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

Fausto Neri da Silva Vanin[1], **Rodrigo da Rosa Righi**[1], **Cristiano André da Costa**[1]

[1]SOFTWARELAB – Universidade do Vale do Rio dos Sinos (Unisinos)
São Leopoldo 93022-000

{faustovanin,rrrighi,cac}@unisinos.br

Abstract. *Blockchain technology in healthcare is gaining attention for addressing data privacy, interoperability, and health record integrity issues. Standards like HL7 FHIR and OpenEHR ensure data consistency, but privacy concerns persist under regulations like HIPAA, GDPR, and LGPD. Existing methods often store only data hashes, raising validation risks. The MEPCA model introduces a blockchain-based framework for secure health record management, focusing on on-chain EHR data processing. Key elements include Data Steward, Shared Data Vault, and Zero-Knowledge Proofs of HL7 FHIR fields. Experiments with Fully Homomorphic Encryption show enhanced security and reliability for health records, offering a robust alternative to traditional off-chain approaches.*

1. Introduction

The adoption of blockchain solutions for Electronic Health Records (EHR) recently gained significant attention in the scientific community and in the healthcare industry, 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, interoperability, and could be very helpful in scenarios such as the COVID-19 pandemics, medical research, counterfeit prevention, and management of medical supply chains [Taherdoost 2023].

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), or Distributed Hash Table (DHT), over the on-chain alternative, frequently sending only a hash representation of a given input data to the blockchain. 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.

In this work, 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. The model presents an algorithm for on-chain hash validation of HL7 FHIR data, based on Merkle Trees and Zero-Knowledge Proofs. We evaluated the proposed model using a data set of 10,000 registries and evaluated the Fully Homomorphic

Encryption calculation on a data set of 1.3M cases from the United States Center for Disease Control and Prevention (CDC).

2. Problem Statement

This work relies on the use of digital approach to health records, having as motivation a context where most data processing occurs off-chain, data is frequently processed in unencrypted format, and there is lack of control from patients over their corresponding data [Perera et al. 2020, Shuaib et al. 2021]. Our work approaches such a context with the given research problem: **How to improve smart healthcare for the exchange and interoperability of EHR, protecting privacy, meeting scalability and regulatory requirements?**. 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.

3. MEPCA Model

We propose the MEPCA model to improve the use of on-chain resources to process EHR data. Our approach aims to improve decision making based on on-chain data, and 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: [M] maximize on chain usage; [E] encrypt data whenever possible; [P] prove if incoming data is valid; [C] comply with regulation; [A] adapt to the multiple use cases.

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. In Figure 1 we present all the main components of the architecture and introduce two elements that allow better segregation of responsibilities regarding patient data and clinic treatment, called Data Steward and Shared Data Vault.

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.

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

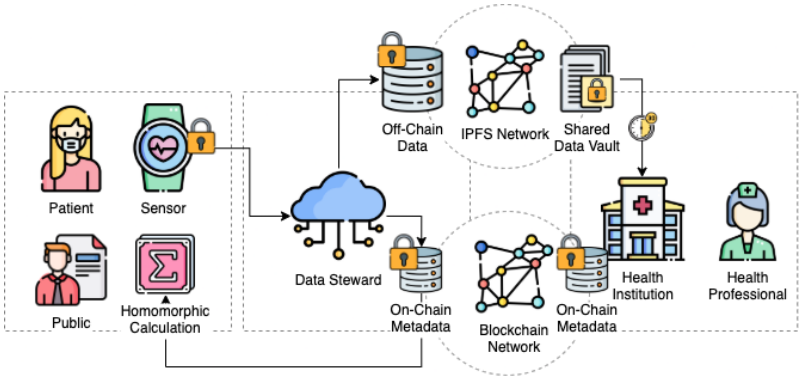

Figure 1. MEPCA model architecture components.

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 [Health Level Seven International 2019] or OpenEHR [openEHR 2020], as each institution runs a different Health Management System (HMS)[de Mello et al. 2022].

We introduce an algorithm called Hash Proof to verify whether a hash digest preimages to a file containing a predefined set of HL7 FHIR required fields, adopting Zero-Knowledge Proofs as a way to verify the incoming hash data. Assume an input HL7 FHIR file J in JSON format with a set of fields $F = \{f_1, f_2, ..., f_n\} \in L$, where L is the language schema that describes J. There is a subset $F' \subset F$ with all k 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 for the selected hash function H that satisfy $H(w, \pi) = h$, having:

$$M \leftarrow H(\{F' \cup \phi, (F - F'), V\}) \tag{1}$$

, where M is a Sparse Merkle Tree, ϕ is a set of NULL values with $\lfloor 2^{\sqrt{k}+1} - k \rfloor$ elements, M_1 is the leftmost sub-tree of M, with leaf nodes $\in F' \cup \phi$, w is the root of M_1, V is the set of fields in J, Π_f is the Merkle Proof of $f \in F'$ in M, and π is the proof of h in the form: $\Pi_f - M_1$.

4. Experiments and Results

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. We createed a test data set with 10 thousand synthetic patients using Synthea [Walonoski et al. 2018]. For testing, we used Hyperledger Caliper with the following configuration: Fixed Rate (100 TPS read, 100 TPS write), Fixed Load (40 transaction load and 100 TPS write, 500 transaction load and 1,000 TPS read), Maximum Rate (100 starting TPS, 5s TPS increasing interval, and 20s sample interval write, 1,000 starting TPS, 5s TPS increasing interval, and 20s sample interval read).

After three rounds of testing with each configuration scenario, we collected results for read and write process separately for each approach, comparing results for Transac-

tions per Second (TPS), latency, error rate and pending transactions in Table 1. Fixed Rate write reached highest TPS and also the highest latency, with an average of 40 pending 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 1. Evaluation results for data write comparing Fixed Rate, Fixed Load, and Maximum Rate strategies.

	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 2 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.

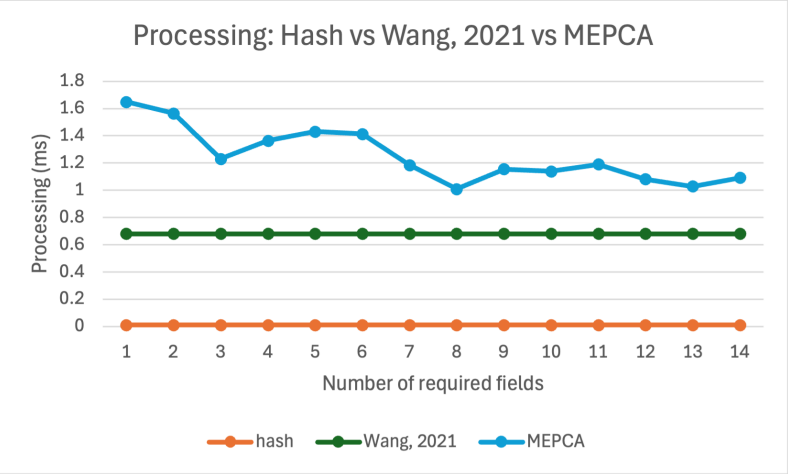

Figure 2. Hash processing in milliseconds compared to [Wang et al. 2021] and MEPCA. As the number of required fields increases, the processing time reduces.

For FHE, we selected an open data set from the Centers for Disease Control and Prevention[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. We used the SEAL library, which implements the BFV (Brakerski, Fan, and Vercauteren) algorithm for FHE. To calculate block propagation time, we used a Blockchain Network Simulation tool called Simblock [Aoki et al. 2019]. Also, all experiments were run on a 3.2 MHz 8 cores computer with 16 GB RAM.

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

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

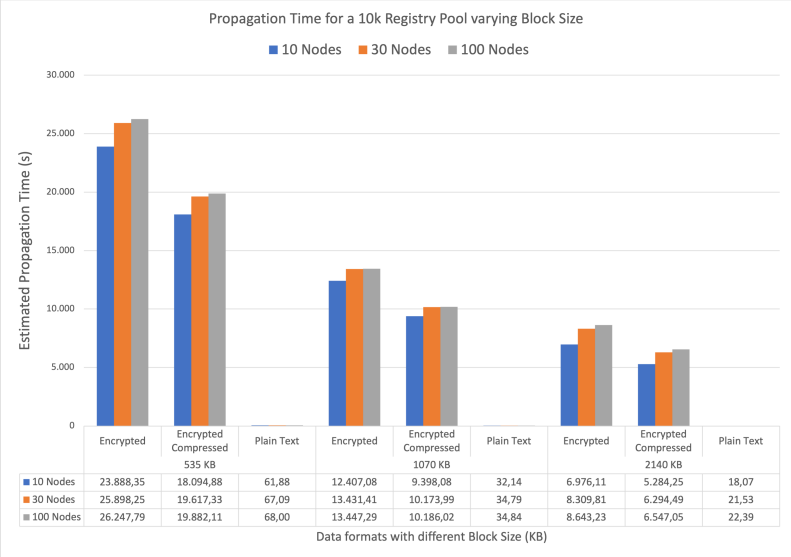

Figure 3. Blockchain Network Simulation results varying block size, data format and network size.

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. 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 $\approx 24\%$, which could reduce about 1.7GB of space consumption. Each simulation round generated 100 blocks with each corresponding propagation time in seconds. In Figure 3 we demonstrate the results for each scenario.

5. 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.

The MEPCA model contributes to map 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. 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. 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 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 FHE 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.

References

Aoki, Y., Otsuki, K., Kaneko, T., Banno, R., and Shudo, K. (2019). Simblock: A blockchain network simulator. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 325–329. IEEE.

Chen, Y., Chen, H., Zhang, Y., Han, M., Siddula, M., and Cai, Z. (2022). A survey on blockchain systems: Attacks, defenses, and privacy preservation. *High-Confidence Computing*, 2(2):100048.

de Mello, B. H., Rigo, S. J., da Costa, C. A., da Rosa Righi, R., Donida, B., Bez, M. R., and Schunke, L. C. (2022). Semantic interoperability in health records standards: a systematic literature review. *Health and Technology*, pages 1–18.

Health Level Seven International (2019). *FHIR Release 4.0.1*. Accessed: 2024-08-12.

Jiang, Y., Noguchi, T., Kanno, N., Yasumura, Y., Suzuki, T., Ishimaki, Y., and Yamana, H. (2020). A privacy-preserving query system using fully homomorphic encryption with real-world implementation for medicine-side effect search. In *Proceedings of the 21st International Conference on Information Integration and Web-Based Applications & Services*, iiWAS2019, page 63–72, New York, NY, USA. Association for Computing Machinery.

Misbhauddin, M., AlAbdulatheam, A., Aloufi, M., Al-Hajji, H., and AlGhuwainem, A. (2020). Medaccess: A scalable architecture for blockchain-based health record management. In *2020 2nd International Conference on Computer and Information Sciences (ICCIS)*, pages 1–5. IEEE.

openEHR (2020). openehr - an open domain-driven platform for developing flexible e-health systems. `https://www.openehr.org/`.

Perera, S., Nanayakkara, S., Rodrigo, M., Senaratne, S., and Weinand, R. (2020). Blockchain technology: Is it hype or real in the construction industry? *Journal of Industrial Information Integration*, 17:100125.

Shuaib, M., Alam, S., Alam, M. S., and Nasir, M. S. (2021). Compliance with hipaa and gdpr in blockchain-based electronic health record. *Materials Today: Proceedings*.

Taherdoost, H. (2023). Blockchain and healthcare: A critical analysis of progress and challenges in the last five years. *Blockchains*, 1(2):73–89.

Walonoski, J., Kramer, M., Nichols, J., Quina, A., Moesel, C., Hall, D., Duffett, C., Dube, K., Gallagher, T., and McLachlan, S. (2018). 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*, 25(3):230–238.

Wang, Y., Zhang, A., Zhang, P., Qu, Y., and Yu, S. (2021). Security-aware and privacy-preserving personal health record sharing using consortium blockchain. *IEEE Internet of Things Journal*.