



Programa de Pós-Graduação em

Computação Aplicada

Mestrado/Doutorado Acadêmico

Fausto Neri da Silva Vanin

MEPCA: a technical model to improve on-chain Electronic Health
Records processing

São Leopoldo, 2024

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL DOUTORADO

FAUSTO NERI DA SILVA VANIN

MEPCA: A TECHNICAL MODEL TO IMPROVE ON-CHAIN ELECTRONIC HEALTH
RECORDS PROCESSING

SÃO LEOPOLDO
2024

Fausto Neri da Silva Vanin

MEPCA: A TECHNICAL MODEL TO IMPROVE ON-CHAIN ELECTRONIC HEALTH
RECORDS PROCESSING

Thesis presented as a partial requirement to
obtain the Doctor's degree by the Applied
Computing Graduate Program of the
Universidade do Vale do Rio dos Sinos

Advisor:
Prof. Dr. Rodrigo R. Righi

Co-advisor:
Prof. Dr. Cristiano A. Costa

São Leopoldo
2024

V258m Vanin, Fausto Neri da Silva.
MEPCA : a technical model to improve on-chain
electronic health records processing / Fausto Neri da Silva
Vanin. – 2024.
77 f. : il. ; 30 cm.

Tese (doutorado) – Universidade do Vale do Rio dos
Sinos, Programa de Pós-Graduação em Computação
Aplicada, 2024.

“Orientador: Prof. Dr. Rodrigo R. Righi
Coordenador: Prof. Dr. Cristiano A. Costa”

1. Blockchain. 2. Distributed Hash Tables. 3. Distributed
Network. 4. Electronic Health Records. 4. Homomorphic
Encryption. 5. Personal Health Records. I. Título.

CDU 004.4

*O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de
Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001*

*This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de
Nível Superior - Brasil (CAPES) - Finance Code 001*

ATA DE BANCA EXAMINADORA DE TESE DE DOUTORADO Nº 11/2024

Aluna: Fausto Neri da Silva Vanin

Título da Tese: “MEPCA: A TECHNICAL MODEL TO IMPROVE ON-CHAIN ELECTRONIC HEALTH RECORDS PROCESSING”

Banca: Prof. Dr. Rodrigo da Rosa Righi (Orientador) – Unisinos
Prof. Dr. Cristiano Andre da Costa (Coorientador) – Unisinos
Prof. Dr. Rodolfo Stoffel Antunes (Avaliador) – Unisinos
Prof. Dr. Luciano Gaspary (Avaliador) – UFRGS
Prof. Dr. Daeyoung Kim (Avaliador) – KAIST

Aos vinte e três dias do mês de setembro do ano de 2024, às 20h30 reuniu-se a Comissão Examinadora de Defesa de Tese composta pelos professores: Prof. Dr. Rodrigo da Rosa Righi (Orientador) – Unisinos (por webconferência); Prof. Dr. Cristiano Andre da Costa (Coorientador) – Unisinos (por webconferência); Prof. Dr. Rodolfo Stoffel Antunes (Unisinos) (por webconferência); Prof. Dr. Luciano Gaspary (UFRGS) (por webconferência) e Prof. Dr. Daeyoung Kim (KAIST) (por webconferência) para analisar e avaliar a Tese apresentada pelo(a) aluno(a) Fausto Neri da Silva Vanin (por webconferência).

Considerações da Banca:

After the Fausto's presentation, the jury have decided by the Approve evaluation. Prof. Rodrigo took notes about all the referees' comments, so the idea is to incorporate all of them in the final version of the Ph.D. document in order to improve its quality.

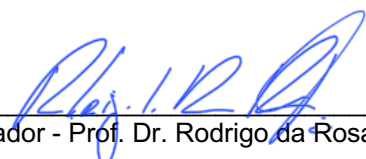
Ocorreu alteração do título? (☒) Não (☐) Sim
Indicar o novo título:

A Banca Examinadora, em cumprimento ao requisito exigido para a obtenção do Título de Doutor em Computação Aplicada, julga esta tese:

(☒) APROVADA (☐) REPROVADA

Conforme Artigo 75 do Regimento do Programa o texto definitivo, com aprovação do Orientador, deverá ser entregue no prazo máximo de sessenta (60) dias após a defesa. O resultado da banca é de consenso entre os avaliadores. A emissão do Diploma está condicionada a entrega da versão final da Tese.

São Leopoldo, 23 de setembro de 2024.



Orientador - Prof. Dr. Rodrigo da Rosa Righi

Aos ancestrais, originários e amadinhos.

*Até que os leões contem suas próprias histórias
todos os contos sobre a caça irão glorificar os caçadores.*
— PROVÉRBIO AFRICANO

ABSTRACT

The integration of blockchain technology within the healthcare industry has garnered significant attention due to its potential to address critical challenges such as data privacy, interoperability, and the integrity of health records. Although electronic health record (EHR) standards such as HL7 FHIR and OpenEHR have established frameworks for data consistency and system interoperability, concerns remain about the privacy and security of sensitive patient information, particularly in light of regulations such as the Health Insurance Portability and Accountability Act (HIPAA), the General Data Protection Regulation (GDPR), and the Lei Geral de Proteção de Dados (LGPD). Most related work stores only data hash on blockchain nodes, making data validation impossible from a blockchain perspective, which raises the risks of invalid or malicious data being provided. This work introduces the MEPCA model, a novel framework grounded in five core principles that explore the application of blockchain and cryptographic technologies in the management of health records, focusing on maximizing the use of on-chain resources for the processing of EHR data. Our main contribution is to provide guidance and techniques to maximize the adoption of decentralized solutions in the healthcare industry, with practical use cases and technical analysis. Our model introduces novel elements for secure data sharing, called Data Steward and Shared Data Vault, and proposes an innovative method that generates Zero-Knowledge Proofs of HL7 FHIR required fields for hash digests. We run technical experiments with Fully Homomorphic Encryption (FHE) algorithms to evaluate on-chain data analysis using a dataset with 1.3 million records and evaluates on-chain data processing and storage with a 10 thousand HL7 FHIR dataset with plain and hash representation. Our findings suggest that maximizing on-chain processing can improve the security and reliability of health records, offering a robust alternative to traditional off-chain data processing approaches. The adoption of the MEPCA model can bring an evolution to the healthcare industry, allowing society and institutions to have a more secure and efficient digital infrastructure for EHR.

Keywords: Blockchain. Electronic Health Records. Personal Health Records. Homomorphic Encryption. Distributed Hash Tables. Distributed Network.

LIST OF FIGURES

1	Solution contribution: in the traditional model, the blockchain layer has restricted contribution to the solution, while in the MEPCA model, we promote enhanced on-chain data processing and end-to-end encryption, in a scenario where the blockchain layer is able to validate incoming data, even when it is a hash digest.	19
2	EHR and PHR Overview: an evolution from a institution-centered data management to a patient-centered approach	22
3	Blockchain strategies and on-chain usage comparison. First, the traditional approach, where only a hash representation is processed on-chain, with no data validation. Second, MEPCA recommended approaches: a) a minimum reliable model has on-chain data validity check to ensure incoming data complies with standards; b) a hybrid approach is an alternative where part of the data is stored into blockchain nodes and another part in traditional infrastructure; c) the recommended approach for most use cases has data analysis and proof generation on-chain, even on encrypted data.	40
4	Solution Data Flow: Patient authorizes a Data Steward (a) to keep an encrypted version of PHR off-chain (b) on an IPFS network (c); A Health Institution requires a patient for PHR (d) and, whenever authorized, PHR metadata can be shared on-chain (e) in a blockchain network (f).	45
5	Architecture Components: IPFS Network for off-chain data, Blockchain Network for on-chain metadata, Data Steward to keep encrypted version of PHR and Shared Data Vault to support temporary data sharing encrypted with requester public key.	46
6	Solution Sequence Diagram with three primary use cases represented: 1) Patient stores encrypted health records in Data Steward nodes (IPFS network); 2) Patient stores encrypted metadata in Blockchain Nodes; 3) Patient authorizes health institution and consortium to access a portion of data in a Shared Data Vault for a predefined period.	48
7	On-chain hash proof algorithm for HL7 FHIR data: a Merkle Tree based system for Zero Knowledge Proof generation. The system is able to produce verify if a given hash digest contains required fields in the input. Required fields map deterministically to specific leaves in a Merkle Tree, making it possible to use the complementary values from a Merkle Proof to act as proof to a verifier.	51
8	Experiments network architecture using Hyperledger Fabric 2.5 and Caliper 2.0	57
9	Transactions per second for read and write. Fixed Rate delivered better results for write, while Fixed Load delivered the best overall performance. Maximum Rate reached worst result of the tree strategies. Fixed Rate resulted in failing transactions due to system overload.	59
10	Hash processing in milliseconds compared to (WANG et al., 2021) and MEPCA. As the number of required fields increases, the processing time reduces. The baseline hash calculation for SHA-256 took 0.00993ms to calculate, and the reference value from (WANG et al., 2021) took 0.68ms.	60

11	Hash storage compared to (WANG et al., 2021) and MEPCA. Proof calculation takes $364B$ for one required field and decreases as the number of required fields increases, as the proof size requires less data for validation. It reaches better results compared to (WANG et al., 2021) for higher number of required fields.	61
12	Blockchain Network Simulation: 10k pool size with 500B of block header and 5 seconds block time. Scenarios compare different block sizes (535KB, 1070KB, and 2140KB) in three formats (raw data, encrypted cyphertext, and compressed encrypted cyphertext) with 10, 30, and 100 nodes distributed in 3 different regions. The chart summarizes the time in seconds for 10k registries to propagate in the network.	62

LIST OF TABLES

1	Related work, organized by use case, focus and support to health data standards.	32
2	Gaps and Opportunities: a summary of gaps found in related work and that raise opportunities for contribution in the proposed model	37
3	Key requisites for MEPCA model implementation	41
4	MEPCA components and use cases	42
5	Applied Example	49
6	Evaluation aspects for the proposed model	55
7	Evaluation results for data write: Fixed rate reached best results, but with higher number of pending transactions and some failing transactions. Fixed Load reached 64.3 TPS with 12 pending transactions, while Maximum Rate generated no pending transactions, while delivering 51.6 TPS.	58
8	Performance evaluation: FHE calculation on different dataset sizes comparing the raw data to encrypted data (in seconds). The second and third columns show the overall calculation time for addition in seconds comparing raw data and ciphertext respectively.	60
9	Algorithm profiling: parameters n and q for Polynomial Modulus Degree and Coefficient Modulus respectively in BFV encryption. The steps of encryption, addition and decryption are in the following rows.	60
10	Storage consumption: as the (n, q) pair increases, the ciphertext string size in KB also increases. Each column S has the average size of each registry in plain, encrypted and compressed format, while each column T has the total storage amount for a set of 100k registries	61
11	Security and Privacy scenarios: comparison	63

LIST OF ACRONYMS

ABE	Attribute-Based Encryption
DHT	Distributed Hash Table
DS	Data Steward
EHR	Electronic Health Record
FHE	Fully Homomorphic Encryption
FHIR	Fast Healthcare Interoperability Resources
GDPR	General Data Protection Regulation
HE	Homomorphic Encryption
HIPAA	Health Insurance Portability and Accountability Act
HL7	Health Level Seven International
IoHT	Internet of Health Things
IPFS	Interplanetary File Systems
LGPD	Lei Geral de Proteção de Dados
PHE	Partially Homomorphic Encryption
PHR	Personal Health Record
SC	Smart Contract
SDV	Shared Data Vault
ZKP	Zero-Knowledge Proof

CONTENTS

1 INTRODUCTION	17
1.1 Research Problem and Objectives	18
1.2 Research Hypothesis	20
1.3 Text Organization	20
2 BACKGROUND	21
2.1 Electronic Health Records	21
2.2 Healthcare Data Standards	22
2.3 Distributed Networks and Cryptography Components	25
2.3.1 Blockchain Networks	25
2.3.2 Network architecture and configuration	26
2.3.3 Distributed Hash Table and IPFS	27
2.3.4 Asymmetric Cryptography and Digital Signatures	27
2.3.5 Hash functions and Hash Structures	28
2.3.6 Zero-Knowledge Proofs	28
2.3.7 Proxy Re-encryption and Attribute-Based Encryption	29
2.3.8 Homomorphic Encryption	29
3 RELATED WORK	31
3.1 Selection Criteria	31
3.2 State-of-the-art	32
3.3 Open Gaps and Opportunities	36
4 MEPCA MODEL	39
4.1 Design Decisions and Use cases	39
4.1.1 Use Case 1: International Patient Summary (IPS)	42
4.1.2 Use Case 2: Digital Public Infrastructure (DPI-H)	43
4.1.3 Use Case 3: Internet of Health Things (IoHT)	43
4.2 End-to-end data protection for PHR	44
4.2.1 Data Steward	46
4.2.2 Shared Data Vault	47
4.3 On-chain hash proofs	47
5 RESULTS AND DISCUSSION	53
5.1 Methodology	53
5.1.1 On-chain data processing	55
5.1.2 Data analysis on encrypted data	57
5.2 Experiments and Results	58
5.2.1 Test Set 1: HL7 FHIR on-chain processing	58
5.2.2 Test Set 2: Fully Homomorphic Encryption	59
5.3 Discussion	64
6 CONCLUSION	67
6.1 Scientific Contribution	68
6.2 Publications	69
6.3 Limitations and Future Work	69
BIBLIOGRAPHY	71

1 INTRODUCTION

The adoption of blockchain solutions for healthcare recently gained significant attention in the scientific community and in the healthcare industry (NAMASUDRA, 2024), as the population and institutions increase the demand for efficient services, increased privacy protection, and a higher level of integration between the actors in the industry. Blockchain technologies provide many elements that contribute significantly to important topics, such as security and privacy (YANG et al., 2023), interoperability (SENTAUSA; HAREVA, 2023), and could be very helpful in scenarios such as the COVID-19 pandemics (NG et al., 2021), medical research, counterfeit prevention, and management of medical supply chains (TAHERDOOST, 2023). In this work, we focus on the use of blockchain and cryptography technologies for health records.

Electronic Health Record (EHR) refers to an electronic structure for patient health records, generally collected and stored by health institutions (ISO, 2021). Patient Health Record (PHR) refers to an electronic structure for patient health records, collected by devices such as IoT sensors (ARCHER et al., 2011; DA COSTA et al., 2018) and stored in a repository that supports sharing in different digital formats (ISO, 2021; ROEHRS; COSTA; ROSA RIGHI, 2017). The healthcare industry has evolved over the last few decades in defining standards for Electronic Health Records (EHR), such as HL7 FHIR (Health Level Seven International, 2019) and OpenEHR (OPENEHR, 2020). These standards provide benefits such as system interoperability, data consistency, efficiency, and cost savings (SETYAWAN et al., 2021). However, from a privacy and security perspective, they should also offer means to protect patient privacy and comply with data protection regulations (FINCK, 2019) such as the Health Insurance Portability and Accountability Act (HIPAA) (HIPAA, 1996), the General Data Protection Regulation (GDPR) (UNION, 2016), and the Lei Geral de Proteção de Dados (LGPD) (REPÚBLICA, 2018).

Although scalability is a challenge in blockchain solutions for EHR, considering the amount of data, it could not become feasible in terms of computational resources and cost (MISBHAUDDIN et al., 2020). Thus, most existing solutions adopt off-chain data processing, such as Cloud Service Providers (CSP) (REEGU et al., 2023; YAZDINEJAD et al., 2020), the Interplanetary Protocol File System (IPFS) (MADINE et al., 2020a; MISBHAUDDIN et al., 2020; VANIN et al., 2023) or Distributed Hash Table (DHT) (ROEHRS; COSTA; ROSA RIGHI, 2017), over the on-chain alternative, frequently sending only a hash representation of a given input data to the blockchain (TAHERDOOST, 2023). Such strategies face the risk of cyberattacks that could cause data leaks (CHEN et al., 2022), and the introduction of incorrect or malicious data into blockchain nodes if there are no means of on-chain validation of incoming data in terms of structure, cardinality, and value domains (Health Level Seven International, 2024).

In this work, we analyze many technical aspects in the use of blockchain technologies for the EHR lifecycle. We introduce the MEPCA model, a combination of five principles (Maximize, Encrypt, Prove, Comply, and Adapt) divided into multiple architectural building blocks from

blockchain and cryptography, aimed at enhancing the use of EHR on-chain processing. We propose a set of key requirements and provide technical background for adoption, detailed in three different use cases, and levels of adoption. The model presents an algorithm for on-chain hash validation of HL7 FHIR JSON data, based on Merkle Trees and Zero-Knowledge Proofs. We applied the MPECA model for PHR interoperability, proposing a hybrid strategy for data and metadata, using Hyperledger Fabric and IPFS network, end-to-end encryption, and Fully Homomorphic Encryption (FHE) techniques to support data analysis on encrypted data.

We evaluated the proposed model, analyzing technical aspects of on-chain storage of HL7 FHIR documents, with a data set of 10,000 registries, compared to the hash-based model, considering processing time for proof calculation, along with security aspects of on-chain validation of hash digests. We also evaluated the FHE calculation on a data set of 1.3M cases from the United States Center for Disease Control and Prevention (CDC). We provide analysis of algorithm performance, cryptography calculation, and security and privacy scenarios to identify impacts on each building block in the proposed model. Our main contribution is to provide guidance and techniques to maximize the adoption of decentralized solutions in the healthcare industry, with practical use cases and technical analysis.

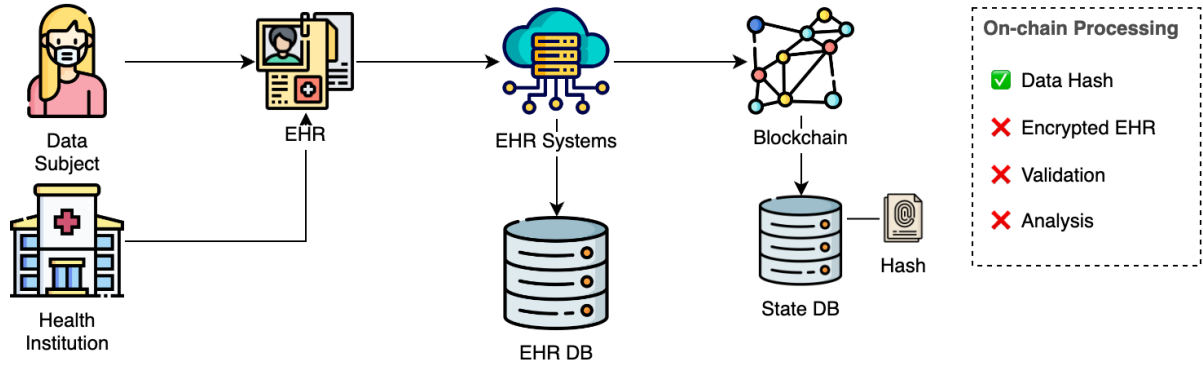
1.1 Research Problem and Objectives

The digital approach to health records has many applications in the health industry (CHUKWU; GARG, 2020; YAQOOB et al., 2022). When it comes to data sharing and interoperability, solutions must address issues such as scalability, privacy protection, and compliance with regulation (PERERA et al., 2020; SHUAIB et al., 2021; PAIK et al., 2019). Blockchain solutions for EHR recently gained significant attention in the scientific community (TAHERDOOST, 2023), although most approaches face the following problems:

- Most data processing off-chain: the use of the blockchain only to store data, especially hash digest, raises risks to data consistency, as whenever blockchain nodes do not run any business logic or validation, malicious or invalid information might be stored on-chain;
- Unencrypted data: whenever during transport or at rest, unencrypted data processing raises privacy issues, especially unauthorized access;
- Lack of control from Data Subjects: especially with PHR, allowing data subjects to manage access to data is key for a secure blockchain EHR system. Most models do not implement access management systems that include data subjects as responsible for data, or any delegation model to service providers.

Our work approaches the context of EHR with the given research problem: **How to improve smart healthcare for the exchange and interoperability of EHR, protecting privacy,**

Traditional Model



MEPCA Model

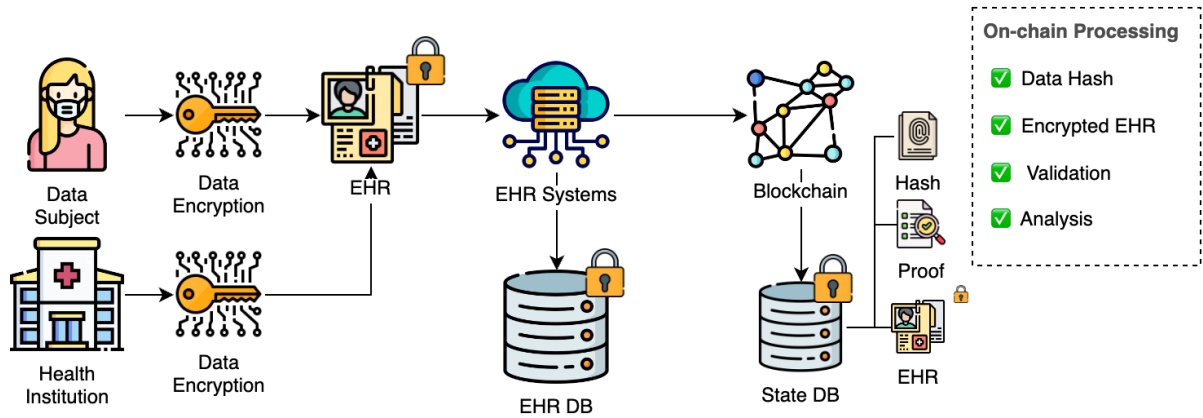


Figure 1: Solution contribution: in the traditional model, the blockchain layer has restricted contribution to the solution, while in the MEPCA model, we promote enhanced on-chain data processing and end-to-end encryption, in a scenario where the blockchain layer is able to validate incoming data, even when it is a hash digest.

meeting scalability and regulatory requirements? The problem relates to the EHR lifecycle, considering the challenges related to access policy, data security, and support for decision-making. To approach the research question, our main objective is to **create a blockchain-based model to maximize on-chain EHR data processing, promoting privacy protection and data interoperability**. To achieve this objective, we establish the following sub-objectives:

1. Propose a model to drive the adoption of on-chain strategies and support decision-making;
2. Design blockchain and cryptography strategies that support multiple use cases for EHR, with technical components;
3. Apply the model to PHR interoperability, promoting end-to-end encryption and data analysis on encrypted data;
4. Evaluate technical aspects regarding storage occupation, data synchronization, cryptography methods and data analysis.

1.2 Research Hypothesis

To approach the proposed research problem and deliver the objectives of this work, we establish some hypotheses that combine techniques for data protection, network interoperability, and data analysis. In the following, we describe each of the research hypotheses.

1. There is an opportunity to increase the adoption of on-chain strategies and cryptography compared to the existing literature;
2. A consistent set of design principles and a mapping between relevant use cases in health care and key building-blocks in blockchain and cryptography can drive decision-making and technology adoption;
3. The ability to add proofs to hash data for on-chain validation can improve auditability of data existing in the network;
4. End-to-end encryption techniques can support data analysis in an acceptable processing time when compared to raw data processing and reduce the demand for unencrypted data.

1.3 Text Organization

This document is organized as follows: we start with a background on the main research topics on Chapter 2. In Chapter 3 we analyze the most representative related work, on PHR data sharing and interoperability using distributed networks, describing gaps and opportunities in the existing literature. Chapter 4 introduces the proposed model, its components, and use case applications. In Chapter 5, we describe technical evaluation criteria for the proposed method and in Chapter 5.2 we present experiments and results, with two data sets and the hash proof algorithm. In Chapter 6 we analyze results and project future work and opportunities for research.

2 BACKGROUND

In this chapter, we cover the main building-blocks of the proposed model, including health records, blockchain components, and cryptography techniques. We start by covering Electronic Health Records (EHR), conceptualizing important elements, such as Personal Health Records (PHR), and data standards, including OpenEHR and HL7 FHIR, as these standards represent an important design aspect in the proposed model. Next, we describe multiple concepts in decentralized network and blockchain, covering Distributed Hash Tables, blockchain networks, and technical aspects related to network performance, scalability, and security, such as transaction finality and consensus algorithms. The chapter also describes multiple cryptography techniques that will be applied in the model for data protection, data analysis, and data validation, including Zero-Knowledge Proofs, Attribute-Based Encryption, and Hash data structures, such as Merkle Trees, which play an important role in the proposed model.

2.1 Electronic Health Records

Digitization is an important topic in the healthcare industry, as many processes still rely on paper and manual work. Such scenario opens opportunities for standards and methods in Information Technology to tackle issues regarding topics like regulation compliance, interoperability, data standards, system integration (sensors, IoT, applications) and data privacy.

Electronic Health Records (EHR) refer to an electronic representation of patient health records, collected and stored in a repository, that can be shared in different digital formats (CHUKWU; GARG, 2020; ROEHRS et al., 2017; MELLO et al., 2022). EHR can contain several data groups, such as allergies, vital signs, medical appointments, laboratory exams results, medical imaging and diagnoses. Usually, EHR origination, distribution, and storage, run by Health Institution, so patients don't have access to such records.

Patient Health Records (PHR) are an evolution of EHR (ISO, 2021). In a PHR context, patients have control over data and can grant - and revoke - access permission to third parties, such as Health Institutions (MADINE et al., 2020b). Data could be obtained by sensors, IoT, and mobile technologies, for example. PHR can receive data from healthcare providers, stored in a repository where the patient has access (ROEHRS et al., 2019; SHE et al., 2019). We illustrate the differences between EHR and PHR on Figure 2, having a convergence of data coming from the patient side and from the health institution side, generating a large set of records related to a specific patient.

As PHR involves multiple actors during the data lifecycle, it raises the need for semantic interoperability standards (MELLO et al., 2022), such as HL7/FHIR (SARIPALLE; RUNYAN; RUSSELL, 2019) and OpenPHR (OPENEHR, 2020). It becomes more complex to comply with existing regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), the General Data Protection Regulation (GDPR), and the Brazilian equivalent LGPD (SHUAIB

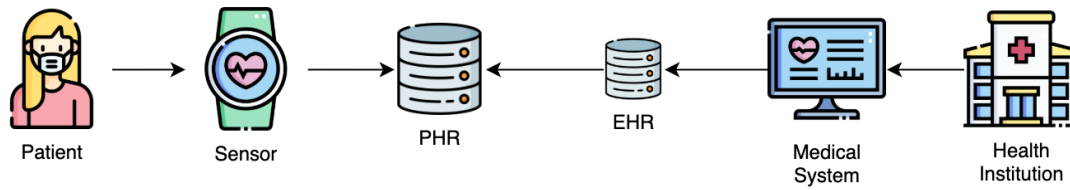


Figure 2: EHR and PHR Overview: an evolution from a institution-centered data management to a patient-centered approach

et al., 2021) as PHR can be transmitted in insecure connections, and there is a challenge to track which actor has access to a given portion of data.

2.2 Healthcare Data Standards

Health data standards are a set of established rules, conventions and specifications designed to ensure consistency and interoperability in the collection, storage, exchange and use of health-related information within the healthcare ecosystem. These standards play a pivotal role in creating a unified and structured framework for organizing diverse health data, enabling seamless communication between different health information systems and stakeholders. The purpose of health data standards is to establish a common language and format that facilitates the exchange of information between healthcare providers, institutions, and systems, ultimately improving the quality, safety, and efficiency of patient care.

At its core, health data standards encompass various dimensions, including data formats, terminologies, and communication protocols. The standardization of data formats ensures that health information is uniformly structured, making it easier to interpret and exchange across different platforms. The adoption of health data standards brings several benefits. Interoperability, a key advantage, allows disparate healthcare systems to seamlessly exchange information, reducing data silos, and improving care coordination. Standardized health data promotes the accuracy and integrity of data, minimizing errors and discrepancies in patient records.

HL7 FHIR (Fast Healthcare Interoperability Resources) is a modern and widely adopted healthcare interoperability standard developed by Health Level Seven International (HL7) (Health Level Seven International, 2019). FHIR represents a set of specifications that standardize the exchange of healthcare information in a more accessible and web-friendly format. FHIR leverages contemporary web standards such as RESTful APIs (Representational State Transfer), JSON (JavaScript Object Notation), and XML (eXtensible Markup Language) to facilitate seamless data exchange and interoperability across diverse healthcare systems.

One of the defining features of HL7 FHIR is its modular and resource-based approach. FHIR resources are discrete units of standardized clinical or administrative information that represent entities such as patients, medications, observations, etc. These resources are designed to be modular, making it easy to assemble and exchange only the necessary information for specific use cases. FHIR also supports a RESTful architecture, allowing for simple and efficient com-

munication over the web. This approach to standardization promotes flexibility, making FHIR adaptable to various healthcare scenarios and enabling agile development and implementation.

OpenEHR is an open standard framework that defines specifications for the construction of electronic health records (EHRs) and related systems (OPENEHR, 2020). Unlike traditional health information systems that are often constrained by rigid formats and proprietary structures, OpenEHR takes a fundamentally flexible and modular approach. At its core, OpenEHR employs a dual-model architecture, distinguishing between the information model for data storage and retrieval and the knowledge model for defining clinical content. This separation allows for the creation of standardized, yet adaptable, clinical archetypes and templates, providing a foundation for interoperable and customizable EHR systems.

One of the key strengths of OpenEHR lies in its use of archetypes and templates. Archetypes represent standardized clinical concepts, while templates combine multiple archetypes to define specific clinical documents or forms. This approach enables healthcare organizations to tailor their EHR systems to meet unique clinical requirements without sacrificing interoperability. OpenEHR's emphasis on semantic interoperability ensures that health information can be consistently interpreted and shared across different systems and settings, promoting a more connected and collaborative healthcare ecosystem.

OpenEHR is well-suited for data modeling and long-term data persistence, making it suitable for systems requiring extensive data analysis and complex querying. In contrast, HL7 FHIR prioritizes interoperability and real-time data exchange, offering a modular and flexible approach that supports rapid implementation and integration across various healthcare applications (KRYSZYN et al., 2023). As HL7 FHIR facilitates seamless data sharing and integration, we will adopt this standard for the experiments in our model.

Standardized health data should be meet the FAIR principles, with are Findability, Accessibility, Interoperability, and Reusability. This concept applies to data and metadata (GROUP, 2021). Metadata are known as "data about data", as they play a key role on making existing data in a health system easily to locate and process. In the context of our work, we consider both data and metadata as key elements to establish on-chain strategies in the proposed model. We provide an example of an HL7 FHIR file in JSON format in Listing 2.1.

Listing 2.1: Example HL7 FHIR file in JSON format

```
{
  "resourceType": "Patient",
  "id": "example",
  "text": {
    "status": "generated",
  },
  "identifier": [
    {
      "use": "usual",
      "system": "urn:oid:1.2.36.146.595.217.0.1",
      "value": "12345"
    }
  ],
  "active": true,
  "name": [
    {
      "use": "official",
      "family": "Doe",
      "given": ["John"]
    }
  ],
  "gender": "male",
  "birthDate": "1974-12-25",
  "address": [
    {
      "use": "home",
      "line": [
        "123 Main St"
      ],
      "city": "Anytown",
      "state": "CA",
      "postalCode": "12345"
    }
  ]
}
```

2.3 Distributed Networks and Cryptography Components

As a significant part of the population and companies operate over the Internet, it is now natural that different industries, including healthcare, make use of Wide-Area Networks (WAN) to gain scale and integrate with multiple partners and actors. The remarkable work of Satoshi Nakamoto in 2008 (NAKAMOTO, 2008) has established the basis for distributed tamper-proof blockchain networks, making it practically viable to address issues regarding privacy, auditability and tracking of transactions involving sensitive elements, from currency to assets, from retail products to health records. In the following chapters, we describe Blockchain Networks and Distributed Hash Networks and relevant concepts in the context of the proposed model.

2.3.1 Blockchain Networks

Blockchain technologies represent a unique design of ledger structure, distributed network, consensus protocol, and cryptographic mechanisms to promote transparency, data immutability, consistency, equal rights and data availability (PAIK et al., 2019). Data processing and sharing in a blockchain are triggered by transactions. Transactions are grouped into blocks and each block must reach specific consensus rules to be accepted by network peers (GEBREMEDHIN, 2018). Security in blockchain transactions is promoted by the use of cryptography algorithms and techniques such as Asymmetric Key Pairs e.g. RSA and Elliptic Curves, Hashing Algorithms, Hashing Hierarchy (e.g. Merkle and PATRICIA trees), Zero-Knowledge Proofs, and Homomorphic Encryption (ZHANG et al., 2018).

Networks are made up of peer nodes that exchange data in transactions. Transaction finality may vary according to the adopted Consensus Algorithm, which is the method used by the nodes to agree on each new accepted block of transactions. Blockchain networks are usually divided in two major types: public and permissioned. In public networks, any computing node can enter the network with no need for permission of any participant beside the need to comply with the protocol e.g. block size, block timing, agreed by existing members. Permissioned networks adopt governance protocols that depend outside the protocol itself. Thus, new nodes in the network need some kind of approval in order to operate in the network. The concept of Blockchain Consortium emerges in this context, as consortia represent well how a permissioned blockchain network might work.

Depending on the existing network protocol, richer forms of transactions might take place, like the concept of Smart Contracts (SC). SC are programmable idempotent pieces of software that run on top of network peers and allow specific-purpose solutions to generate transactions in the network. Such transactions might carry extended data fields that meet the specific needs of a SC interface¹ and consume storage area in each network peers, as they are stored in the

¹Smart Contract interfaces are equivalent to Application Programmable Interfaces (API) in terms that each interface defines a set of incoming parameters, returning data, and error messages. Blockchain network peers usually expose such interfaces via protocols like Remote Procedure Call (RPC).

Distributed Ledger, the synchronized transaction history shared among network peers. This characteristic raises a debate on which data should be stored in the ledger (on-chain) and which data should not, as the ledger file size could compromise network performance through time.

Most blockchain networks adopt an economic model to pay for transactions and prevent malicious behavior. Public blockchain protocols, by default, have a crypto currency as the underlying economic element. Such a currency works as a network incentive, as it may work for taxation and rewarding. In networks that support Smart Contracts, taxes may vary according to the demand for computational resources of each SC interface. Thus, it is of great importance to consider the economic model and its corresponding costs when designing a blockchain network or a blockchain-based solution.

2.3.2 Network architecture and configuration

Consensus algorithms and transaction finality are two highly coupled concepts in blockchain with a great influence on the processing of EHR data on-chain. Networks such as Bitcoin have probabilistic finality, since their consensus algorithm (proof-of-work) results in higher transaction finality probability as more blocks enter the ledger. Proof-of-stake algorithms, such as Ethereum and most compatible networks, have financial finality, as the cost of tampering with a transaction of older blocks increases drastically. Most private and permissioned networks adopt different types of consensus algorithms, such as Practical Byzantine Fault Tolerance (PBFT), which result in instant finality. Granular EHR data writing operations suit well with higher transaction finality, while EHR interoperability operations demand smaller or instant transaction finality, as data access might be for critical situations.

Each network protocol establishes a maximum size for data attached to transactions, as they influence the block size and consequently the number of transactions a block can hold. The Bitcoin network allows *80Bytes* arbitrary data attached to each transaction, while Ethereum supports *128KB*. Hyperledger Fabric has no formal limit for transactions, neither blocks, although such parameters can be defined in the network configuration. For EHRs with larger file sizes, on-chain storage strategies might require multiple transactions to complete a file transfer. We further explore this aspect in detail on Chapter 5.2.

Subnetworks allow specialized access control to blocks and transactions, higher performance, and lower costs. This technique is applied by using different approaches, such as Shardings and Rollups in Ethereum, Parachains in Polkadot, Channels and Private Data in Hyperledger Fabric. As a general rule, each network layer $L2$ has its own node infrastructure and ledger, and according to a set of rules, provides proofs of its transactions to another network layer $L1$. It allows for lower costs and network scalability, as $L2$ can be a smaller network with a fast consensus algorithm, while relying on a larger network $L1$ to keep proofs of its transactions. Such an approach might help protecting EHR with the formation of networks exclusive to authorized participants.

2.3.3 Distributed Hash Table and IPFS

Distributed Hash Tables (DHT) represent a content routing system in which data is distributed in a key-value format (IPFS Distributed Hash Table, 2022; MUBASHAR et al., 2021). Differently from a Blockchain, a DHT shares Content Identifiers (CID) and peer lists that provide such files. There are multiple algorithms to manage DHT routing, including Content Access Network (CAN), Chord, Kademlia, Pastry and Tapestry (ROEHRS; COSTA; ROSA RIGHI, 2017). One of the most popular public DHT implementations is called Interplanetary File System (IPFS)(LABS, 2024).

IPFS uses the Kademlia algorithm to manage its DHT (MAYMOUNKOV; MAZIERES, 2002). Kademlia uses a binary tree to represent nodes with a given CID in the shortest path. Its protocol ensures each node knows at least one peer in its CID sub-tree, if there is one. The purpose of Kademlia is to build a DHT on top of three system parameters(IPFS Kademlia, 2022):

1. An address space as a way that all of the network peers can be uniquely identified. In IPFS, this is all the numbers from 0 to 2^{256-1} .
2. A metric to order the peers in the address space and therefore; visualize all the peers along a line ordered from smallest to largest. IPFS takes $SHA256(PeerID)$ and interprets it as an integer between 0 and 2^{256-1} ;
3. A projection that will take a record key and calculate a position in the address space where the peer or peers most ideally suited to store the record should be near. IPFS uses $SHA256(RecordKey)$.

2.3.4 Asymmetric Cryptography and Digital Signatures

represent a set of underlying technologies on top of which the security of blockchain systems rely. Let (sk, pk) be a pair of private and public keys, respectively. A Digital Signature consists in applying the private key sk on an input data x to generate an output y , which can be verified using the public key pk , having $V(pk, x, y) = 1$, where V is a verification function. In recent years, new solutions, such as Decentralized Identifiers (DID), Smart Accounts, and signed typed messages standards (ERC-712) added support for more complex key management and digital signatures. Individuals, institutions, government agencies, and smart health devices might have a corresponding key pair as their main identification in a blockchain system, to sign health records and blockchain transactions. In self-sovereign models, users are responsible for protecting private keys, although many scenarios in healthcare require a trusted element, such as a government agency, to provide identity-pegged cryptographic key pairs.

2.3.5 Hash functions and Hash Structures

are one of the most widely adopted techniques for blockchain EHR, as they can represent external information to the blockchain in a format that can protect original data and reduce the use of on-chain storage. Given an arbitrary input M , $H(M) = M'$ is a n -bit hash function if collisions meet a security goal of $2^{n/2}$ work, where M' is a fixed-size representation of M . In addition to collision resistance, secure hash algorithms must have **preimage resistance** in which, given a value for M' , it is computationally infeasible to find M , and **second preimage resistance**, where given M is computationally infeasible to find a second input S with $H(S) = M'$ (KELSEY; SCHNEIER, 2005). Blockchain solutions for EHR can calculate hash functions on-chain using Smart Contracts (e.g. keccak256 in Ethereum), although when it is impractical or not desired to calculate such a value on-chain, it is highly recommended to provide additional information, such as the used hash algorithm and bit size (e.g. SHA-3, 512), along with the hash data.

whenever a batch of health records need to be stored on-chain at once, hash structures, such as Merkle Trees and Merkle Patricia Tries, become useful techniques. Such solutions consist of one root, multiple leaf and inner nodes, where every non-leaf node $N = H(N_i|N_{i+1})$ in a binary Merkle Tree is a hash derivation from its child nodes N_i and N_{i+1} . To proof if an arbitrary hash is part of a Merkle Tree, it requires a $\log_2(n)$ subset of its nodes to reach an arithmetically verifiable output, called Merkle Proof, where n is the number of leaf nodes in the tree. Such solutions are applicable to health monitoring scenarios, with potentially large volumes of health records generated from devices, such as remote monitoring devices, IoT, and wearables. For instance, let's consider a health monitoring device that collects $1KB$ of data every second; in one hour, it generates $3,600KB$; assuming a $256bits$ hash function, it generates a $32bytes$ digest for each record, $115.2KB$ total; for a well-balanced binary Merkle Tree, it requires twice the area for leaf and non-leaf nodes, resulting in a total of $230.4KB$ area, which is 93.6% lower than the original data size; each proof, in worst case, demands around 12 nodes to produce a valid proof, resulting in $384B$ for each proof.

2.3.6 Zero-Knowledge Proofs

Another prominent cryptographic technique for the protection and validation of EHR data is Zero-Knowledge Proof (ZKP) systems. This technique consists in having a prover P to prove to a verifier V that a statement is true by providing no detail about the statement other than a proof $x \in L$. Being P and V Probabilistic Turing Machines, an interactive proofing system $P(x) \leftrightarrow V(x, z)$ is considered valid if there is a system $S(x, z)$ that can reproduce the interactions between P and V in a probabilistic polynomial time, where z represents any prior knowledge P and S have from previous interactions. Non-interactive protocols such as the zk-SNARKS and zk-STARKS have become popular recently for blockchain solutions as

well. ZKP systems can be applied to validation of data against standards such as HL7 FHIR and OpenEHR, where proofs are provided to the blockchain along with hash data to ensure that the incoming data represent a valid and standard health record.

2.3.7 Proxy Re-encryption and Attribute-Based Encryption

access control is a key topic in healthcare. Considering that health records are potentially encrypted for privacy protection, such techniques become useful in managing data delivery upon request without compromising privacy (MHIRI et al., 2024). It consists of having a proxy $P(sk, pk)$ that is able to transform data encrypted by another key pair $A(sk, pk)$ using a re-encryption key $rk_{A \rightarrow P} = G(sk_A, pk_P)$, where G is the re-encryption key generator function. Thus, given an arbitrary message m and its corresponding encrypted version $C_A = E(m, pk_A)$, the proxy is able to re-encrypt it to $C_P = E(C_A, rk_{A \rightarrow P})$ and to decrypt it using its private key, having $m = D(C_P, sk_P)$. Such a technique has great utility for health records, as it allows data owners to delegate decryption permission to a third party, such as a hospital or a relative.

Attribute-Based Encryption (ABE) allows setting up a key pair with a master private key $A(Msk, pk)$ and encrypt m adding a set of attributes, having $C_A = E(m, pk, \{A_1, A_2, \dots, A_n\})$. The master private key can generate a secret key $sk_i = KeyGen(Msk, T)$ based on an access policy T , and this key can be used to decrypt C_A and obtain $m = D(C_A, sk_i)$. Such a technique is well suited for institutional access management to data, where a set of permissions can control access to EHR data.

2.3.8 Homomorphic Encryption

Represent a set of cryptography techniques that allow the execution of arithmetic operations using encrypted values. It has a good fit with EHR data, as it makes it possible to run data analysis without compromising data privacy (VANIN et al., 2023). Let E and D be encryption and decryption functions, respectively, a valid homomorphic encryption operation, with input messages m_1 and m_2 , has the following property for the addition operation $D(c_1 \oplus c_2) = m_1 + m_2$ and $D(c_1 \otimes c_2) = m_1 \times m_2$ for multiplication, where $c_1 = E(m_1)$ and $c_2 = E(m_2)$. Partially Homomorphic Encryption techniques (PHE), such as Paillier Cryptosystem, support only one operation type (addition or multiplication), while Fully Homomorphic Encryption (FHE), such as Brakerski-Gentry-Vaikuntanathan (BGV), support both addition and multiplication for the same set of encrypted values. Such techniques might be used to support decision-making in medicine and bioinformatics, as it is possible to design Machine Learning mechanisms that learn based on encrypted data (WOOD; NAJARIAN; KAHROBAEI, 2020).

A homomorphic cryptosystem in a given message space M is a quadruple (K, E, D, A) of probabilistically expected time-based algorithms conforming to the following conditions (AL-LOGHANI et al., 2019):

- Key Generation (K): a key pair $(k_e, k_d) = k \in K$ where K represents the key space;
- Encryption (E): consist in applying the key k_e on a message $m \in M$, producing a cipher-text c in the cipher-space C where $c \in C$;
- Decryption (D): consist in applying the key k_d on an encrypted message c to produce $m \in M$;
- Homomorphic Property (A): is a scheme that requires $c_1, c_2 \in C$ to produce a third element $c_3 \in C$ such that $\forall m_1, m_2 \in M$ holds only when $m_3 = m_1 \dot{m}_2$.

Thus, Homomorphic Encryption could support mathematical operations like summation, multiplication, and logic XOR operation (NAEHRIG; LAUTER; VAIKUNTANATHAN, 2011) directly on encrypted data. Calculation algorithms are highly dependent on the K element. The most popular key pair techniques used in the Blockchain context are Rivest-Shamir-Adleman (RSA) and Elliptic Curves.

There are basically two major groups of HE algorithms; Partially Homomorphic Encryption (PHE) and Fully Homomorphic Encryption. (FHE) (ROCHA; LÓPEZ; FALCÃO DA ROCHA, 2019). PHE algorithms support only one type of mathematical operations, like Paillier (addition), RSA and ElGamal (multiplication). However, FHE algorithms support more than one mathematical operation. The work of Gentry (GENTRY, 2009) set the ground-base for FHE with the concept of bootstrapping and was followed by many works, such as the Brakerski, Fan and Vercauteren (BFV) algorithm for FHE (FAN; VERCAUTEREN, 2012) that addressed performance issues towards a practical implementation of FHE.

HE has great applicability in many use cases, such as Big Data, Cloud Computing, Medical Applications, Financial Applications, Advertising, Pricing, and Smart Home Systems (AL-LOGHANI et al., 2019; NAEHRIG; LAUTER; VAIKUNTANATHAN, 2011; SHE et al., 2019). Data Analysis on encrypted data is a raising trend in technology in the last three years (RIMOL, 2021).

3 RELATED WORK

To design the proposed model, we selected relevant publications regarding the use of blockchain for PHR with distributed networks, especially focusing on data sharing and interoperability. We adopted a bottom-up selection strategy, mapping relevant use cases for blockchain and EHR to the state-of-the-art. In the following, we detail the three selected use cases for our work.

An **International Patient Summary (IPS)** document is an extract of an electronic health record that contains essential healthcare information about a subject of care. As specified in EN 17269 and ISO 27269, it is designed to support the use case scenario for 'unplanned cross-border care', but it is not limited to it. As it presents the challenge of compiling multiple historical data in healthcare, such access to data is key for a proper implementation of IPS. One important design principle in this use case is that the data set must be "minimal and non-exhaustive patient summary dataset, specialty-agnostic, condition-independent, but readily usable by clinicians for the cross-border unscheduled care of a patient". In the following, we describe the on-chain strategy recommended for this use case.

Digital Public Infrastructure for Health (DPI-H) is a concept that transcends the healthcare industry, although having significant contribution in this field. It describes a series of digital services provided by government agencies to support citizens and private industry. In healthcare, it can contribute in multiple areas, such as citizen and entities identification, data exchange and interoperability, personal health data, prescription and medicament, reporting, and supply chain tracking.

Internet of Health Things (IoHT) allows health devices equipped with sensors and actuators to interact and communicate over the Internet, working together to better understand the individual health of each person (DA COSTA et al., 2018). The term IoHT comprises many devices and solutions, including wearables and mobile applications, integrated with ad hoc sensors/actuators. Such devices contribute significantly to the collection of PHR, which play an important role in health records management, and might benefit from blockchain and cryptography components.

In the following, we describe the article selection criteria and related work analysis.

3.1 Selection Criteria

To create the proposed model, we selected relevant articles in Computer Science and Bioinformatics based on the following criteria.

- The healthcare use case
- Use of distributed network for EHR
- Use of cryptography techniques for EHR

Table 1: Related work, organized by use case, focus and support to health data standards.

Reference	Use Case	Focus	Data Standard
(ROEHRS; COSTA; ROSA RIGHI, 2017)	Interoperability	Data Exchange	OpenEHR
(GHADAMYARI; SAMET, 2019)	IPS	Data Exchange	-
(NIU et al., 2019)	Interoperability	Data Exchange	-
(SHARMA; HALDER; SINGH, 2020)	DPI-H	Data Exchange	-
(YAZDINEJAD et al., 2020)	Care Services	Access Management	-
(ZHUANG et al., 2020)	Care Services	Data Exchange	-
(WANG et al., 2021)	IoHT	Data Exchange	-
(SONKAMBLE et al., 2021)	Interoperability	Data Exchange	HL7 FHIR, OpenEHR
(ZHANG et al., 2022)	IoHT	Data Exchange	-
(KIM et al., 2022)	Care Services	Data Exchange	-
(ALI et al., 2022)	IoHT	Access Management	-
(REEGU et al., 2023)	Interoperability	Data Exchange	HL7 FHIR
(YANG et al., 2023)	Interoperability	Access Management	HL7 FHIR

- Year of publication
- Article relevance (publisher, citations)

We analyzed a total of 53 articles and selected the 13 most relevant considering aspects such as privacy protection, data access control, and the use of on-chain and off-chain resources. We organized the selected related work according to their use case, technical focus, and supported data standard in Table 1. Most works focus on data exchange, as interoperability is a major topic in the segment.

3.2 State-of-the-art

The data strategy is a major evaluation criteria, as it defines the role a blockchain solution can play in each proposed model. Most selected publications process encrypted EHR in off-chain components (ROEHRS; COSTA; ROSA RIGHI, 2017; NIU et al., 2019; SHARMA; HALDER; SINGH, 2020; WANG et al., 2021; ZHANG et al., 2022; ALI et al., 2022; YANG et al., 2023), which can set the basis for an end-to-end encryption model. However, some models rely on unencrypted EHR (GHADAMYARI; SAMET, 2019; ZHUANG et al., 2020; KIM et al., 2022; REEGU et al., 2023), which may raise privacy protection issues, especially when managing PHR, as in (KIM et al., 2022). In (SONKAMBLE et al., 2021), they adopt a data size

criterion, having smaller data on-chain, and larger files off-chain.

The use of on-chain resources to store the hash digest is a common strategy (SHARMA; HALDER; SINGH, 2020; ZHUANG et al., 2020; REEGU et al., 2023), although it might reduce the importance of blockchain, if there are no means for data validation. Some models run on-chain data validation, either for keyword search or for permission (NIU et al., 2019; SHARMA; HALDER; SINGH, 2020; WANG et al., 2021), using Zero-Knowledge Proof as a main technique. No method uses on-chain components to validate and keep proofs of health records, however, the work of (WANG et al., 2021) implements a mechanism to validate IPFS hash that could be applicable for EHR.

The data exchange technique is an important evaluation criterion as it drives how participating actors in the system will request and receive data. Access management techniques, such as control-based (Access Control List, Mandatory Access Control) (GHADAMYARI; SAMET, 2019; ALI et al., 2022), Attribute-Based Encryption (NIU et al., 2019; WANG et al., 2021; ZHANG et al., 2022), Proxy Re-encryption (SHARMA; HALDER; SINGH, 2020), and Secret Sharing (YANG et al., 2023). However, some models do not implement any cryptographic system for data exchange (ZHUANG et al., 2020; SONKAMBLE et al., 2021; KIM et al., 2022), which could lead to privacy protection issues. The work of (ROEHRS; COSTA; ROSA RIGHI, 2017) and (REEGU et al., 2023) delegates the task of data exchange to the data owner.

In terms of decentralized infrastructure, most related works adopt Ethereum (SHARMA; HALDER; SINGH, 2020; ZHUANG et al., 2020; SONKAMBLE et al., 2021; YANG et al., 2023) or Hyperledger Fabric (GHADAMYARI; SAMET, 2019; WANG et al., 2021; ZHANG et al., 2022; ALI et al., 2022; REEGU et al., 2023). However, the work of (NIU et al., 2019) and (YAZDINEJAD et al., 2020) does not specify which blockchain network was used, while in (KIM et al., 2022) a network simulator was used to evaluate the model. Roehrs et al.~(ROEHRS; COSTA; ROSA RIGHI, 2017) propose an architecture model called OmniPHR, which does not use blockchain for decentralized data processing. Instead, a Distributed Hash Table (DHT) was implemented to share small portions of health records, which they called "data blocks", adopting the Chord algorithm (STOICA et al., 2001).

DHT solutions represent an alternative for off-chain data storage due to their scalability and data availability mechanisms. In the related work, IPFS emerged as a common solutions for such models (SHARMA; HALDER; SINGH, 2020; WANG et al., 2021; YANG et al., 2023). Data analysis on encrypted data is also an important aspect we analyzed in the related work, as it can improve the decision-making support, while mitigating privacy issues. The work of (GHADAMYARI; SAMET, 2019) adopts the Paillier cryptosystem for data analysis, supporting count and mean operations. The work of (YANG et al., 2023) also set the basis for Homomorphic Encryption by adopting the ElGamal cryptosystem, which supports HE calculations. The work of (JIANG et al., 2020) adopts FHE to reduce the query time for health records, while preserving the privacy of patients.

The support for Decision-making in blockchain and cryptography adoption is an element in

our research. The work of (REEGU et al., 2023) presents a framework for blockchain adoption in healthcare called BCIF-EHR. They propose a patient-centric and HIPAA-based model with support for interoperability among different blockchain networks, by the use of Hash Locking, a system consisting of a secret/hash pair $(x, H(x))$, where $H(x)$ is public and x is kept secret. The method consists in locking an operation on the blockchain B_1 , while a related operation executes on the blockchain B_2 . In the occurrence of a transaction that provides the preimage x to B_1 , the lock is completed. The model covers the following use cases: Registration Phase, Pre-Agreement and Verification, Bank Fund Transfer and Certificate Generation, Health Record Exchange between Two Hospitals.

The work of (WANG et al., 2021) proposes a method for sharing PHR with privacy preservation, using a blockchain consortium. The model uses IPFS to store encrypted data and Attribute-Based Encryption for data search. They apply Zero-Knowledge Proof to create a "storage proof", to allow on-chain validation that the incoming data were effectively stored on IPFS. They tested the model with 500 attributes, using Hyperledger Fabric. The model does not adopt a data standard.

Reference	Off-chain Data	On-chain Data	On-chain validation	Data Exchange Technique	Blockchain Network	DHT
(ROEHR; COSTA; ROSA RIGHI, 2017)	Encrypted EHR	N/A	-	N/A	-	Chord
(GHADAMYARI; SAMET, 2019)	Unencrypted EHR	Encrypted EHR	-	Access Control List	Hyperledger Fabric	-
(NIU et al., 2019)	Encrypted EHR	Encrypted keywords	Keyword search (ABE)	Attribute-Based Encryption	Unknown	-
(SHARMA; HALDER; SINGH, 2020)	Encrypted EHR, re-encryption cipher	Identity hash, Encrypted EHR hash	Identity (ZKP)	Proxy Re-encryption	Ethereum	IPFS
(YAZDINEJAD et al., 2020)	-	-	-	-	-	-
(ZHUANG et al., 2020)	Unencrypted EHR	EHR Hash	-	URI to encrypted EHR	Ethereum	-
(WANG et al., 2021)	Encrypted PHR	Encrypted keywords, access policy, incentive mechanism	Permission, keyword search (ZKP)	Attribute-Based Encryption	Hyperledger Fabric	IPFS
(SONKAMBLE et al., 2021)	Larger EHR	Light-weight EHR	Hash calculation	Unencrypted data share	Ethereum	-
(ZHANG et al., 2022)	Encrypted PHR	Encrypted PHR	-	Ciphertext-Policy Attribute-Based Keyword Search	Hyperledger Fabric	-
(KIM et al., 2022)	Unencrypted PHR	Consent Information	-	Unencrypted data share	N/A	-
(ALI et al., 2022)	Encrypted PHR	-	-	Mandatory Access Control	Hyperledger Fabric	-
(REGU et al., 2023)	Unencrypted EHR	EHR Hash	-	Data encryption by patient	Hyperledger Fabric	-
(YANG et al., 2023)	Encrypted EHR	Secrets	-	Secret Sharing	Ethereum	IPFS
(MHIRI et al., 2024)	Encrypted EHR	-	-	Proxy Re-encryption	N/A	-

3.3 Open Gaps and Opportunities

As presented in this chapter, some models consider that whenever patients participate in the model, access to their data is automatically granted (ROEHRS; COSTA; ROSA RIGHI, 2017; GHADAMYARI; SAMET, 2019; SHE et al., 2019; SUN et al., 2020) without further authorization needed. When a third party generates a patient's private key, privacy issues may occur, which is why the encryption key pair generation is a crucial step, and some models consider patients as responsible for generating their own keys (ROEHRS; COSTA; ROSA RIGHI, 2017; MAHDY, 2020; MISBHAUDDIN et al., 2020; MADINE et al., 2020b; GHANI; ZINEDINE; EL MOHAJIR, 2021; WANG et al., 2021). Most models do not provide ways for patients, once granted access to PHR data, to formally revoke such access, except for the work of Sonkamble et al. (SONKAMBLE et al., 2021), Misbhauddin et al. (MISBHAUDDIN et al., 2020) and Ghani et al. (GHANI; ZINEDINE; EL MOHAJIR, 2021). Some models allow patients to encrypt PHR data (ROEHRS; COSTA; ROSA RIGHI, 2017; SHE et al., 2019; MADINE et al., 2020b; WANG et al., 2021), and only the work of Roehrs et al. with the OmniPHR model (ROEHRS; COSTA; ROSA RIGHI, 2017), and Wang et al. (WANG et al., 2021) provides ways for patients to manage data location, as most of the related articles consider health institutions as the only responsible for managing data location.

Most related work combine on-chain with off-chain storage because storing all medical records on-chain could raise scalability issues, as data is replicated in most/all nodes in the network. Most models put only the hash of PHR data on-chain (SHE et al., 2019; YAZDINEJAD et al., 2020; ZHUANG et al., 2020; MISBHAUDDIN et al., 2020; MADINE et al., 2020b; SUN et al., 2020), while the work of Sonkamble et al. (SONKAMBLE et al., 2021) with the MyBlockEHR model put lightweight data on-chain and bigger data files, like scan images and medical reports, off-chain. The work of Muizz Mahdy (MAHDY, 2020) considers medical records to be on-chain, whereas most models consider medical records to be off-chain. Some articles adopt decentralized file storage, such as IPFS or DHT (MADINE et al., 2020a; MISBHAUDDIN et al., 2020; ROEHRS; COSTA; ROSA RIGHI, 2017; SUN et al., 2020), as a solution for scalability.

The work of (REEGU et al., 2023) presents a blockchain-based framework to guide the decision-making for EHR. It focuses on the interoperability of HL7 and HIPAA, for entities such as hospitals, clinics, and insurance companies. For interoperability between different blockchain networks, their model adopts a Hash Lock technique. This technique consists in locking a given data in blockchain A, while it becomes available in blockchain B for usage. However, with this approach, a EHR cannot be updated while in another blockchain network. The model proposes the use of cloud infrastructure to store EHR. They framework develops four scenarios, including care service provisioning, healthcare payment, and data exchange between hospitals. A government agency is responsible for providing identity to the patients.

We summarize gaps and opportunities in Table 2.

Table 2: Gaps and Opportunities: a summary of gaps found in related work and that raise opportunities for contribution in the proposed model

Gap	Opportunity
Low or non-existing on-chain data processing for EHR, only data storage	Propose models to enhance the adoption of on-chain data processing, for data validation and analysis, on hash or encrypted data
Low specialization of cryptography and blockchain techniques for different EHR use cases	Identify key use cases for EHR and provide guidance for blockchain adoption
Lack of on-chain validation of incoming data, especially in hash format R	Provide methods methods to improve on-chain data validation, even on hash data
Patients are not able to manage access and data location. Once data is in the network, it is assumed that all members have access to PHR	Give patients the condition to decide who has access to PHR, where data will be stored and for how long
Centralized off-chain data location e.g. cloud servers	Adopt decentralized network, such as Distribute Hash Tables, for on-chain and off-chain records

4 MEPCA MODEL

In this chapter we introduce our scientific contribution for the use of blockchain and cryptography technologies for EHR. First, we introduce the MEPCA model with a map of key requirements for blockchain and EHR to use cases in health care industry. Next, we introduce a hash proof algorithm that allows for on-chain validation of HL7 FHIR required fields, without exposing sensitive data. The third element is a specification of the MEPCA model to promote end-to-end data protection for PHR, with Fully Homomorphic Encryption techniques for data analysis.

4.1 Design Decisions and Use cases

We propose the MEPCA model to improve the use of on-chain resources to process EHR data. Our approach aims to avoid risks of inserting malicious or invalid data into the blockchain, as most existing models use the blockchain only to store hash representations, with no proper data validation or analysis. MEPCA is an acronym of the five principles in our proposed approach (Maximize, Encrypt, Prove, Comply, Adapt), aimed to improve the use of blockchain and cryptography for health data processing, describe in the following.

We compare such strategies in Figure 3, considering important steps in the EHR, from data generation to decommissioning. In our recommended approaches, we include a step of on-chain data validation, where even if data hash is provided to the blockchain, it has means to validate if the provided hash corresponds to a valid EHR record. The steps in the lifecycle are described as follows.

1. Generation: when data are obtained from health devices and systems in a standard format;
2. Hash Calculation: calculates a digest based on each obtained file;
3. Hash storage: file contents are stored in the file system, mainly in a database;
4. Data proof: the generation of a cryptographic proof related to the input file, specially for structure, cardinality, and value domains;
5. Data analysis: knowledge obtaining from existing data, using techniques such as symbolic artificial intelligence and machine learning;
6. Validation: test a given data proof against a predefined set of criteria;
7. Encryption: apply a cryptosystem to convert the input data into an encrypted version;
8. Anonymisation: techniques to remove sensitive data from the input files, to protect the data subject identity;

9. Decommissioning: when a given data becomes invalid, unavailable or is removed from the system.

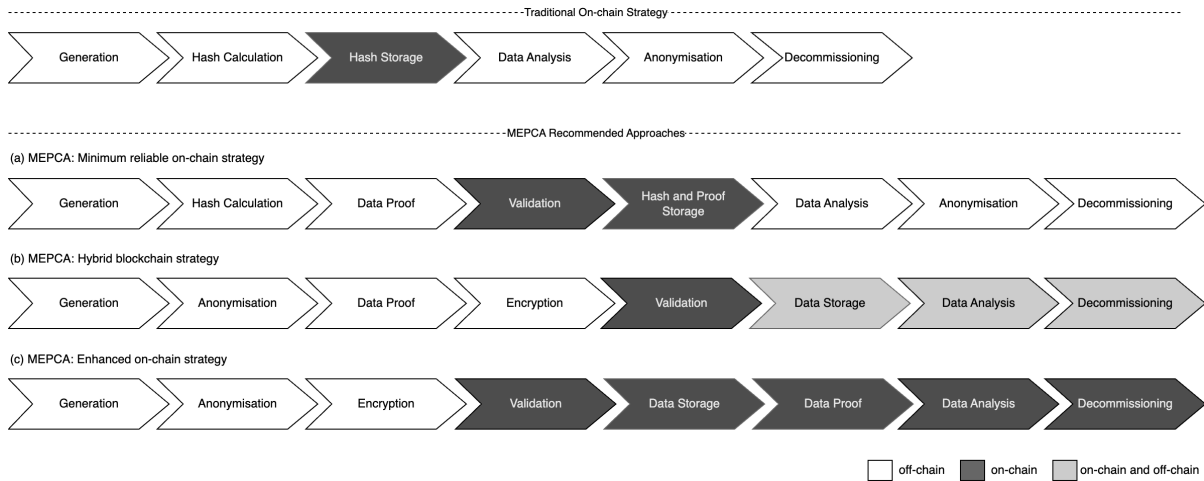


Figure 3: Blockchain strategies and on-chain usage comparison. First, the traditional approach, where only a hash representation is processed on-chain, with no data validation. Second, MEPCA recommended approaches: a) a minimum reliable model has on-chain data validity check to ensure incoming data complies with standards; b) a hybrid approach is an alternative where part of the data is stored into blockchain nodes and another part in traditional infrastructure; c) the recommended approach for most use cases has data analysis and proof generation on-chain, even on encrypted data.

1. Maximize on chain usage: to take advantage of the benefits the blockchain can provide, it is important to maximize the use of its elements whenever possible, otherwise it might become irrelevant;
2. Encrypt data whenever possible: privacy protection is a design principle that improves the reliability of health systems, and the existing data encryption techniques allow access management and data analysis, even on encrypted data;
3. Prove, dot not trust: from a blockchain perspective, incoming data must provide proof of its validity, not only a hash representation;
4. Comply with regulation: data subjects and healthcare institutions must comply with existing regulations such as GDPR, LGPD, and HIPAA, thus our proposed approach must consider it as a major guideline;
5. Adapt to the use case: there is no one-size-fits-all strategy for blockchain and cryptography adoption, multiple variations of architecture are possible according to factors, such as the segment, data scope, business requirements, technical restrictions, and roles involved.

#	Key Requisites for EHR	MEPCA Implementation Guidelines
KR-1	Input data should be in standard format (e.g. HL7 FHIR, OpenEHR)	Data validation mechanisms are necessary to ensure incoming data meet the desired standard; Data validation should occur on-chain
KR-2	Unique data subject identification	Adopt asymmetric keys to identify data subjects
KR-3	Data privacy and security	Data encryption techniques should take place for data in transit and at rest; Roles such as of Data Controller and their responsibilities should be unequivocally implemented; EHR metadata should indicate which fields are considered sensitive data
KR-4	Scalability and availability to support a high number of records and transactions	Data replication among peers in the network should respect a minimum of copies available
KR-5	Access control mechanisms should keep track of every read or right operation on existing data	Consent on data access should be recorded on-chain; Access privileges must have an expiration period
KR-6	Storage strategy should provide feasible hardware allocation, data availability and protection	Provide support to storing data in multiple formats, such as hash digest, hash with proof, encrypted, and compacted.

Table 3: Key requisites for MEPCA model implementation

Our work relies on a set of state-of-the-art techniques for distributed networks in a combination of cryptography, distributed systems, deeply inspired by the work of Satoshi Nakamoto (NAKAMOTO, 2008). Cryptography provides design elements regarding key management, data encryption and decryption, digital signatures, Homomorphic Encryption, Zero-Knowledge Proofs, Merkle Trees and hashing algorithms. Distributed systems contribute to the design by adding elements for data availability, synchronization, fault tolerance and parallel processing. The MEPCA proposed model respects the following design principles:

1. Data subject (or any caregiver/authorized personnel) has control of their corresponding health records;
2. All data access requires a corresponding express authorization from patient (or responsible personnel) auditable on the blockchain
3. Any granted data access will have a predefined time frame;
4. Data subjects can, at any time, revoke permission to data for any participant;
5. Regardless the incoming data format (raw, encrypted, hash digest), there are means to validate such data on-chain;
6. End-to-end data encryption elevates the security level, while supporting data analysis by the use of Homomorphic Encryption;

Technique/Use Case	International Patient Summary	Digital Public Infrastructure	Internet of Health Things
Asymmetric Cryptography and Digital Signature	Transaction signature	DID	Signed Typed Messages
Hash Functions	Represent granular data	N/A	N/A
Hash Structures	Merkle Proofs for EHR verification	Merkle Proofs for PHR verification	Merkle Proofs for PHR verification
Zero-Knowledge Proofs	Data validation	Data validation	N/A
Homomorphic Encryption	Data analysis	Public interest reports	Data analysis and reporting
Proxy Re-encryption and Attribute-Based Encryption	PRE for summary exchange	PRE for data subjects, ABE for institutions	ABE for purpose-based role permissions
Consensus Algorithms and Transaction Finality	Permissioned network with public access. Instant finality for on-chain access management	Layer-1 network for general purpose transactions	Layer-1 public network for permission management
Subnetworks	N/A	Layer-2 networks for specialized applications	Layer-2 network for metadata
Block Data	IPS data, Merkle Trees, Merkle Proofs	Operations metadata on layer-1, granular data on layer-2	Roles, permissions, and metadata

Table 4: MEPCA components and use cases

7. Some data might be publicly accessible in encrypted format to provide public interest information;
8. Access to data should have an expiration date, to allow data owners to define in advance the period they wish to concede access to their health records.

The MEPCA model is a multipurpose blockchain adoption strategy for EHR. To support practical implementations of the model, we describe three use cases in health care. In Table 4 we present a summary of the application of MEPCA components to the proposed use cases.

4.1.1 Use Case 1: International Patient Summary (IPS)

We recommend using on-chain storage for the IPS dataset instead of for granular data that generate the summary. However, every granular record should have a corresponding proof of its usage, thus a Modified Patricia Trie with hashes from all input data is recommended to generate proofs and be provided along with the summary. If granular data need to be on the chain for arbitrary reasons, we recommend encrypting all registries and applying homomorphic encryption techniques to generate the patient summary whenever possible. Homomorphic Encryption might also be applied to obtain information from an encrypted patient summary, reducing the need for plain data. When there is a demand to access plain data, Proxy Re-encryption techni-

ques are recommended, as they allow for IPS to stay encrypted at rest within blockchain nodes and be provided to requester for their exclusive reading when authorized by the data owner.

In terms of blockchain architecture and protocol, we recommend using a permissioned network, where all participating nodes have previous authorization to hold IPS data and control access to it. The nodes might be publicly accessible for data reading, as IPS aims to be on a worldwide scale, thus reinforcing the need for end-to-end data encryption. The consensus protocol and transaction finality are highly influenced by whether access authorization is held on-chain or off-chain. Whenever access management runs on-chain, it is recommended that the network have fast consensus and instant transaction finality. When it is held off-chain, the network might operate with probabilistic or financial finality, although it is recommended that proofs of each request/authorization pair are stored on-chain for auditing and transparency purposes.

4.1.2 Use Case 2: Digital Public Infrastructure (DPI-H)

MEPCA components can set the basis for multiple DPI-H services. For digital identification services, we recommend the use of asymmetric key cryptography and the adoption of DID, where the government can generate claims for individuals and institutions. Key rotation or Proxy Re-encryption mechanisms are recommended for individuals to manage access to EHR. For care services infrastructure, we recommend ABE, where government can establish standard roles for health professionals to deliver interoperable access management.

For data strategy, we recommend the adoption of a hybrid approach with subnetworks, where DPI-H services provide layer-1 nodes to keep metadata and proofs of EHR operations in subnetworks, maintained by medical institutions. Such an approach can also deliver greater performance and scalability, as most transactions will occur in layer 2. For PHR, we recommend storing Modified Patrice Tries to keep data proof on layer 1, while granular data might reside in off-chain infrastructure or at a layer 2 network. Public reporting services can receive encrypted data from medical institutions to produce public interest information using Homomorphic Encryption techniques for on-chain data analysis and privacy protection.

4.1.3 Use Case 3: Internet of Health Things (IoHT)

A blockchain network for IoHT and PHR can improve care services with data interoperability, health monitoring, and data analysis (BENNACER et al., 2023). This network will have the main purpose of keeping control of permissions over data and providing metadata for search. It will not store PHR data due to the fact that they reside in heterogeneous applications and in high volume, making it not viable to replicate efficiently in a blockchain network. However, we recommend that data access permissions and revocations have an on-chain proof counterpart, along with metadata that allow to query for existing data and players with access to it. A pu-

blic layer-1 network to maintain access permissions and a layer-2 network for metadata are the recommended approach.

We recommend the definition of purpose-based role set for data access, e.g. care services, data analysis, research, data steward (VANIN et al., 2023). Such roles will drive the design of an ABE system, where data subjects can manage which keys have access to data, and for which purpose. Permission events should be in the form of Signed Typed Messages, in accordance to EIP-712. Thus, each authorization can be validated and stored on-chain, in accordance with MEPCA principles, despite the fact that the data are off-chain. Every time a new data set becomes available for a role, a metadata transaction on such data should become available on the layer-2 network.

Metadata should be in standard format, such as FAIR Data Objects (GROUP, 2021) to facilitate queries. Such queries might help requesters identify which type of data (e.g. heart rate, blood pressure), coverage period, and supported roles are available for a given data subject. Based on this information, a data access request can be created, so a data subject can add the requester to a role, making it possible to have interoperability from other players in the same role.

4.2 End-to-end data protection for PHR

In this chapter, we apply the MEPCA model to address two significant issues regarding PHR/EHR interoperability in distributed health systems: How to give patients complete control over their data and obtain relevant information without exposing individual data. The proposed method combines Blockchain Architecture with IPFS and Fully Homomorphic Encryption (FHE) to deliver a distributed health system in which patients control their data and support relevant information calculations on encrypted data.

The main data flow in our model is summarized in Figure 4. Our focus is to provide patients complete control over PHR data throughout the lifecycle, applying end-to-end encryption to improve privacy, separating data from metadata to improve scalability, applying role segregation to prevent conflicts of interest, and Fully Homomorphic Encryption to provide flexibility with calculations even on encrypted data. We introduce a new role called Data Steward, which will run IPFS nodes to manage PHR data via temporary areas called Data Vault, triggered by requesters and approved by patients.

Patient control over PHR is a crucial characteristic in the proposed model. One of our design principles is that “patient (or any caregiver/authorized personnel) has control of all data access”. To accomplish that, patients are responsible for creating their cryptographic key pair. The private key will encrypt data and sign transactions, while health institutions will use the public key to encrypt patient data. Patients could store data in many different ways, be it on their premises or in specialized service providers, like a Data Steward as we introduce further.

The model considers health institutions as data consumers grouped into consortia with spe-

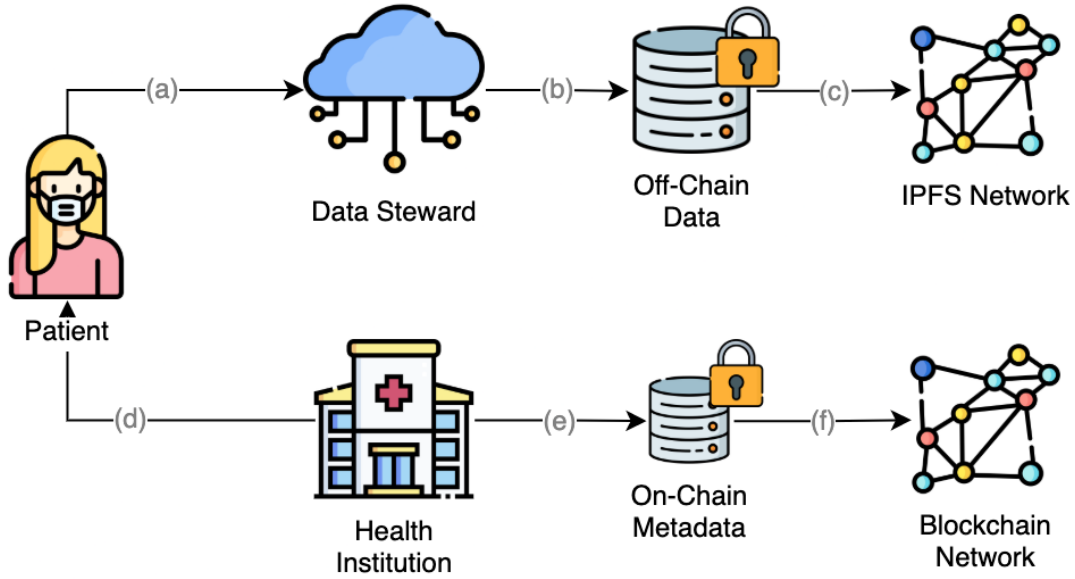


Figure 4: Solution Data Flow: Patient authorizes a Data Steward (a) to keep an encrypted version of PHR off-chain (b) on an IPFS network (c); A Health Institution requires a patient for PHR (d) and, whenever authorized, PHR metadata can be shared on-chain (e) in a blockchain network (f).

cific purposes (e.g., clinic, research, data analysis). Whenever a health institution generates data about a particular patient, it should be provided to the corresponding patient to store and protect new records. In every circumstance in which there is demand for a patient's data (b), the patient (or a previously authorized person) will formally grant access to requested data for a predefined period and a specific consumer (c).

All requested data are provided to a specific health consortium, encrypted using the requesting institution's public key (d). Requesting institution can fetch data and decrypt it outside the Blockchain and provide the data to other consortium members whenever requested. Patients have the right to revoke access to data or retire data at any time.

Data stored outside the blockchain network are said to be off-chain, while data already shared among participants in a distributed network are said to be on-chain. Our model considers metadata to be on-chain and PHR to be off-chain. The responsibility over data is delegated from the patient to Data Stewards and from Data Stewards to health institutions in the consortium to respect the design principles to guarantee end-to-end encryption and control over data access, scope, and availability for a predefined period.

Our model relies on the gaps and opportunities identified in the related work (Chapter 3) to propose and architecture formed by two complementary distributed networks: an IPFS network, as in (MADINE et al., 2020b; WANG et al., 2021; MUBASHAR et al., 2021), for off-chain data dedicated to plain text PHR and one blockchain network for on-chain data, focused on metadata, as in (MADINE et al., 2020b).

In Figure 5 we present all the main components of the architecture and introduce two elements that allow better segregation of responsibilities regarding patient data and clinic treat-

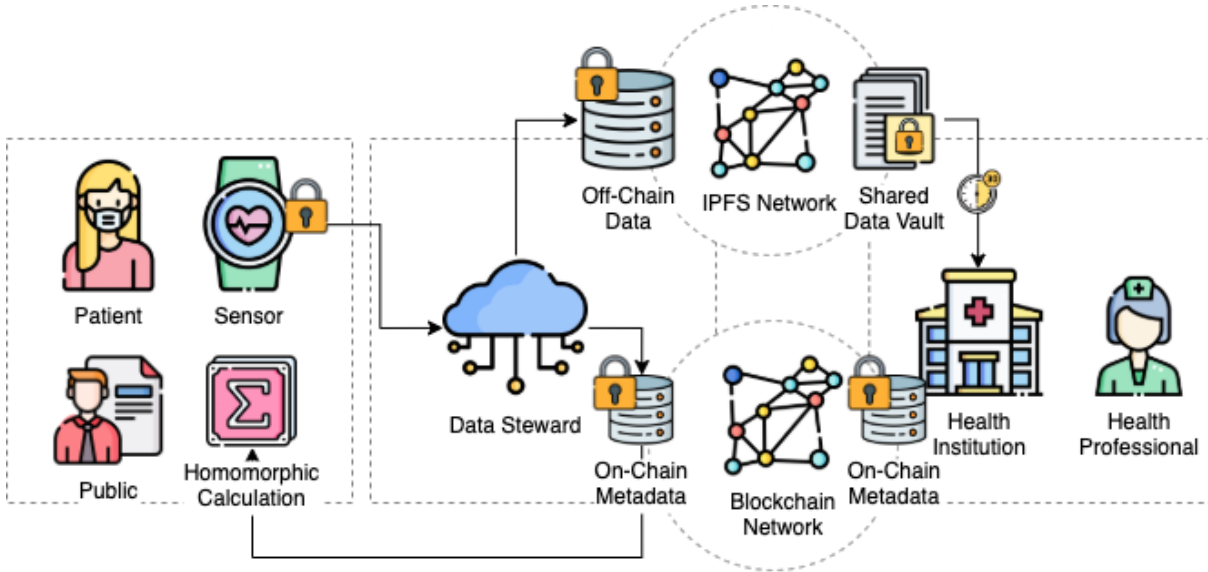


Figure 5: Architecture Components: IPFS Network for off-chain data, Blockchain Network for on-chain metadata, Data Steward to keep encrypted version of PHR and Shared Data Vault to support temporary data sharing encrypted with requester public key.

ment, called Data Steward and Shared Data Vault. These elements respond to health records in different phases of the lifecycle, allowing data sharing with end-to-end protection. We describe both elements in the following chapters.

4.2.1 Data Steward

Data Steward is a role in our model focused on storing data on patient's behalf. They are service providers and their role can be performed by public or private institutions. Each Data Steward represents a distributed network of IPFS nodes that store data on behalf of patients, originated by third-party health solutions, such as sensors, monitoring devices or mobile applications, and encrypted using the public key of the corresponding patient.

Health data will be made available in a Shared Data Vault whenever requested and approved by the patient. Requests are data payloads in the format: **{date, consortium-public-key, institution-public-key, requester-public-key, data-scope, finality, requested-period}**. The payload can be in the form of a QR Code, accessible by patients through a mobile application with their private key, to sign the payload and send a transaction to the Data Steward authorizing data access for a period. Requested data will be re-encrypted by patients using the requester public key and posted to a Shared Data Vault.

Patients can give permission to a specific health institution or a health consortium. When allowing access to a consortium, we could adopt a technique such as attribute-based encryption (NIU et al., 2019) to provide access to all members of the consortium.

4.2.2 Shared Data Vault

Shared Data Vaults (SDV) represent temporary file Content Identification (CID) distributed on the IPFS network and made available to fulfill a specific request from a health institution. They are created by a Data Steward, only with express authorization from the patient, using encrypted data sent directly from them. For interoperability, SDVs respect a market standard such as HL7 FHIR (SARIPALLE; RUNYAN; RUSSELL, 2019) or OpenEHR (OPENEHR, 2020), as each institution runs a different Health Management System (HMS)(MELLO et al., 2022). For the FHE calculation, the data should be in numeric array format. To meet the design guideline of “all data access will have a predefined time frame”, each data vault should have a predefined expiration time. We present the access management process in Figure 6 and describe the steps in the following.

1. Patient authorizes a given consortium to access their data
2. Patient requires the creation of a Data Vault and a data scope from Data Steward;
3. Patient decrypts data with private key and encrypt with health institution’s public key
4. Data Steward creates a Shared Data Vault and returns the CID to patient
5. Patient shares CID with health institution
6. Once time window expires, Data Steward removes (unpin) the file from IPFS

As Shared Data Vaults store data encrypted with a specific institution’s public key, the given institution will be able to decrypt data using their private key. This situation does not represent any security issue related to privacy exposure, as only patients can authorize such access for a predefined period.

The data analysis process runs on data from an SDV. In Table 5, we present an example with the main steps that start when a health institution requires access to a set of PHR data until the execution of FHE algorithms for data analysis. Covers technical and clinical aspects involved in the process, which can drive the design of systems with the proposed architecture.

4.3 On-chain hash proofs

A minimum reliable on-chain approach should be able to meet the key requirements described in Table 3, and some use cases could benefit from the increased adoption of cryptography and blockchain technologies. Having the ability to validate incoming hash data on-chain is a major scientific contribution in our model. It allows for a higher level of auditability and validation from a blockchain perspective, as many models opt to store EHR data off-chain and only send hash digest to the blockchain.

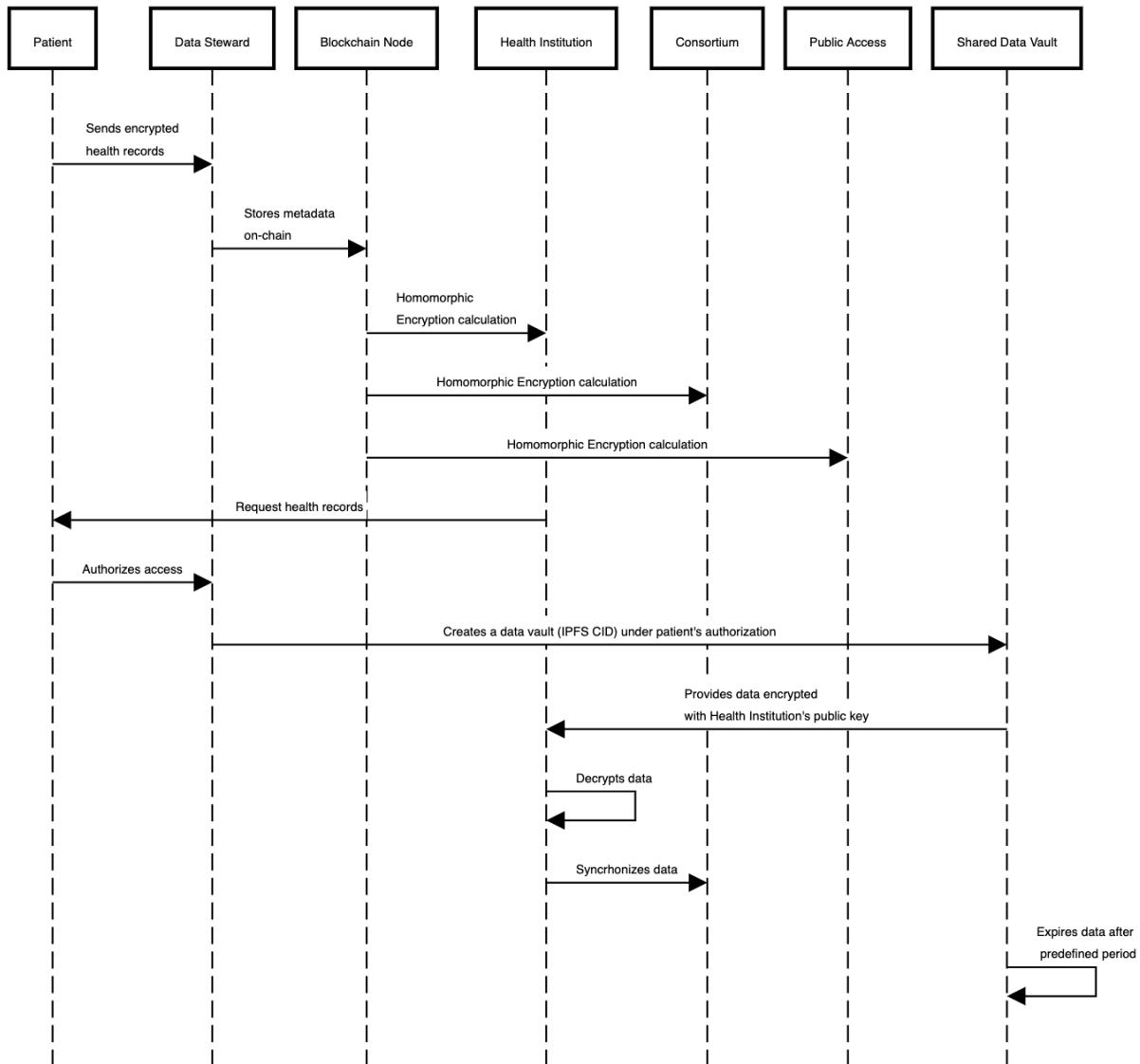


Figure 6: Solution Sequence Diagram with three primary use cases represented: 1) Patient stores encrypted health records in Data Steward nodes (IPFS network); 2) Patient stores encrypted metadata in Blockchain Nodes; 3) Patient authorizes health institution and consortium to access a portion of data in a Shared Data Vault for a predefined period.

Table 5: Applied Example

Medical Event	Technical Aspects
Patient enters a medical institution for care	Patient shares their data with a Data Steward; a previous virus test result is already shared to the network via Data Steward; patient carries their Key Management app along to the medical institution
Health professionals request access to medical records	Medical institution communicates with Data Steward to request data access; the patient receives notification and signs transaction;
Patient allows access for ten days	A Shared Vault is created between Medical Institution and Data Steward (Zero-Knowledge Proof)
The screening process is conducted	Health professionals, patient, and medical institutions sign a transaction with data
The medical record is updated	Blockchain is updated with encrypted data
A new statistics round is started	Homomorphic encryption computation is executed to obtain updated statistic data

We introduce an algorithm to verify whether a hash digest preimages to a file containing a predefined set of HL7 FHIR required fields. In Algorithm 1, we detail the on-chain data validation process in the recommended approach (a), adopting Zero-Knowledge Proofs as a way to verify the incoming hash data. We detail this strategy in the following.

Assume an input HL7 FHIR file J in JSON format with a set of fields $F = \{f_1, f_2, \dots, f_n\} \in L$, where L is the language schema that describes J . There is a subset $F' \subset F$ with all required fields in L . The proposed hash validation method consists of a prover P being able to prove to a verifier S that he knows F' (the secret) by providing only the hash h and a proof π , without revealing any value in J . Thus, for each arbitrary set F' , there is a maximum 2^k possible proofs that satisfy:

$$H(w, \pi) = h \quad (4.1)$$

$$M \leftarrow H(\{F' \cup \phi, (F - F'), V\}) \quad (4.2)$$

where:

H = a hash function

M = a Sparse Merkle Tree

h = the root of M

F' = the set of required fields in L

k = the size of F'

ϕ = a NULL set with $\lfloor 2^{\sqrt{k}+1} - k \rfloor$ elements

M_1 = the leftmost sub-tree of M , with leaf nodes $\in F' \cup \phi$

w = the root of M_1

V = the set of field values in J

Π_f = the Merkle Proof of $f \in F'$ in M

π = the proof of h in the form: $\Pi_f - M_1$

Notice that any field combination in F can produce a valid proof. M is a sparse tree because we add a set of null values ϕ to its leaf nodes to meet the quadratic structure of Merkle Proof calculations and allow for a deterministic construction of M_1 . The proof π is a set of hash values with size $\log(n+m) - \log(k)$. Removing M_1 hashes from Π_f to obtain π strictly reduces the size of the $\log(k)$ elements. For instance, an arbitrary file J with 50 fields where 10 are required, the proof π will be composed of 4 hash values. In Algorithm 1 we describe the verification process to be executed on-chain and validate h against π .

We provide a diagram to demonstrate the flow for Hash Proof in Figure 7. The flow from the bottom to the top of the diagram, starts with the input JSON file, with the list of required fields. These fields will be used to set up the witness w in the verifier. The prover generates a Sparse Merkle Tree M_1 having required fields and phi at the leftmost ramification. The remaining leaf nodes will be composed by non-required fields and values from the file. The file hash will be the Merkle Root that, along with the generated proof, allows for a verification if the hash contains or not the required fields. I the demonstration $\Pi_{f1} = H(H_A, H_B, H_{CD}, H_{EFG})$ and $\pi = H_{EFG}$ is the proof.

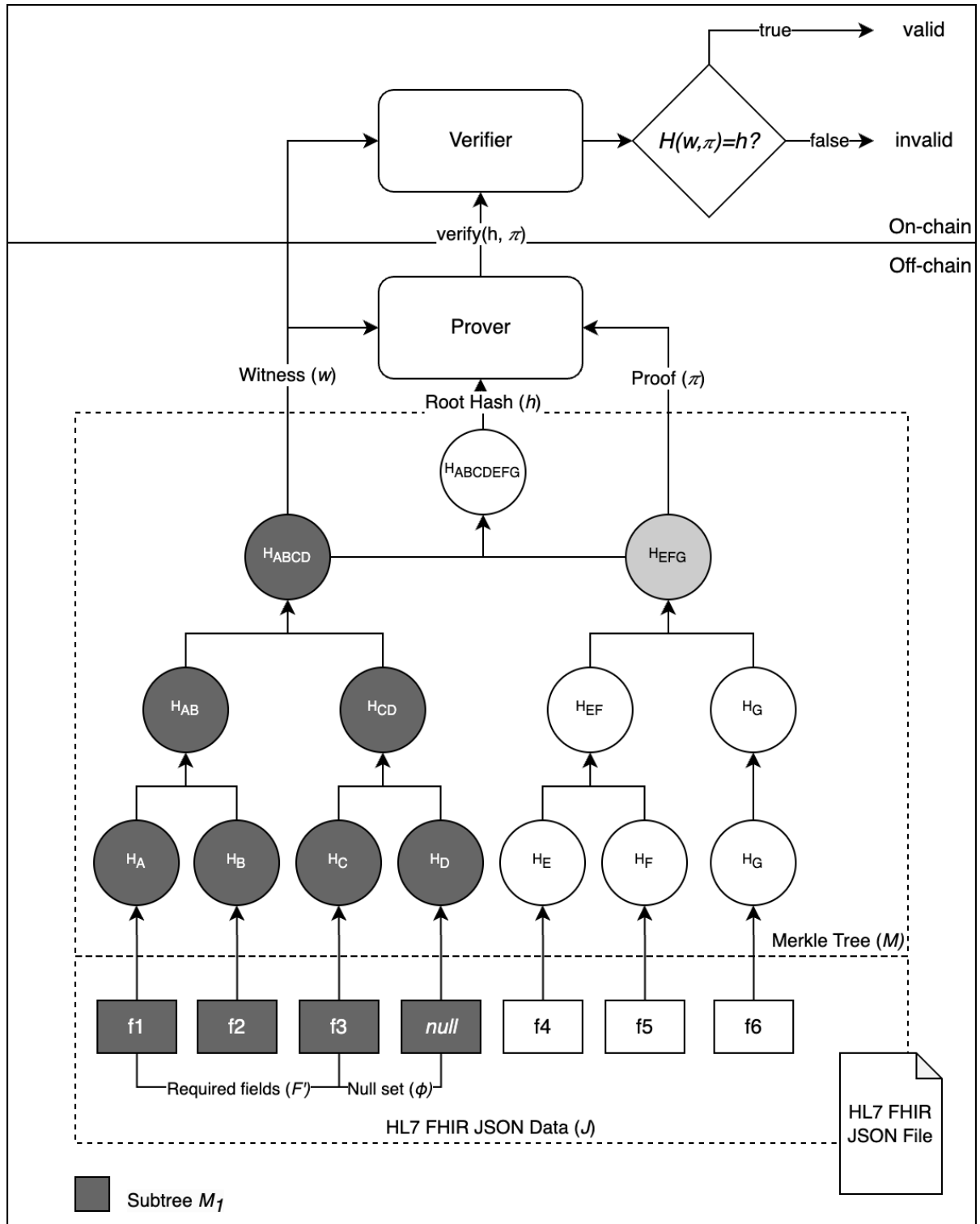


Figure 7: On-chain hash proof algorithm for HL7 FHIR data: a Merkle Tree based system for Zero Knowledge Proof generation. The system is able to produce verify if a given hash digest contains required fields in the input. Required fields map deterministically to specific leaves in a Merkle Tree, making it possible to use the complementary values from a Merkle Proof to act as proof to a verifier.

Algorithm 1 EHR on-chain validation. Requests need to provide the used hash function and a Zero Knowledge Proof of data validity along with hash data. It validates hash size and algorithms, then verifies the proof against a pre-defined witness.

$w \leftarrow \text{GenerateWitness}(F')$

$r \leftarrow \text{random}()$

procedure VERIFY(h : Hash, π : Proof)

if $\text{len}(x) \notin [32, 40, 64, 128]$ **then**

return *false*

▷ Invalid hash length

end if

$h \leftarrow H(r)$

▷ Generate hash of a random value

if $\text{len}(h) \neq \text{len}(x)$ **then**

return *false*

▷ x and H not compatible

end if

if $H(w, \pi) = h$ **then**

return *true*

▷ Proof is correct

end if

return *false*

▷ Proof is not correct

end procedure

5 RESULTS AND DISCUSSION

In this chapter, we present the evaluation methodology and experiment results for the MEPCA model. We have two test sets, one to evaluate the impacts of on-chain strategy for HL7 FHIR data, in raw and hash format, with a special focus on the Hash Proof algorithm and its impact on processing time and storage to obtain a proof of validity for the input data. The second test set is aimed to evaluate data analysis on encrypted data. We set a simulated blockchain network and run a FHE algorithm to extract information from an encrypted dataset with COVID-19 registries. Both experiments can provide a deep technical understanding of the practical applicability of the MEPCA model for real healthcare use cases.

5.1 Methodology

The proposed model considers the following phases in data lifecycle: a) patient authorizes a Data Steward to protect PHR in an off-chain IPFS network; b) Data Steward keeps a version of encrypted PHR; c) Data Steward synchronize with IPFS network peers; d) Health Institution requests access to a set of PHR; e) patient formally authorizes or rejects the request; f) if authorized, Data Steward creates a Shared Data Vault instance with a predefined duration. We can divide these steps into two different moments, the first focused on setting up a new Data Steward, as described in Algorithm 2, and the second focused on the request for PHR from a Health Institution to a patient, as described on Algorithm 3.

Algorithm 2 starts with the patient P sending a transaction to the blockchain notifying that the Data Steward D will hold their data for future requests. The method *postPHR* is responsible for sending PHR to the authorized Data Steward. The first step consists on the patient encrypting the data with using their private key. After the encryption, data is sent to the Data Steward, that returns with the content identification hash from IPFS. With such a data, a transaction is sent to the blockchain notifying that a new set of PHR data became available at the Data Steward.

In Algorithm 3 we describe the process of PHR data request from a Health Institution H . The request comes in the form of a blockchain transaction from the Health Institution H , with the requested patient public key, the requested data scope and a duration t . If approved by the patient, a Shared Data Vault will be created, containing the requested data encrypted with the requester public key. The method returns the Universal Resource Identifier (URI), with the location where the requested data can be accessed.

Here we present a methodology to analyze the proposed method in its main aspects, like data management, process and calculations. We assume that such calculations need to process millions of records in a matter of hours, even on encrypted data. To have a detailed evaluation of the proposed model, some technical aspects are relevant, as described in Table 6.

Algorithm 2 Procedure for access authorization for PHR sharing. It takes as input a Data Steward and as output, shares encrypted PHR with Data Steward if authorized by patient

Patient P
 Data Steward D
 Blockchain B
 PHR p
 Blockchain Transaction tx
procedure AUTHORIZEDATASTEWARD(D) ▷ Data Steward
 $tx \leftarrow P.authorize(D)$
 $B.sendTransaction(tx)$
 return $true$ ▷ Data Steward authorized
end procedure
procedure POSTPHR(p, D) ▷ PHR instance
 $encryptedPhr \leftarrow P.encrypt(p)$
 $pProof \leftarrow P.signData(timestamp)$
 $p \leftarrow \mathbf{new} \text{ PHR}(encryptedPhr, pProof, P.publicKey)$
 $cid \leftarrow D.postPhr(p)$
 $tx \leftarrow P.signTransaction(p.proof, p.sig, p.pubKey, dProof, D.publicKey)$
 $B.sendTransaction(tx)$
 return $[cid, tx]$ ▷ CID and transaction hash
end procedure

Algorithm 3 PHR request from Health Institution to patient

Patient P
 Data Steward D
 Health Institution H
 Shared Data Vault V
 Duration t
 Blockchain Transaction tx
 Requested data description d
procedure REQUESTPHR(tx)
 $[H, d, t] \leftarrow tx.extractParams()$
 $authorization \leftarrow P.checkAuthorization()$
 if $authorization = \mathbf{true}$ **then**
 $CIDs \leftarrow D.searchData(d)$
 while $cid \in CIDs$ **do** $PHRs \leftarrow D.getPhr(cid)$
 end while
 $rawPhr \leftarrow P.decrypt(PHRs, P.privateKey)$
 $encryptedPhr \leftarrow \text{Encrypt}(rawPhr, H.publicKey)$
 $V \leftarrow \mathbf{new} \text{ SharedDataVault}(encryptedPhr, H, t)$
 return V ▷ Shared Data Vault
 end if
end procedure

Table 6: Evaluation aspects for the proposed model

Evaluation Aspect	Evaluation Criteria
Hash data validation	Proof calculation time; Proof size
End-to-end encryption	Processing time; Algorithm profiling
Data Analysis on encrypted data	Processing time
Storage occupation raw data compared to encrypted data	Comparison analysis over a data set
Network performance	Transactions per second (TPS)
Privacy protection during all steps	Security analysis

5.1.1 On-chain data processing

To evaluate the practical applicability of blockchain solutions for on-chain management of healthcare data, we propose a method that processes HL7 FHIR data on a Hyperledger Fabric network (HYPERLEDGER, 2024a). Hyperledger is a solution for permissioned networks. Each network can be constructed with a custom set of parameters that suits specific needs for decentralized data processing, including the consensus algorithm (Practical Byzantine Fault Tolerance or Raft) and the State Database for smart contracts (MongoDB or CouchDB). It supports Smart Contracts, also called chaincodes, in Go, Java or Javascript language. Participants in the network are called organizations, and each organization can utilize its own Certification Authority (CA) solution to generate key pairs for users.

A Hyperledger Fabric infrastructure has two types of processing units, called peers and orderers. Peers are responsible for receiving and processing transactions, while orderers are responsible for reaching consensus with other peers, with instant finality. After consensus is reached, peers can run logic in Smart Contracts, update its State DB and return to the caller with the requested data. We created a four-node Hyperledger Fabric 2.5 (HF) network and used Hyperledger Caliper 2.0 (HYPERLEDGER, 2024b) to benchmark performance using the test data set. Hyperledger Caliper is the official benchmarking solution for Hyperledger Fabric. It supports multiple approaches to stress test networks.

To create the test data set, we used Synthea (WALONOSKI et al., 2018), a synthetic patient generator that models the medical history of random patients. We generated 10 thousand synthetic patients. The files are in JSON format and have a file size ranging from 1MB to 5MB. The total data set size is 48GB.

Each node in the network has a 8 cores 3.00GHz Xeon E5-2600 processor, and 16GB of memory, running HF peers and orderers using Docker containers. The network design is presented in Figure 8. It has two organizations (Org1, Org2) with two servers each. Each server runs a Hyperledger peer and observer. The experiment consisted in sending all registries in the data set to the blockchain and then querying each registry by its identification number. We configured Hyperledger Caliper in three different scenarios, as follows:

- Fixed Rate: aims to reach a given TPS as much as possible. Might incur into pending transactions
 - Write Parameters:
 - * TPS: 100
 - Read Parameters:
 - * TPS: 1000
- Fixed Load: aims to keep a given number of transactions presented to the system by adjusting the TPS during the test. Might incur into pending transactions.
 - Write Parameters:
 - * Transaction Load: 40
 - * Starting TPS: 100
 - Read Parameters:
 - * Transaction Load: 500
 - * Starting TPS: 1000
- Maximum Rate: aims to reach the best TPS without pending transactions
 - Write Parameters:
 - * Starting TPS: 100
 - * TPS increase for each interval: 5
 - * Sample Interval (seconds): 20
 - Read Parameters:
 - * Starting TPS; 1000
 - * TPS increase for each interval: 5
 - * Sample Interval (seconds): 20

We tested the Hash Proof algorithm using the same dataset, extracting patient registries from each HL7 FHIR file, as the Synthea library generates bulk files by default. Then, each patient record was submitted to test the Hash Proof algorithm using SHA-256 as the hash function, varying the number of required fields from 1 to 14, to measure to processing time in milliseconds and the storage occupation in Bytes, using the hash digest (64 Bytes) as baseline, and the proof calculation in (WANG et al., 2021) as a reference value.

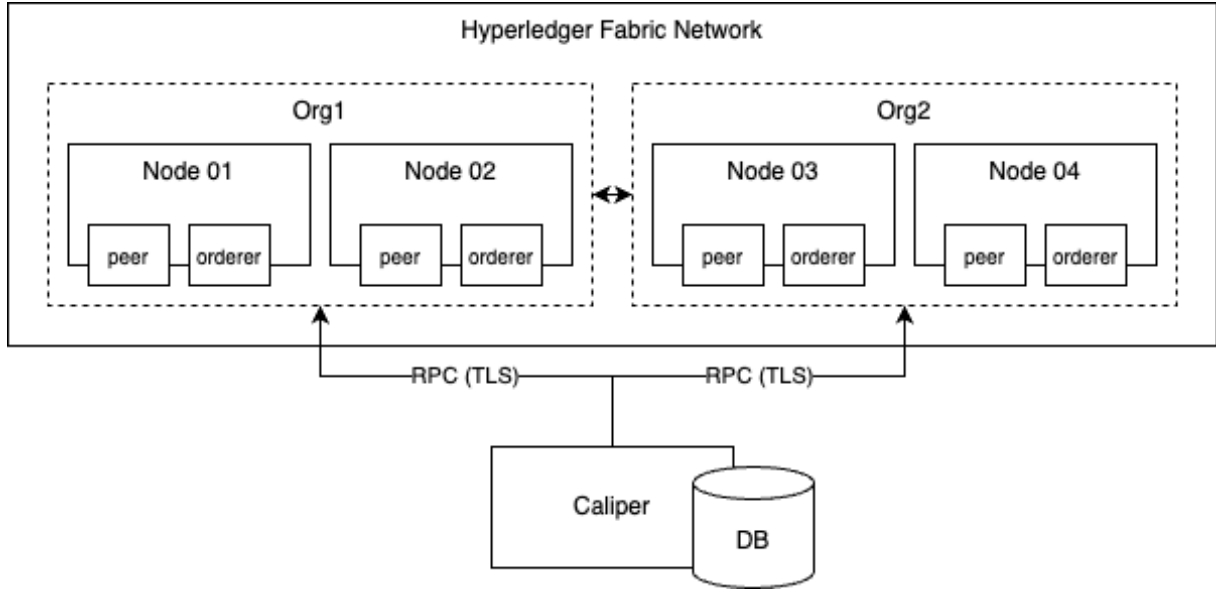


Figure 8: Experiments network architecture using Hyperledger Fabric 2.5 and Caliper 2.0

5.1.2 Data analysis on encrypted data

End-to-end encryption is an important element in the proposed model, as it focuses on data privacy throughout all phases of the PHR lifecycle. Data encryption, decryption and encrypted calculations with Homomorphic Encryption consume more processing and storage area when compared to raw data processing, so it is necessary to understand which step consumes more resources and time (profiling). We will analyze the algorithm with multiple data sets to compare its performance on different payloads and data format, such as HL7/FHIR, OpenEHR and plain text format.

In terms of materials, we selected an open data set from the Centers for Disease Control and Prevention (Centers for Disease Control and Prevention, COVID-19 Response, 2021)¹. The data set includes 22.5 million records of anonymized patient data. We chose a subset of data related to people between 60 and 69 years, which resulted in a total of 1.285 million records. For development, we used the SEAL library (MICROSOFT SEAL, RELEASE 3.6), which implements the BFV (Brakerski, Fan, and Vercauteren) algorithm for FHE (FAN; VERCAUTEREN, 2012). The application code is written in Javascript, using Node.js version of SEAL Library, and data is loaded from CSV files in raw data and encrypted during execution. The application simulates a Blockchain node that calculates the number of infected patients from the data set. For data compression, we used the standard **zlib** package from Node.js. To calculate block propagation time, we used a Blockchain Network Simulation tool called Simblock (AOKI et al., 2019).

¹The dataset is available on the CDC Website: <https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf>

5.2 Experiments and Results

We conducted 2 different test sets to evaluate different aspects of MEPCA model. In the following, we present each experiment configuration and the obtained results. In Chapter 5.3 we discuss the obtained results.

5.2.1 Test Set 1: HL7 FHIR on-chain processing

After three rounds of testing with each configuration scenario, we collected results for read and write process separately for Fixed Rate, Fixed Load and Maximum Rate approaches. We compare the results for Transactions per Second (TPS), latency, error rate and pending transactions in Table 7. Fixed Rate write reached highest TPS (around 83 transactions), and also the highest latency (around 1.15s). It also had a 40 pending transaction average during the experiments, and a percentage of 0.1% failed transactions. The fixed load resulted in a 12 pending transaction average, with no failed transactions. Maximum Load method generated no pending or failed transactions, as the method is designed to prevent such behaviors.

Table 7: Evaluation results for data write: Fixed rate reached best results, but with higher number of pending transactions and some failing transactions. Fixed Load reached 64.3 TPS with 12 pending transactions, while Maximum Rate generated no pending transactions, while delivering 51.6 TPS.

	Write TPS	Latency (s)	Error Rate	Pending Transactions
Fixed Rate	83.2	0.39	0.090%	40
Fixed Load	64.3	0.27	0.000%	12
Maximum Rate	51.6	0.15	0.000%	0

In Figure 9 we present TPS results for write and read operations, comparing the selected strategies. Fixed Rate delivered 83 write and 987 read TPS, Fixed Load achieved 64 write and 1022 read, and Maximum Load achieved 51 write and 642 read TPS.

In Figure 10 we present the results for the evaluation of the Hash Proof algorithm. We adopt the hash calculation as the baseline for our method, as it is the best performing technique, achieving $0.009ms$ processing time. We use the work of (WANG et al., 2021) as reference value with $0.68ms$ for the proof calculation, while our method achieved a processing time of less than $1.7ms$ for the worst case (1 required field) and achieved $1.2ms$ for a higher number of required fields. For storage occupation, we present the results in Figure 11. MEPCA demanded $384Bytes$ in the worst case (1 required field), achieving better results than (WANG et al., 2021) ($192Bytes$) for a higher number of required fields. In our model, as the number of required fields increases, the processing time reduces, as well as the proof size in Bytes.

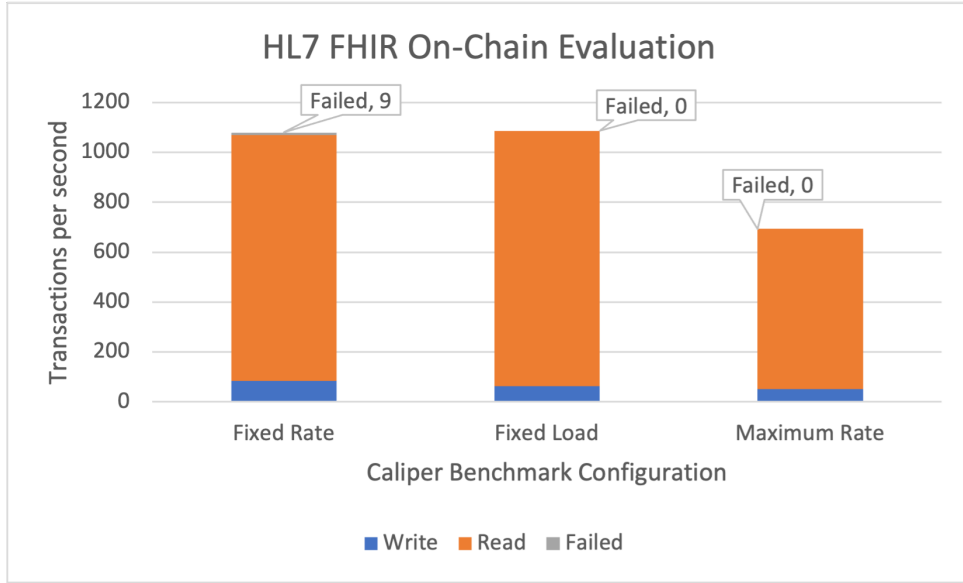


Figure 9: Transactions per second for read and write. Fixed Rate delivered better results for write, while Fixed Load delivered the best overall performance. Maximum Rate reached worst result of the three strategies. Fixed Rate resulted in failing transactions due to system overload.

5.2.2 Test Set 2: Fully Homomorphic Encryption

We evaluated the processing time performance on encrypted data to the plain data processing as in (ALABDULATIF; KHALIL; YI, 2020), organized in groups of 100k, 300k, 600k, and 1.2M records. The encryption process analyzes different key sizes for Polynomial Modulus $n \in \{1024, 2048, 4096\}$, equivalent to (GHADAMYARI; SAMET, 2019), while security and privacy scenarios are analyzed with a set of experiment scenarios as in (WANG et al., 2021; SHE et al., 2019). The network simulation is based on (SHE et al., 2019) and simulates a network with three different regions running 10, 30, and 100 nodes with block sizes varying from 535KB to 2140KB. Also, all experiments were run on a 3.2 MHz 8 cores computer with 16 GB RAM.

In this chapter, we analyze the performance of the FHE algorithm in the context of the proposed model. We aim to measure the required time to encrypt, process, and decrypt PHR data. Thus, we set up a scenario where a node in the network needs to calculate over a whole set of encrypted data.

BFV algorithm setting is as follows: Polynomial Modulus Degree $n = 4096$, Coefficient Modulus $q = 109$ (according to default recommendation in (MICROSOFT SEAL, RELEASE 3.6)) with 128 bits Security Level. Results comparing raw data and encrypted data processing are presented in Table 8 and show that, for more than one million records, computation on encrypted data executes in less than 3 minutes, even when processed by regular hardware.

The work of (JIANG et al., 2020) reached around 60s to query an encrypted dataset with 10 thousand registries, using FHE. Our results reached 1.3M registries in less than 30s. After

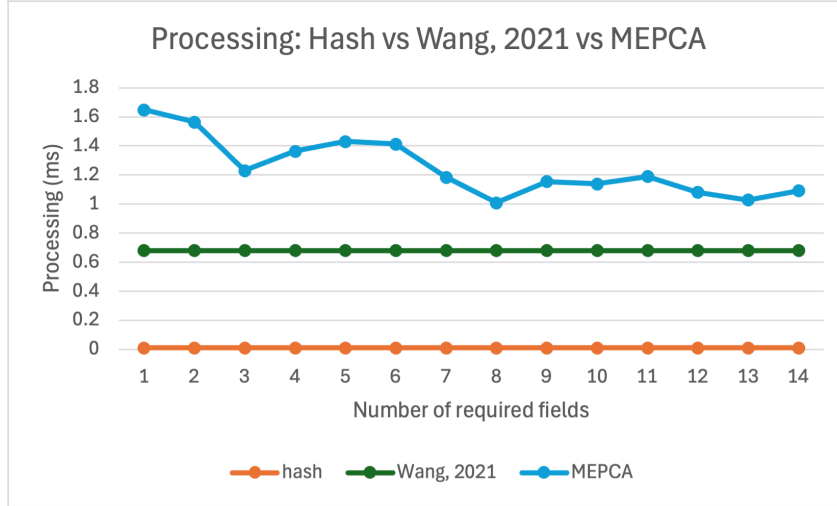


Figure 10: Hash processing in milliseconds compared to (WANG et al., 2021) and MEPCA. As the number of required fields increases, the processing time reduces. The baseline hash calculation for SHA-256 took $0.00993ms$ to calculate, and the reference value from (WANG et al., 2021) took $0.68ms$.

Table 8: Performance evaluation: FHE calculation on different dataset sizes comparing the raw data to encrypted data (in seconds). The second and third columns show the overall calculation time for addition in seconds comparing raw data and ciphertext respectively.

Records	Average Time Plain (s)	Average Time Encrypted (s)
100k	0,0130	11,158
300k	0,0404	33,079
600k	0,0774	65,826
1.3m	0,1673	138,559

applying FHE to reduce the dataset size for queries to 800 registries, (JIANG et al., 2020) was able to run the calculation in $7s$ average time.

Table 9: Algorithm profiling: parameters n and q for Polynomial Modulus Degree and Coefficient Modulus respectively in BFV encryption. The steps of encryption, addition and decryption are in the following rows.

n	q	Encryption (ms)	Addition (ms)	Decryption (ms)
1024	27	2.367	0.011	2.579
2048	52	2.355	0.011	2.465
4096	86	2.435	0.011	2.662

We analyze storage consumption considering the same parameters as in Table 9 and for each pair (n, q) we measured the string size in raw data, encrypted data, and compressed encrypted data. For compression in Node.js, we used the library **zlib**. Each test round had 100k registries. The results are presented in Table 10.

The total space necessary to store 100k registries in encrypted format could reach almost 7GB of storage with a higher level of encryption (n, q) . We reached a compression rate of

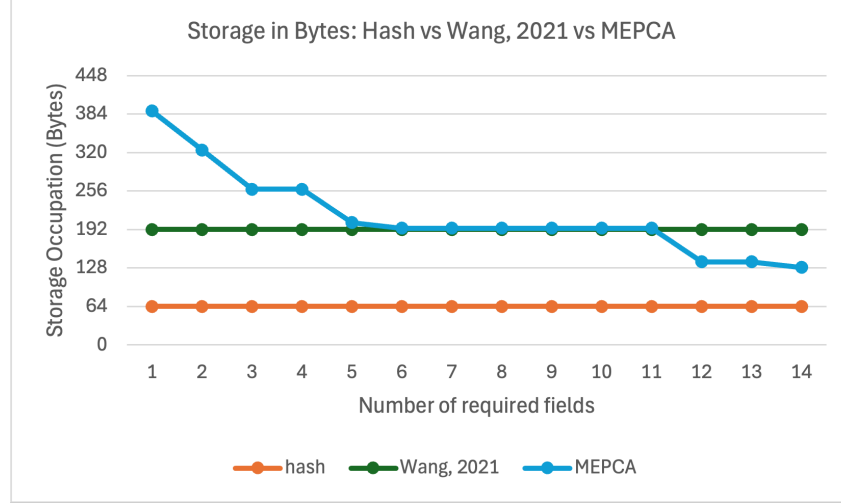


Figure 11: Hash storage compared to (WANG et al., 2021) and MEPCA. Proof calculation takes $364B$ for one required field and decreases as the number of required fields increases, as the proof size requires less data for validation. It reaches better results compared to (WANG et al., 2021) for higher number of required fields.

Table 10: Storage consumption: as the (n, q) pair increases, the ciphertext string size in KB also increases. Each column S has the average size of each registry in plain, encrypted and compressed format, while each column T has the total storage amount for a set of 100k registries

n	q	Plain Text		Encrypted		Compressed	
		S_{plain}	T_{plain}	S_{enc}	T_{enc}	S_{comp}	T_{comp}
1024	27	212B	21MB	11.5KB	1.12GB	8.7KB	851MB
2048	52	212B	21MB	40.6KB	3.96GB	30.7KB	3GB
4096	86	212B	21MB	71.5KB	6.98GB	54.2KB	5.29GB

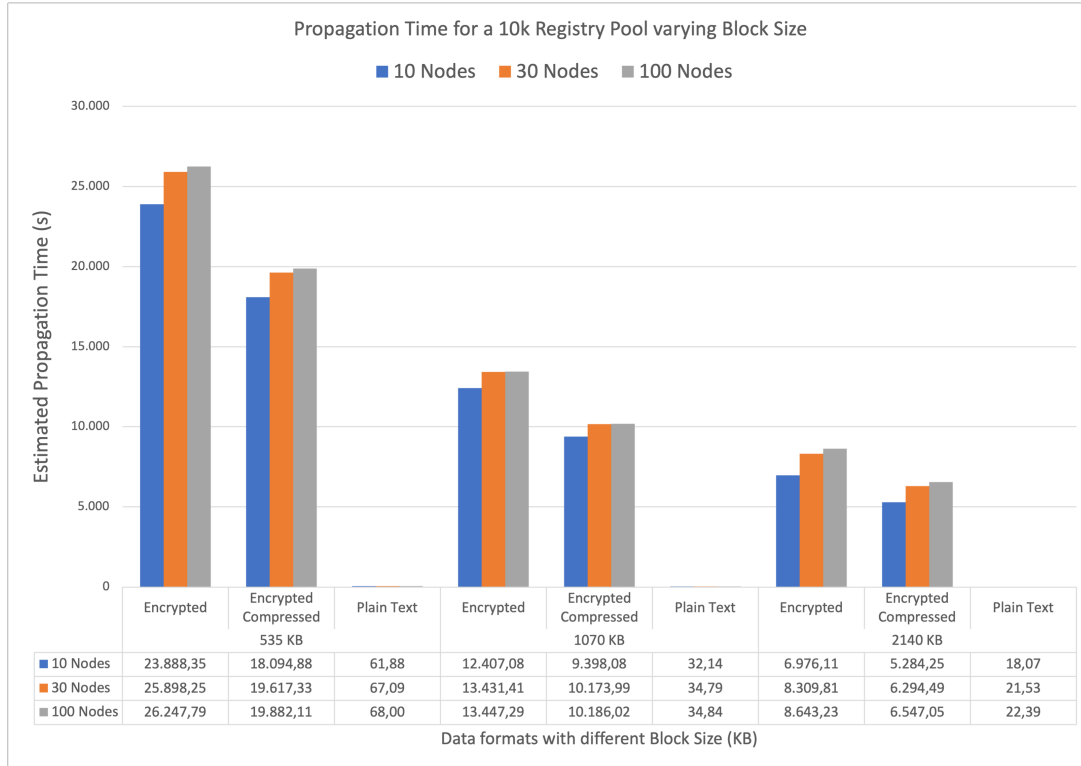


Figure 12: Blockchain Network Simulation: 10k pool size with 500B of block header and 5 seconds block time. Scenarios compare different block sizes (535KB, 1070KB, and 2140KB) in three formats (raw data, encrypted cyphertext, and compressed encrypted cyphertext) with 10, 30, and 100 nodes distributed in 3 different regions. The chart summarizes the time in seconds for 10k registries to propagate in the network.

$\approx 24\%$, which could reduce about 1.7GB of space consumption.

To estimate the time to propagate a given block pool size to a blockchain network, we simulated a network using Simblock, a Blockchain Network Simulator (AOKI et al., 2019). The network parameters are as follows: three different regions (North America, Europe, and South America), a block time of 5 seconds, a blockchain header size of 500 Bytes, and an average pool size of 10.000 registries.

We tested the network by varying the following parameters: the number of nodes in the network $n \in \{10, 30, 100\}$, the block size (in Bytes) $b \in \{535, 1070, 2140\}$ and the median record size for a $n = 4096$ encryption with 212B for raw data, 84,924B for encrypted ciphertext and 64,328B for compressed encrypted ciphertext. Each simulation round generated 100 blocks with each corresponding propagation time in seconds. In Figure 12 we demonstrate the results for each scenario. With a pool size of 10 thousand transactions with 500B of block header and 5s block time, scenarios compare different block sizes (535KB, 1070KB, and 2140KB) in three formats (raw data, encrypted ciphertext, and compressed encrypted ciphertext) with 10, 30, and 100 nodes distributed in 3 different regions. Encrypted data require higher propagation time, especially compared to plain text. Compressed encrypted format achieved a better result compared to encrypted format.

Table 11: Security and Privacy scenarios: comparison

Indicator	Compromised Data	Data Format	Extension	Mitigation
Exp-1	Single patient's PHR	Encrypted/Plain	Patient's history; new records	Two-factor authentication; Trusted Peers list
Exp-2	Data Stewards off-chain data	Encrypted	A group of patients' PHR	Not necessary
Exp-3	Medical Institution's off-chain data	Encrypted/Plain	A group of patient's PHR	Reduced Shared Data Vault duration
Exp-4	Consortium node	Encrypted	Incoming transactions; synchronized ledger	Not necessary

Our network is based on the work of She et al. (SHE et al., 2019) but, as the article does not provide numbers regarding propagation time, it is not possible to compare the results. Nevertheless, raw data take a matter of seconds to propagate, while encrypted and compressed encrypted ciphertexts take a matter of hours to propagate 10k registries. Larger block sizes provide more efficient propagation time, and compressed ciphertext provides a 20% to 25% time reduction in block propagation.

Data security and patient privacy are critical to PHR interoperability. Based on (SHE et al., 2019; YAZDINEJAD et al., 2020), we propose a set of vulnerability scenarios to analyze how each scenario impacts each model building block from a security and privacy point of view. In Table 11 we summarize the scenarios, highlighting the type of data at risk, the extension of the risk, and the appropriate mitigation strategy for each scenario. In the following, we provide a detailed description of each scenario.

1. Patient privacy is compromised when someone obtains unauthorized access to the patient's key k_e or any other means of communication with the Data Steward or any third-party Health Platform (off-chain data). Considering: a) all PHR are encrypted with the patient's key; b) Requests for Shared Data Vault should provide a public key; c) One given patient does not have access to other patient's data. In this case, Data Stewards could implement means of two-factor authentication (2FA) to mitigate access to sensitive data, and requesters could demand Trusted Peers list participation in order to mitigate unauthorized access;
2. Data Steward's security compromised: occurs when off-chain records or private key k_e of a given Data Steward have unauthorized access. Considering: 1) all PHR are encrypted with the patient's key k_e ; 2) Data Stewards cannot create a Shared Data Vault without authorization (signed transaction). In this case, there is no PHR exposition as they are encrypted, and Data Stewards do not have access to the patient's private key;

3. Medical Institution's private key compromised: occurs when unauthorized access to key k_e requests Data Stewards for a Data Vault. Considering: 1) all PHR are encrypted by each corresponding patient's key k_e ; 2) Data Stewards cannot create a Shared Data Vault without patient's authorization (signed transaction); 3) Any open Shared Data Vault has a limited duration. In this case, only data in open Shared Data Vaults is subject to exposition for some time;
4. A node in the consortium compromised: occurs when the server running the node or the private key suffers unauthorized access. Considering: 1) nodes receive transactions with encrypted data; 2) its corresponding counterpart signs each transaction; 3) each node synchronizes blocks with other peers with encrypted data; 4) nodes peer in a consortium. In this case: all data available to the compromised node in data vaults become accessible for a while; peered nodes in the consortium can ban the compromised node from the network and remove access to data.

5.3 Discussion

The proposed model proved that it is possible to process a new PHR for a given patient in less than 1 second to become fully encrypted, consuming less than 10KB of storage. This is computed by a Data Steward and shared with a Blockchain Consortium with 100 institutions across three different continents in less than 30 seconds. It is a significant benefit to the segment because it protects individual records and supports interoperability among health institutions at the same time without prohibitive technical limitations in performance or storage allocation.

Encryption, calculation, and decryption occur in different moments, considering how the data flow works in the proposed model. Thus, we analyze the impact of Homomorphic Encryption for each situation. Encryption is the most time-consuming step and occurs whenever a patient sends data to the Data Steward, or there is demand for data of a specific patient from the health consortium. Decryption takes around half of the encryption process. This process occurs whenever a node in the health consortium needs to obtain a result from a calculation or receive encrypted data from a patient in a Shared Data Vault. Calculation tends to be the most frequent operation in the model and is also the best performing cryptographic step. When calculations run on high volumes of data, it takes less than 3 minutes to process more than one million records.

The work of (JIANG et al., 2020) reached around 60s to query an encrypted dataset with 10 thousand registries, using FHE. Our results reached 1.3M registries in less than 30s. After applying FHE to reduce the dataset size for queries to 800 registries, (JIANG et al., 2020) was able to run the calculation in 7s average time. Such results demonstrate how FHE calculation can support real use cases and improve performance with proper pre-processing.

Regarding security and vulnerability scenarios, our work offers a differentiation compared to the related work of Yazdinejad et al. (YAZDINEJAD et al., 2020) and She et al. (SHE et al.,

2019), where private keys are managed by health institutions and not by patients. This scenario raises security issues where the institutions could generate transactions on the patient's behalf without consent.

Write requests demand more hardware, as transactions need to reach consensus among peers, and then persisted in the State Database, before returning to the requester. On the other hand, reading requests only validate access permissions before returning the requested data. It explains the fact that read requests throughput is more than 10 times faster than the write requests.

In a practical scenario, health institutions are not able to keep incoming transactions at a controlled level, as with fixed-load and maximum-load methods in Hyperledger Caliper. Thus, it is important to consider that, with higher number of concurrent transactions, the system might experience pending transactions, and even failed transactions, due to overload.

The results show that with 3 organizations, the system is able to reach an average of 83 TPS, which means that less than 28 transactions are made for each organization a second. As the number of participating organizations grows, TPS tends to reduce due to latency or the minimum required approvals for a transaction to reach consensus.

The Hash Proof algorithm demonstrated practical applicability, considering the benefit it brings to hash-based models. To validate a hash digest on-chain, it requires $1.6ms$ off-chain processing to calculate the proof and additional $384B$ on-chain storage to keep it, which can be acceptable in most use cases, considering its benefits.

Thus, to build a practical use case for health data that comply with existing standards such as HL7 FHIR or OpenEHR, and supports higher TPS amount, one might consider strategies that combine off-chain and on-chain strategies, having raw data in a different infrastructure, like cloud storage, Directed Acyclic Graphs (DAG) such as the Interplanetary File System (IPFS) and Arweave, and having a partial representation of the each registry on-chain.

Compared to (REEGU et al., 2023), the MEPCA model proposes different approaches to interoperability. While the BCIF-EHR model adopts Hash Lock for interoperability, we propose the creation of a Shared Data Vault, where data can be made available under patient approval. Our model is more suited for PHR and IoHT use cases, where data is more granular. Granular data in the BCIF-EHR model might become impossible to track, as a significant number of hash locks would be necessary to fulfill a request, while in our model, an SDV will be created on demand, including only the data scope from the requester. Nevertheless, for less granular data use cases, the MEPCA model suggests adopting layer-2 blockchains, that can keep interoperability with other networks with more scalability.

Compared to (WANG et al., 2021), our Hash Proof algorithm achieved a close performance, with higher degree of validation, having values around $1ms$ for proof generation, compared to $0.68ms$ in the reference model. While in (WANG et al., 2021) they validate only if the provided hash was really added to the IPFS, our model allows to verify if the given file has the required fields and avoid invalid or malicious data entering the blockchain. The reference model lacks

such a validation, thus we consider that an average higher performance cost around $0.3ms$ might be acceptable for most cases, due to the benefit it generates.

We conducted the experiments using a permissioned blockchain (Hyperledger Fabric), but the same model can also be implemented in public blockchains. However, with public blockchains, some technical parameters, such as block size and consensus algorithms, cannot be modified, which might imply into higher processing time. Transaction costs might also impact public blockchain strategies although, we indicate adopting Layer-2 networks in our model to reach higher performance and lower costs compared to Layer-1 networks.

It is important to discuss the impact of some elements and their impact on the test performance. The adoption of container servers for the nodes makes it easier to manage a larger infrastructure and run the experiments, but add a performance overhead to the processing time. The adoption of NodeJS for the chaincodes makes it easier to develop and test, but also might add a performance overhead, when compared to other languages supported by Hyperledger Fabric, such as Go. However, the strategy of storing data or metadata on the blockchain might have no significant impact on the system performance. Metadata tends to require less storage space, so the any impact will be perceived in the storage occupation, depending on each file size.

6 CONCLUSION

This work proposes the MEPCA model to improve the on-chain processing of EHR. We propose a set of five principles for improved blockchain adoption for EHR, and introduce new technical elements to support our model: a set of design principles and uses cases to drive blockchain adoption for EHR, a hash proof algorithm for on-chain HL7 FHIR hash data validation, Data Steward and Shared Vaults to segregate responsibilities related to patient data and. With an end-to-end encryption model, it is possible to support the exchange and calculation of information regarding healthcare without exposing individuals due to the Homomorphic Encryption technique.

We explored different aspects regarding privacy, performance, and node communication on the blockchain. A technical evaluation of Homomorphic Encryption algorithms demonstrate the applicability of such techniques. The algorithm proved to calculate over more than one million records in less than 3 minutes and allow sharing a new PHR entry in less than 30 seconds, which could support calculations and publication of pandemic outbreak data in practical applications. Homomorphic Encryption provides a set of techniques to support the calculation of statistics on encrypted data as a mechanism to protect data privacy and provide public interest information at the same time.

The results show that a fully on-chain strategy might not be viable in a practical scenario with multiple health institutions generating and validating data, as the highest TPS is around 83, with a peak of 40 pending transaction. Such a scenario might have a direct impact on the user experience (waiting lines, service time) and be not acceptable.

In Chapter 1.2 we presented the research hypothesis for the MEPCA model. In the following, we analyze each hypothesis against the observed results.

1. There is an opportunity to increase the adoption of on-chain strategies and cryptography compared to the existing literature: the MEPCA model demonstrated that operations such as data validation and data analysis can be performed on-chain, without exposing data privacy;
2. A consistent set of design principles and a mapping between relevant use cases in health care and key building-blocks in blockchain and cryptography can drive decision-making and technology adoption: the MEPCA model provided a set of key requirements, design principles and use case mapping to support decision-making. However, the model could provide more decision-making tools, such as a data governance model and an assessment tool;
3. The ability to add proofs to hash data for on-chain validation can improve auditability of data existing in the network: the Hash Proof algorithm proved to be well-suitable to validate required fields in HL7 FHIR JSON format. It can be improved to add support to other rules in JSON schema validation, such as data format;

4. End-to-end encryption techniques can support data analysis in an acceptable processing time when compared to raw data processing and reduce the demand for unencrypted data: the use of FHE algorithms for data analysis demonstrated great potential for data analysis, with performance able to process more than 1 million registries in less than 30 seconds.

6.1 Scientific Contribution

The MEPCA model contributes in the healthcare segment with a combination of design guidelines and applied algorithms to improve on-chain EHR processing. We focus on giving full control of data to Data Subjects, from origin to sharing, from interoperability to data monetization, from permission to revocation, addressing current regulation, such as HIPAA, GDPR, and LGPD. In the following, we summarize the specific contributions of this work.

1. The MEPCA model maps key requirements for blockchain and EHR to relevant use cases, with guidelines for decision-making. The five principles in the MEPCA model (Maximize, Encrypt, Prove, Comply, Adapt) have the potential to drive a consistent adoption of on-chain EHR processing, reducing the risk of introducing invalid or malicious data into blockchain nodes
2. The Hash Proof algorithm is a significant advance in the construction of cryptographic tools to enhance on-chain data processing, by providing proofs of data instead of arbitrary hash counterparts
3. An application of the MEPCA model for end-to-end data protection of PHR, adopting FHE algorithms to allow data analysis on encrypted data, with proven performance

We strongly believe that our model can support decision making and promote an improved adoption of blockchain and cryptography for EHR, by the providing of a profound analysis of important use cases in the segment, with a technical analysis of key requirements and a technical evaluation of the model building-blocks. We introduce a novel technique to validate HL7 FHIR JSON files based on hash digest and a zero-knowledge proof, which promotes the auditability of incoming data and prevent invalid or malicious data to enter the blockchain components.

We expect our work to raise awareness on end-to-end encryption, including key pair generation, as most related work propose a centralized agent to issue key pairs which, by design, expose patient's private key to unauthorized access. End-to-end encryption with HE can support data analysis on encrypted data and support, at the same time, decision-making and privacy protection, which can motivate the sharing of public interest health information, such as pandemic data, without exposing individuals.

With a well-structure technical model for on-chain EHR processing, the MEPCA model has the potential of becoming a reference in terms of applying cutting-edge cryptography and distributed network technologies to real use cases in healthcare, with a bottom-up strategy, where key

requirements from stakeholders can drive technology adoption. The adoption of the MEPCA model can bring an evolution to the healthcare industry, allowing society and institutions to have a more secure and efficient digital infrastructure for EHR.

6.2 Publications

We published three articles and one book chapter in the course of the development of the MEPCA model. The first entitled "A blockchain-based end-to-end data protection model for personal health records sharing: a fully homomorphic encryption approach" (VANIN et al., 2023), published in 2023 by the Multidisciplinary Digital Publishing Institute the Sensors Journal, in an special issue for Internet of Health Things. The article was produced in collaboration with the Instituto Colaborativo de Blockchain and with Hospital de Clínicas de Porto Alegre, proposing a model for end-to-end PHR protection, with the adoption of FHE for on-chain data analysis.

Two articles were published in partnership with the Auto-ID Lab from the Korea Advanced Institute of Science and Technology (KAIST). Both articles explore advancements with the use of blockchain technology for version 2.0 of the Electronic Product Code Information Services (EPCIS) and the Core Business Vocabulary (CBV). The article "Decentralized Ledger Technology for EPCIS 2.0: Utilizing NFTs for Enhanced Product Traceability" (VANIN et al., 2024) was presented at the 7th IEEE International Conference on Blockchain in Copenhagen, Denmark. The article "Enhancing Supply Chain Security and Interoperability with GS1 ISO EPCIS/CBV Open Standards using Decentralized Ledgers" (TOLCHA et al., 2024) was presented at the IEEE Global Blockchain Conference in Shanghai, China. All articles were published in 2024.

The chapter "Internet of Things and Machine Learning for Smart Healthcare" was published in the book "IoT and ML for Information Management: A Smart Healthcare Perspective" (NAMASUDRA, 2024). Covers the applicability of multiple technologies for Smart Healthcare, including blockchain technologies, and the challenges related to its adoption. The book was published in 2024.

6.3 Limitations and Future Work

Our work focuses on the application of blockchain and cryptography techniques in the healthcare segment, introducing new elements to separate data management responsibilities, and improving on-chain data processing. To accomplish this, we introduce an element called Data Steward, responsible for managing PHR outside the health institution's environment, and an element called Shared Data Vault to manage temporary access to fulfill requests. We also introduced a novel model to allow the validation of HL7 FHIR data. In this chapter, we describe limitations and future work in the MEPCA model.

As Data Steward is not a traditional element in the health industry, companies need an incentive to provide such a service. Patients must also enroll in a Data Steward to share their data with health institutions. It could incur costs to patients or a demand to share their data as a way to cover the infrastructure costs.

Another limitation is that Data Stewards must be part of each specific consortium to exchange data from patients with health institutions. It could result in situations where a given patient needs to share their data with a health institution, but the Data Steward is not part of the same consortium as the institution.

The data scope considers only numeric data in a format that supports the Homomorphic Encryption calculations. It does not consider more complex data formats and standards like HL7/FHIR. Data in this format are still suitable for sharing in encrypted format, but not for calculations.

Some measures might positively affect the design for a better performance, such as the use of physical servers instead of containerized ones. Hyperledger Fabric also perceives better performance with Go-based chaincodes, instead of Javascript. However, a design with Wide Area Network (WAN) communication can experience higher latency than the test results.

For future work, we recommend expanding the MEPCA model to more use cases in health care, with the provision of data governance resources. The model can also provide programming libraries to help developers create a solution in accordance with the five principles of the model. The model can also provide a set of assessment criteria to help stakeholders measure the maturity level of their solution according to the design principles of the model.

Extending the model to work with IoHT devices can provide more scalable and secure solutions, with the adoption of Decentralized Physical Infrastructure (DePin) techniques, to use smart devices as the infrastructure for on-chain data processing and validation. As more heterogeneous devices and data formats come to market, it increases the demand for data protection, semantic interoperability, and data analysis, thus the MEPCA model can be extended to cover this scenario with more design elements and technical building blocks.

The Hash Proof algorithm can be improved to provide complete on-chain support JSON schema validation rules, extending models such as (ATTOUCHE et al., 2021) to the healthcare industry. The addition of proofs to encrypted data is also a future work for the MEPCA model, as it can increase transparency and relevance of blockchain components in the segment.

BIBLIOGRAPHY

- ALABDULATIF, A.; KHALIL, I.; YI, X. Towards secure big data analytic for cloud-enabled applications with fully homomorphic encryption. **Journal of Parallel and Distributed Computing**, [S.l.], v. 137, p. 192–204, 3 2020.
- ALI, A.; PASHA, M. F.; ALI, J.; FANG, O. H.; MASUD, M.; JURCUT, A. D.; ALZAIN, M. A. Deep learning based homomorphic secure search-able encryption for keyword search in blockchain healthcare system: a novel approach to cryptography. **Sensors**, [S.l.], v. 22, n. 2, p. 528, 2022.
- ALLOGHANI, M.; ALANI, M. M.; AL-JUMEILY, D.; BAKER, T.; MUSTAFINA, J.; HUSSAIN, A.; ALJAAF, A. J. A systematic review on the status and progress of homomorphic encryption technologies. **Journal of Information Security and Applications**, [S.l.], v. 48, 10 2019.
- AOKI, Y.; OTSUKI, K.; KANEKO, T.; BANNO, R.; SHUDO, K. Simblock: a blockchain network simulator. In: IEEE INFOCOM 2019-IEEE CONFERENCE ON COMPUTER COMMUNICATIONS WORKSHOPS (INFOCOM WKSHPS), 2019. **Anais...** [S.l.: s.n.], 2019. p. 325–329.
- ARCHER, N.; FEVRIER-THOMAS, U.; LOKKER, C.; MCKIBBON, K. A.; STRAUS, S. E. Personal health records: a scoping review. **Journal of the American Medical Informatics Association**, [S.l.], v. 18, n. 4, p. 515–522, 2011.
- ATTOUCHE, L.; BAAZIZI, M.-A.; COLAZZO, D.; FALLENI, F.; GHELLI, G.; LANDI, C.; SARTIANI, C.; SCHERZINGER, S. A tool for JSON schema witness generation. In: INTERNATIONAL CONFERENCE ON EXTENDING DATABASE TECHNOLOGY, 24., 2021. **Anais...** [S.l.: s.n.], 2021. p. 694–697.
- BENNACER, S. A.; SABIRI, K.; AAROUD, A.; AKODADI, K.; CHERRADI, B. A comprehensive survey on blockchain-based healthcare industry: applications and challenges. **Indones. J. Electr. Eng. Comput. Sci.**, [S.l.], v. 30, n. 3, p. 1558–1571, 2023.
- Centers for Disease Control and Prevention, COVID-19 Response. **COVID-19 Case Surveillance Public Data Access, Summary, and Limitations**.
<https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Publi>
 Last accessed on 2021-03.
- CHEN, Y.; CHEN, H.; ZHANG, Y.; HAN, M.; SIDDULA, M.; CAI, Z. A survey on blockchain systems: attacks, defenses, and privacy preservation. **High-Confidence Computing**, [S.l.], v. 2, n. 2, p. 100048, 2022.
- CHUKWU, E.; GARG, L. A systematic review of blockchain in healthcare: frameworks, prototypes, and implementations. **IEEE Access**, [S.l.], v. 8, p. 21196–21214, 2020.
- DA COSTA, C. A.; PASLUOSTA, C. F.; ESKOFIER, B.; DA SILVA, D. B.; ROSA RIGHI, R. da. Internet of Health Things: toward intelligent vital signs monitoring in hospital wards. **Artificial intelligence in medicine**, [S.l.], v. 89, p. 61–69, 2018.

FAN, J.; VERCAUTEREN, F. Somewhat practical fully homomorphic encryption. **IACR Cryptol. ePrint Arch.**, [S.l.], v. 2012, p. 144, 2012.

FINCK, M. **Blockchain and the General Data Protection Regulation**: can distributed ledgers be squared with european data protection law? [S.l.]: European Parliamentary Research Service (EPRS), 2019. Study. (PE 634.445).

GEBREMEDHIN, T. A. **Blockchain as a Technology to Facilitate Privacy and Better Health Record Management**. 2018. Dissertação (Mestrado em Ciência da Computação) — The University of Bergen, 2018.

GENTRY, C. Fully homomorphic encryption using ideal lattices. In: ACM SYMPOSIUM ON THEORY OF COMPUTING, 2009. **Proceedings...** [S.l.: s.n.], 2009. p. 169–178.

GHADAMYARI, M.; SAMET, S. Privacy-Preserving Statistical Analysis of Health Data Using Paillier Homomorphic Encryption and Permissioned Blockchain. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 5474–5479.

GHANI, A.; ZINEDINE, A.; EL MOHAJIR, M. A Blockchain-based secure PHR data storage and sharing framework. In: IEEE CONGRESS ON INFORMATION SCIENCE AND TECHNOLOGY (CIST), 2020., 2021. **Anais...** [S.l.: s.n.], 2021. p. 162–166.

GROUP, H. L. S. I. S. W. **Metadata and data - FHIR for FAIR - FHIR Implementation Guide v1.0.0**. Accessed: 2024-08-24, <https://build.fhir.org/ig/HL7/fhir-for-fair/metadata.html>.

Health Level Seven International. **FHIR Release 4.0.1**. [S.l.: s.n.], 2019. Accessed: 2024-08-12.

Health Level Seven International. **Validation - FHIR v5.0.0**. Accessed: 2024-09-10.

HIPAA. **Health Insurance Portability and Accountability Act of 1996**. 1996.

HYPERLEDGER. **Hyperledger Fabric**. Accessed: 2024-09-11, <https://www.hyperledger.org/projects/fabric>.

HYPERLEDGER. **Hyperledger Caliper**. Accessed: 2024-09-11, <https://hyperledger.github.io/caliper/>.

IPFS Distributed Hash Table. **Distributed Hash Tables (DHT)**. <https://docs.ipfs.io/concepts/dht/>, Last accessed on 2022-03-6.

IPFS Kademlia. **IPFS Kademlia Algorithm**. <https://docs.ipfs.io/concepts/dht/kademlia>, Last accessed on 2022-03-6.

ISO, H. I. **Capacity-Based ehealth Architecture Roadmap – Part 2**: architectural components and maturity model, technical report (iso/tr tr14639-2). 2021.

JIANG, Y.; NOGUCHI, T.; KANNO, N.; YASUMURA, Y.; SUZUKI, T.; ISHIMAKI, Y.; YAMANA, H. A Privacy-Preserving Query System using Fully Homomorphic Encryption with Real-World Implementation for Medicine-Side Effect Search. In: INTERNATIONAL CONFERENCE ON INFORMATION INTEGRATION AND WEB-BASED APPLICATIONS & SERVICES, 21., 2020, New York, NY, USA. **Proceedings...** Association for Computing Machinery, 2020. p. 63–72. (iiWAS2019).

KELSEY, J.; SCHNEIER, B. Second preimages on n -bit hash functions for much less than 2^n work. In: ADVANCES IN CRYPTOLOGY–EUROCRYPT 2005: 24TH ANNUAL INTERNATIONAL CONFERENCE ON THE THEORY AND APPLICATIONS OF CRYPTOGRAPHIC TECHNIQUES, AARHUS, DENMARK, MAY 22-26, 2005. PROCEEDINGS 24, 2005. **Anais...** [S.l.: s.n.], 2005. p. 474–490.

KIM, J. W.; KIM, S. J.; CHA, W. C.; KIM, T. A Blockchain-Applied Personal Health Record Application: development and user experience. **Applied Sciences**, [S.l.], v. 12, n. 4, p. 1847, 2022.

KRYSZYN, J.; SMOLIK, W.; WANTA, D.; MIDURA, M.; WRÓBLEWSKI, P. Comparison of OpenEHR and HL7 FHIR Standards. **International Journal of Electronics and Telecommunications**, [S.l.], v. 69, n. 1, 2023.

LABS, P. **IPFS - InterPlanetary File System**. Accessed: 2024-09-12, <https://ipfs.tech>.

MADINE, M. M.; BATTAH, A. A.; YAQOOB, I.; SALAH, K.; JAYARAMAN, R.; AL-HAMMADI, Y.; PESIC, S.; ELLAHHAM, S. Blockchain for giving patients control over their medical records. **IEEE Access**, [S.l.], v. 8, p. 193102–193115, 2020.

MADINE, M. M.; SALAH, K.; JAYARAMAN, R.; YAQOOB, I.; AL-HAMMADI, Y.; ELLAHHAM, S.; CALYAM, P. Fully decentralized multi-party consent management for secure sharing of patient health records. **IEEE Access**, [S.l.], v. 8, p. 225777–225791, 2020.

MAHDY, M. M. Semi-Centralized Blockchain Based Distributed System for Secure and Private Sharing of Electronic Health Records. In: INTERNATIONAL CONFERENCE ON COMPUTER, CONTROL, ELECTRICAL, AND ELECTRONICS ENGINEERING (ICCCEEE), 2020., 2020. **Anais...** [S.l.: s.n.], 2020. p. 1–4.

MAYMOUNKOV, P.; MAZIERES, D. Kademlia: a peer-to-peer information system based on the xor metric. In: INTERNATIONAL WORKSHOP ON PEER-TO-PEER SYSTEMS, 2002. **Anais...** [S.l.: s.n.], 2002. p. 53–65.

MELLO, B. H. de; RIGO, S. J.; COSTA, C. A. da; ROSA RIGHI, R. da; DONIDA, B.; BEZ, M. R.; SCHUNKE, L. C. Semantic interoperability in health records standards: a systematic literature review. **Health and Technology**, [S.l.], p. 1–18, 2022.

MHIRI, S.; EGIO, A.; COMPASTIÉ, M.; COSIO, P. Proxy Re-Encryption for Enhanced Data Security in Healthcare: a practical implementation. In: INTERNATIONAL CONFERENCE ON AVAILABILITY, RELIABILITY AND SECURITY, 19., 2024. **Proceedings...** [S.l.: s.n.], 2024. p. 1–11.

MICROSOFT SEAL (release 3.6). Microsoft Research, Redmond, WA., <https://github.com/Microsoft/SEAL>.

MISBHAUDDIN, M.; ALABDULATHEAM, A.; ALOUFI, M.; AL-HAJJI, H.; ALGHUWAINEM, A. MedAccess: a scalable architecture for blockchain-based health record management. In: INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION SCIENCES (ICCIS), 2020., 2020. **Anais...** [S.l.: s.n.], 2020. p. 1–5.

- MUBASHAR, A.; ASGHAR, K.; JAVED, A. R.; RIZWAN, M.; SRIVASTAVA, G.; GADEKALLU, T. R.; WANG, D.; SHABBIR, M. Storage and proximity management for centralized personal health records using an IPFS-based optimization algorithm. **Journal of Circuits, Systems and Computers**, [S.l.], p. 2250010, 2021.
- NAEHRIG, M.; LAUTER, K.; VAIKUNTANATHAN, V. Can homomorphic encryption be practical? In: ACM WORKSHOP ON CLOUD COMPUTING SECURITY WORKSHOP, 3., 2011. **Proceedings...** [S.l.: s.n.], 2011. p. 113–124.
- NAKAMOTO, S. Bitcoin: a peer-to-peer electronic cash system. **Decentralized Business Review**, [S.l.], p. 21260, 2008.
- NAMASUDRA, S. **IoT and ML for Information Management**: a smart healthcare perspective. [S.l.]: Springer, 2024.
- NG, W. Y.; TAN, T.-E.; MOVVA, P. V.; FANG, A. H. S.; YEO, K.-K.; HO, D.; SAN FOO, F. S.; XIAO, Z.; SUN, K.; WONG, T. Y. et al. Blockchain applications in health care for COVID-19 and beyond: a systematic review. **The Lancet Digital Health**, [S.l.], 2021.
- NIU, S.; CHEN, L.; WANG, J.; YU, F. Electronic health record sharing scheme with searchable attribute-based encryption on blockchain. **IEEE Access**, [S.l.], v. 8, p. 7195–7204, 2019.
- OPENEHR. **openehr - An Open Domain-Driven Platform for Developing Flexible e-Health Systems**. <https://www.openehr.org/>.
- PAIK, H.-Y.; XU, X.; BANDARA, H. D.; LEE, S. U.; LO, S. K. Analysis of Data Management in Blockchain-Based Systems: from architecture to governance. **IEEE Access**, [S.l.], v. 7, p. 186091–186107, 2019.
- PERERA, S.; NANAYAKKARA, S.; RODRIGO, M.; SENARATNE, S.; WEINAND, R. Blockchain technology: is it hype or real in the construction industry? **Journal of Industrial Information Integration**, [S.l.], v. 17, p. 100125, 2020.
- REEGU, F. A.; ABAS, H.; GULZAR, Y.; XIN, Q.; ALWAN, A. A.; JABBARI, A.; SONKAMBLE, R. G.; DZIYAUDDIN, R. A. Blockchain-based framework for interoperable electronic health records for an improved healthcare system. **Sustainability**, [S.l.], v. 15, n. 8, p. 6337, 2023.
- REPÚBLICA, P. da. **Lei Geral de Proteção de Dados Pessoais**. Lei No. 13.709, de 14 de agosto de 2018.
- RIMOL, M. **Gartner Identifies Key Emerging Technologies Spurring Innovation Through Trust, Growth and Change**. 2021.
- ROCHA, V.; LÓPEZ, J.; FALCÃO DA ROCHA, V. **An Overview on Homomorphic Encryption Algorithms**. [S.l.]: Go to reference in article, 2019.
- ROEHRS, A.; COSTA, C. A. da; ROSA RIGHI, R. da. OmniPHR: a distributed architecture model to integrate personal health records. **Journal of biomedical informatics**, [S.l.], v. 71, p. 70–81, 2017.

ROEHRS, A.; COSTA, C. A. da; ROSA RIGHI, R. da; SILVA, V. F. da; GOLDIM, J. R.; SCHMIDT, D. C. Analyzing the performance of a blockchain-based personal health record implementation. **Journal of biomedical informatics**, [S.l.], p. 103140, 2019.

ROEHRS, A.; DA COSTA, C. A.; ROSA RIGHI, R. da; DE OLIVEIRA, K. S. F. Personal health records: a systematic literature review. **Journal of medical Internet research**, [S.l.], v. 19, n. 1, p. e13, 2017.

SARIPALLE, R.; RUNYAN, C.; RUSSELL, M. Using HL7 FHIR to achieve interoperability in patient health record. **Journal of biomedical informatics**, [S.l.], v. 94, p. 103188, 2019.

SENTAUSA, D.; HAREVA, D. H. Exploring Interoperability and Decentralized Applications for Personal Health Data using FHIR Standards. In: INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY: IOT AND SMART CITY, 2023., 2023. **Proceedings...** [S.l.: s.n.], 2023. p. 18–24.

SETYAWAN, R.; HIDAYANTO, A. N.; SENSUSE, D. I.; SURYONO, R. R.; ABILOWO, K. et al. Data integration and interoperability problems of HL7 FHIR implementation and potential solutions: a systematic literature review. In: INTERNATIONAL CONFERENCE ON INFORMATICS AND COMPUTATIONAL SCIENCES (ICICOS), 2021., 2021. **Anais...** [S.l.: s.n.], 2021. p. 293–298.

SHARMA, B.; HALDER, R.; SINGH, J. Blockchain-based interoperable healthcare using zero-knowledge proofs and proxy re-encryption. In: INTERNATIONAL CONFERENCE ON COMMUNICATION SYSTEMS & NETWORKS (COMSNETS), 2020., 2020. **Anais...** [S.l.: s.n.], 2020. p. 1–6.

SHE, W.; GU, Z. H.; LYU, X. K.; LIU, Q.; TIAN, Z.; LIU, W. Homomorphic Consortium Blockchain for Smart Home System Sensitive Data Privacy Preserving. **IEEE Access**, [S.l.], v. 7, p. 62058–62070, 2019.

SHUAIB, M.; ALAM, S.; ALAM, M. S.; NASIR, M. S. Compliance with HIPAA and GDPR in blockchain-based electronic health record. **Materials Today: Proceedings**, [S.l.], 2021.

SONKAMBLE, R. G.; PHANSALKAR, S. P.; POTDAR, V. M.; BONGALE, A. M. Survey of Interoperability in Electronic Health Records Management and Proposed Blockchain Based Framework: myblockehr. **IEEE Access**, [S.l.], 2021.

STOICA, I.; MORRIS, R.; KARGER, D.; KAASHOEK, M. F.; BALAKRISHNAN, H. Chord: a scalable peer-to-peer lookup service for internet applications. In: CONFERENCE ON APPLICATIONS, TECHNOLOGIES, ARCHITECTURES, AND PROTOCOLS FOR COMPUTER COMMUNICATIONS, 2001., 2001. **Proceedings...** [S.l.: s.n.], 2001. p. 149–160.

SUN, J.; YAO, X.; WANG, S.; WU, Y. Blockchain-based secure storage and access scheme for electronic medical records in IPFS. **IEEE Access**, [S.l.], v. 8, p. 59389–59401, 2020.

TAHERDOOST, H. Blockchain and Healthcare: a critical analysis of progress and challenges in the last five years. **Blockchains**, [S.l.], v. 1, n. 2, p. 73–89, 2023.

TOLCHA, Y. K.; VANIN, F. N. d. S.; RIGHI, R.; COSTA, C. A.; KIM, D. Enhancing Supply Chain Security and Interoperability with GS1 ISO EPCIS/CBV Open Standards using

Decentralized Ledgers. In: IEEE GLOBAL BLOCKCHAIN CONFERENCE, 2024. **Anais...** [S.l.: s.n.], 2024.

UNION, E. **Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)**. Accessed: 2024-08-12, Online at <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

VANIN, F. N. d. S.; POLICARPO, L. M.; RIGHI, R. d. R.; HECK, S. M.; SILVA, V. F. da; GOLDIM, J.; COSTA, C. A. da. A blockchain-based end-to-end data protection model for personal health records sharing: a fully homomorphic encryption approach. **Sensors**, [S.l.], v. 23, n. 1, p. 14, 2023.

VANIN, F. N. d. S.; TOLCHA, Y. K.; RIGHI, R.; COSTA, C. A.; KIM, D. Decentralized Ledger Technology for EPCIS 2.0: utilizing nfts for enhanced product traceability. In: THE 7TH IEEE INTERNATIONAL CONFERENCE ON BLOCKCHAIN, 2024. **Anais...** [S.l.: s.n.], 2024.

WALONOSKI, J.; KRAMER, M.; NICHOLS, J.; QUINA, A.; MOESEL, C.; HALL, D.; DUFFETT, C.; DUBE, K.; GALLAGHER, T.; MCLACHLAN, S. Synthea: an approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. **Journal of the American Medical Informatics Association**, [S.l.], v. 25, n. 3, p. 230–238, 2018.

WANG, Y.; ZHANG, A.; ZHANG, P.; QU, Y.; YU, S. Security-Aware and Privacy-Preserving Personal Health Record Sharing using Consortium Blockchain. **IEEE Internet of Things Journal**, [S.l.], 2021.

WOOD, A.; NAJARIAN, K.; KAHROBAEI, D. Homomorphic Encryption for Machine Learning in Medicine and Bioinformatics. **ACM Comput. Surv.**, New York, NY, USA, v. 53, n. 4, aug 2020.

YANG, C.-N.; LI, P.; CHENG, H.-H.; KUO, H.-C.; LU, M.-C.; XIONG, L. A Security Model of Multihospital FHIR Database Authorization Based on Secret Sharing and Blockchain. **IEEE Internet of Things Journal**, [S.l.], 2023.

YAQOOB, I.; SALAH, K.; JAYARAMAN, R.; AL-HAMMADI, Y. Blockchain for healthcare data management: opportunities, challenges, and future recommendations. **Neural Computing and Applications**, [S.l.], p. 1–16, 2022.

YAZDINEJAD, A.; SRIVASTAVA, G.; PARIZI, R. M.; DEGHANTANHA, A.; CHOO, K.-K. R.; ALEDHARI, M. Decentralized Authentication of Distributed Patients in Hospital Networks Using Blockchain. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], v. 24, n. 8, p. 2146–2156, 2020.

ZHANG, J.; YANG, Y.; LIU, X.; MA, J. An efficient blockchain-based hierarchical data sharing for Healthcare Internet of Things. **IEEE Transactions on Industrial Informatics**, [S.l.], 2022.

ZHANG, P.; WHITE, J.; SCHMIDT, D. C.; LENZ, G.; ROSENBLOOM, S. T. FHIRChain: applying blockchain to securely and scalably share clinical data. **Computational and structural biotechnology journal**, [S.l.], v. 16, p. 267–278, 2018.

ZHUANG, Y.; SHEETS, L. R.; CHEN, Y.-W.; SHAE, Z.-Y.; TSAI, J. J.; SHYU, C.-R. A Patient-Centric Health Information Exchange Framework Using Blockchain Technology. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], v. 24, n. 8, p. 2169–2176, 2020.