing certain actions to take place on the blockchain. Unlike a traditional database or ledger, where data can be added, altered or changed by authorized users, a blockchain only allows data to be added. The main reason for this is that any new block added to a blockchain needs to contain the hash of the previous block. Thus, there are no conditions regarded as valid in the consensus protocols where data already stored on the blockchain can be changed or removed [20].

7) Blockchain and file storage: The future presents numerous promising opportunities in regards to blockchain technology's potential to innovate and disrupt the cloud storage arena. Blockchain-based cloud storage solutions enable users to exclusively access and control their own data, without needing to trust the third party to secure data integrity, while maintaining the availability and of access and up time of cloud storage [7].

E. Blockchain technology's proposed use case within EMR storage and access control

Researchers and authors of "Secure Electronic Medical Records Storage and Sharing Using Blockchain Technology" Usman and Qamar [21] suggests that a blockchain-based EMR system is the key solution for dealing with the sharing of EMRs, explaining that the untamperable nature of blockchains is a good fit for maintaining the integrity of such sensitive data.

This assessment is backed up by Westphal and Seitz [22], saying: "Blockchain solutions offer efficient approaches for trustworthy data management, especially in the medical field when storing and processing sensitive patient data".

Furthermore, Usman and Qamar [21] explains that a merge of traditional cloud storage and blockchain-based access management are already being developed and implemented.

Complete and accurate patient data are valuable assets for both patients and doctors Chenthara, Ahmed, Wang, *et al.* [4]. Accessible, safe storage which protects the privacy of medical data are as such important issues.

Usman and Qamar [21] have been successful in developing a blockchain-based EMR management system built on Hyperledger, which protects medical data and allows secure distribution. Their system allows for the individual patients to be in control of managing which doctors are given read and write access to their blockchain-EMRs, giving patients not only access but full control and of their medical data.

F. Summary

Malicious hacks seem to be the most frequent type of attack resulting in health care data breaches, and EMR's are the attacks primary target.

Blockchain technology's unique characteristics seem to have immense potential in disrupting and innovating how EMR's are secured, and may be the key for improving EMR resiliency.

Various iterations of blockchain-based EMR systems have already been developed and tested, with promising results of improving both confidentiality and integrity of EMRs.

The merge of blockchain technology and EMR systems face serious challenges in terms of regulatory compliance, especially in regards to immutability and data subject's right to have their data erased.

For the purpose of this project, which is to deploy a working demonstration of a blockchain-based EMR management system, and test it's capabilities of integrity security mechanisms, the various blockchain types have been assessed, and the Hyperledger Fabric permissioned blockchain seems to be the best platform on which to build such a solution - likely as a hybrid of blockchain-based access control and integrity auditor, together with traditional cloud storage database enabling data storage of the EMRs themselves.

III. Data Gathering/Analysis

A. Security architecture comparison

In order to assess whether or not data integrity has been secured or improved with blockchain-based emr solutions there needs to be a comparison with traditional non-blockchain based emr solutions. For the purpose of this project, at least two and preferably three categories of emr management systems should be compared. One category being the traditional centralised and possibly cloud based but non-blockchain based solution. The second category or example will be the hyperledger blockchain based-emr management system, and

if possible, a third, ethereum based solution will also be included.

The different solutions need to be assessed both theoretically and technically in order to build tables describing their security features. Each category will naturally have many different possible solutions within them, and for this project one specific project will be chosen to represent each category.

An example of such a table illustrating the differences is provided. The table is meant as an example at this stage, and is likely to change in order to more accurately portray future findings.

TABLE I
SECURITY ARCHITECTURE COMPARISON

Security features	**EMR solutions**		
	Centralised	***Hyperledger***	***Ethereum***
Immutability	-	x	x
Single point of failure	x	-	-
Decentralised data storage	-	x	x
Patient data on-chain	-	x	-
Blockchain type	-	consortium	permissionless
Etc	-	-	-
Etc	-	-	-
Etc	-	-	-

[a]Proposed comparison table. "-" represents no while "x" represents yes.

B. Deployment

Deployment of a working existing blockchain based EMR management system has proven to be challenging. During the past months, more than twelve different deployments have been attempted, largely unsuccessfully. The common issues that persist through iterations are that of outdated software dependencies. There are many attempts to create these systems, however most of them require substantial changes and updates in order for them to function with recent updates to the platforms. With the blockchain space growing at such a rapid pace, being in its early stages of maturity, it seems decentralised applications need constant maintenance in order to still function. Surely, there are decentralised applications that do get these updates and that are maintained, but for this project a free to use open source application is needed, both because of cost and the ability to see what is happening under the hood.

The option to create a blockchain based EMR management system from scratch could seem like an obvious solution, however when attempted this quickly proved to be far out of scope for this project, as it would require immense levels of experience in the chosen platform's given programming language, as well as extensive experience in front- and back end development. Most if not all the open source iterations that have been attempted deployed have been created as collaborative masters degree projects, and even these are in truth imperfect and mostly proofs of concept.

In conclusion, the best option seems to be to find a working already developed open source project and use that for testing.

One such project, developed by students at Frankfurt University of Applied Sciences, named "Blockchain Solution to Healthcare Record System using Hyperledger Fabric" has shown great promise Kshitij Yelpale and Kamath [23]. The deployment of this project has not been without challenges, with many of the same issues discussed earlier persisting. However after contacting one of the developers Kshitij Yelpale, significant progress has been made in deployment, although not fully to completion as of yet. However, this Hyperledger fabric based iteration seems possible to deploy as of writing, and will be the main focus of this project going forward.

The Blockchain Solution to Healthcare Record System using Hyperledger Fabric uses Hyperledger fabric's consortium blockchain test net to allow many organisations (or hospitals in this scenario) to collaborate in hosting the blockchain. Each organisation is able to manage the network, add doctors and patients and privileges. This decentralisation, although small in comparison to larger public decentralised blockchains, do provide the system with the increased security of not being centralised, in the event of attacks that would otherwise compromise a system which had a single point of failure. If one hospital was to go down as a result of a natural disaster, power outage or cyber attack, the consortium network would still function, and no data would be lost.

Patients using this Hyperledger based EMR management system could view all records in their own EMR, from start to beginning, providing full transparency of the patients transactions and doctor interaction, which is exactly what provides these systems with increased integrity and security. The ability to know who wrote what data at what time into which patient's records, combined with the immutability of transactions, lets users know that their data has not been tampered with, deleted or changed, and in the event that incorrect data has been appended to their records, the details around this transactions would be available and could then be addressed.

IV. RESULTS

A. Initial results

At the current stage of the project, final results are limited to mostly based on research. The research conducted through literature review of related work indicates that it should be possible to technically demonstrate and describe how blockchain-based EMR management systems improve the data-integrity of patient records. Once the technical issues and challenges related to deployment of the blockchain-based solutions have been sorted out, the results of the comparison can be worked on more efficiently.

V. CONCLUSION AND FUTURE WORK

What conclusions you can draw from your findings? how do these relate you your research objectives? How could other researchers extend your work in the future.

Based on the research conducted thus far, it seems likely that blockchain technology has the potential for mass scale disruption within the health sector. If user friendly solutions of implementing blockchain in EMR management systems are

created and sufficiently tested, this could significantly impact the security and integrity of confidential patient data in a positive way.

This project will continue forwards trying to deploy and compare these iterations, technically describe which functions of the software protects data integrity and compare it to traditional systems to assess whether or not there is an improvement.

ACKNOWLEDGMENT

Special thanks to Kshitij Yelpale from Frankfurt University of Applied Sciences, Germany for assisting in troubleshooting the deployment of the "blockchain-hyperledger-fabric-electronic-patient-records" software which he created and for allowing this project to make use of the demo for research purposes.

REFERENCES

[1] Centers for Medicare & Medicaid Services, *The Health Insurance Portability and Accountability Act of 1996 (HIPAA)*, Online at http://www.cms.hhs.gov/hipaa/, 1996.

[2] P. Cappers and R. Scheer, "American recovery and reinvestment act of 2009: Final report on customer acceptance, retention, and response to time-based rates from consumer behavior studies," Tech. Rep., Mar. 2009. DOI: 10.2172/1424221. [Online]. Available: https://doi.org/10.2172/1424221.

[3] I. Keshta and A. Odeh, "Security and privacy of electronic health records: Concerns and challenges," vol. 22, no. 2, pp. 177–183, Jul. 2021. DOI: 10.1016/j.eij.2020.07.003. [Online]. Available: https://doi.org/10.1016/j.eij.2020.07.003.

[4] S. Chenthara, K. Ahmed, H. Wang, and F. Whittaker, "Security and privacy-preserving challenges of e-health solutions in cloud computing," *IEEE Access*, vol. 7, pp. 74 361–74 382, 2019. DOI: 10.1109/ACCESS.2019.2919982.

[5] F. Daryabar, A. Dehghantanha, and K.-K. R. Choo, "Cloud storage forensics: Mega as a case study," *Australian Journal of Forensic Sciences*, vol. 49, no. 3, pp. 344–357, 2017. DOI: 10.1080/00450618.2016.1153714. eprint: https://doi.org/10.1080/00450618.2016.1153714. [Online]. Available: https://doi.org/10.1080/00450618.2016.1153714.

[6] P. Mell, T. Grance, *et al.*, "The nist definition of cloud computing," 2011.

[7] P. Sharma, R. Jindal, and M. D. Borah, "Blockchain technology for cloud storage: A systematic literature review," *ACM Comput. Surv.*, vol. 53, no. 4, Aug. 2020, ISSN: 0360-0300. DOI: 10.1145/3403954. [Online]. Available: https://doi.org/10.1145/3403954.

[8] *General data protection regulation*, 2016. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%5C%3A32021R0876.

[9] D. Tith, J.-S. Lee, H. Suzuki, *et al.*, "Patient consent management by a purpose-based consent model for electronic health record based on blockchain technology," vol. 26, no. 4, pp. 265–273, Oct. 2020. DOI: 10.4258/hir.2020.26.4.265. [Online]. Available: https://doi.org/10.4258%5C%2Fhir.2020.26.4.265.

[10] A. H. Seh, M. Zarour, M. Alenezi, *et al.*, "Healthcare data breaches: Insights and implications," vol. 8, no. 2, p. 133, May 2020. DOI: 10.3390/healthcare8020133. [Online]. Available: https://doi.org/10.3390/healthcare8020133.

[11] W. MacKinnon and M. Wasserman, "Implementing electronic medical record systems," *IT professional*, vol. 11, no. 6, pp. 50–53, 2009.

[12] J. Yli-Huumo, D. Ko, S. Choi, S. Park, and K. Smolander, "Where is current research on blockchain technology?—a systematic review," *PLOS ONE*, vol. 11, no. 10, pp. 1–27, Oct. 2016. DOI: 10.1371/journal.pone.0163477. [Online]. Available: https://doi.org/10.1371/journal.pone.0163477.

[13] S. Leible, S. Schlager, M. Schubotz, and B. Gipp, "A review on blockchain technology and blockchain projects fostering open science," *Frontiers in Blockchain*, vol. 2, p. 16, 2019, ISSN: 2624-7852. DOI: 10.3389/fbloc.2019.00016. [Online]. Available: https://www.frontiersin.org/article/10.3389/fbloc.2019.00016.

[14] S. Nakamoto, *Bitcoin: A peer-to-peer electronic cash system*, 2009. [Online]. Available: http://www.bitcoin.org/bitcoin.pdf.

[15] V. Buterin, "Ethereum white paper: A next generation smart contract & decentralized application platform," 2013. [Online]. Available: https://github.com/ethereum/wiki/wiki/White-Paper.

[16] Blockchain.com, *Transactions-per-second*, 2021. [Online]. Available: https://blockchain.com/charts/transactions-per-second.

[17] M. Xu, X. Chen, and G. Kou, "A systematic review of blockchain," vol. 5, no. 1, Jul. 2019. DOI: 10.1186/s40854-019-0147-z. [Online]. Available: https://doi.org/10.1186/s40854-019-0147-z.

[18] K. Wüst and A. Gervais, "Do you need a blockchain?" In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, IEEE, 2018, pp. 45–54.

[19] C. Cachin, "Architecture of the hyperledger blockchain fabric," 2016.

[20] F. Hofmann, S. Wurster, E. Ron, and M. Böhmecke-Schwafert, "The immutability concept of blockchains and benefits of early standardization," in *2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K)*, IEEE, 2017, pp. 1–8.

[21] M. Usman and U. Qamar, "Secure electronic medical records storage and sharing using blockchain technology," *Procedia Computer Science*, vol. 174, pp. 321–327, 2020, 2019 International Conference on Identification, Information and Knowledge in the Internet of Things, ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2020.06.093. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920316136.

[22] E. Westphal and H. Seitz, "Digital and decentralized management of patient data in healthcare using blockchain implementations," *Frontiers in Blockchain*, vol. 4, p. 36, 2021, ISSN: 2624-7852. DOI: 10.3389/fbloc.2021.732112. [Online]. Available: https://www.frontiersin.org/article/10.3389/fbloc.2021.732112.

[23] J. S. Kshitij Yelpale and V. V. Kamath, "Blockchain solution to healthcare record system using hyperledger fabric," 2020. [Online]. Available: https://github.com/kshitijyelpale/blockchain-hyperledger-fabric-electronic-patient-records/blob/main/docs/Final%5C%20report.pdf.