

Blockchain-based Electronic Medical Records (EMR) Management with OCR Assisted Summarization

*Term Paper (Final) submitted in fulfillment of the requirements for the Degree
of*

B. Tech

in

Information Technology

By

Vatsal Gupta

ENROLLMENT NO.

17104060

under the guidance of

Mr. Vikas Hassija



Department of Computer Science Engineering and Information Technology

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

(Declared Deemed to be University U/S 3 of UGC

Act) A-10, SECTOR-62, NOIDA, INDIA

October – 2020

Table of Contents

INTRODUCTION	3
LITERATURE REVIEW	4
Blockchain-based EHR	4
Cloud-assisted secure eHealth Systems for tamper-proofing EHR via blockchain.....	4
Blockchain for Secure EHRs Sharing of Mobile Cloud Based E-Health Systems	5
Blockchain-based electronic healthcare record system for healthcare 4.0 applications	6
Blockchain based searchable encryption for electronic health record sharing	7
Access Control for Electronic Health Records with Hybrid Blockchain-Edge Architecture	8
Blockchain-based Interoperable Electronic Health Record Sharing Framework	9
Ancile: Privacy-preserving Framework for Access Control and Interoperability of Electronic Health Records using Blockchain Technology	10
MedBloc: A Blockchain-based Secure EHR System for Sharing and Accessing Medical Data	11
BHEEM: A Blockchain-based Framework for Securing Electronic Health	12
Scalable Architecture for sharing EHR using the Hyperledger Blockchain	13
OCR for Health Records	14
Cloud-assisted secure eHealth Systems for tamper-proofing EHR via blockchain.....	14
Blockchain for Secure EHRs Sharing of Mobile Cloud Based E-Health Systems	15
GAP ANALYSIS (COMBINED)	16
SIGNIFICANCE OF WORK	16
REFERENCES	17

Summary Framework of Individual Papers

- Authors
- Journal/Conference and Publisher
- Clarivate Analytics – 2020 Impact Factor / Core 2020 Conference Ranking
- Month and Year of Publication
- Novelty
- Advantages of the Proposed Solution
- Limitations of the Proposed Solution
- Critical Review

INTRODUCTION

Nowadays, healthcare systems are very complex and expensive. Also, they lack security of patients' sensitive data. Most of the proposed frameworks lack scalability and interoperability.

Blockchain can transform the way the EHRs (electronic health records) of patients are processed and shared by providing secure frameworks for the health information sharing of medical data in the healthcare industry, by protecting it over a decentralized framework. It provides the patients to have better control over their sensitive data. It enables controlled access of patients' sensitive data.

Blockchain technology has the potential to change health care, it can increase the protection, interoperability and privacy of health data. By making electronic health records (EHRs) more effective and safer, this technology could provide a new model for health information sharing (HIE). Blockchain technology, provides enhanced transparency and communication between patients and healthcare providers.

Managing large number of Electronic records is made feasible using Optical character recognition which helps to identify the key words from the patients' records and makes the job easier for the doctor. It saves a lot of time and also gets the job done accurately.

LITERATURE REVIEW

Blockchain-based EHR

1. Cloud-assisted secure eHealth systems for tamper-proofing EHR via blockchain

Authors: Sheng Cao, Gexiang Zhang, Pengfei Liu, Xiaosong Zhang, Ferrante Neri (corresponding author)

Journal and Publisher: Information Sciences, Elsevier

Clarivate Analytics – 2020 Impact Factor: 5.91

Month and Year of Publication: February, 2019

Novelty

In this paper, the authors propose a cloud-assisted eHealth framework for the management of Electronic Health Records (EHR). The authors have used blockchain as the underlying architecture of their framework which protects outsourced EHRs from illegal modification. A comprehensive security analysis and performance evaluation has also been carried out by the authors to stress on their solution's security guarantee.

Advantages of the Proposed Solution

- Confidentiality
- Resistance against EHR forgery
- Resistance against EHR modification
- Resistance against impersonation attacks performed by external adversaries

Limitations of the Proposed Solution

- Decrease in Quality of Service (QoS)
- Latency takes a hit since even basic operations require more time on the blockchain data structure as compared to centralized architectures
- Complexity of the model

Critical Review

The authors have done a good job of articulating the need for decentralized architecture in the healthcare industry. The security analysis performed by the authors covers all possible corner cases including collusion, forgery, impersonation attacks etc. The only thing missing in this work is the comparison of their model to existing decentralized models for the same problem statement. Also, the paper lacks a comprehensive analysis of the QoS and latency of their framework.

2. Blockchain for Secure EHRs Sharing of Mobile Cloud Based E-Health Systems

Authors: Dinh C. Nguyen (corresponding author), Pubudu N. Pathirana, Ming Ding, Aruna Seneviratne

Journal and Publisher: IEEE Access, IEEE

Clarivate Analytics – 2020 Impact Factor: 3.745

Month and Year of Publication: May, 2019

Novelty

In this work, the authors have proposed a novel framework for EHRs sharing that combines blockchain on a mobile cloud platform with the Decentralized Interplanetary File System (IPFS). In particular, using smart contracts, the authors have designed a trustworthy access control mechanism to achieve safe sharing of EHRs among various patients and medical providers.

Advantages of the Proposed Solution

- Reliable data exchange on mobile clouds
- High data privacy levels- the model preserves sensitive health information against potential threats
- Lightweight access control design
- Minimum network latency

Limitations of the Proposed Solution

- Scalability constraints
- Decreased Quality of Service (QoS)

Critical Review

In this work the authors presented a prototype implementation using Ethereum blockchain on a mobile app with Amazon cloud computing in a real data sharing scenario. The authors have included a detailed comparison of their model with existing models. The authors have also evaluated security claims of their model. However, the authors have modelled only a single smart contract deployment over the network. Deployment of more smart contract with more complex interactions would have given a better insight into the performance metrics of the model.

3. Blockchain-based electronic healthcare record system for healthcare 4.0 applications

Authors: Sudeep Tanwar (corresponding author), Karan Parekh, Richard Evans

Journal and Publisher: Journal of Information Security and Applications, Elsevier

Clarivate Analytics – 2020 Impact Factor: 2.327

Month and Year of Publication: February, 2020

Novelty

This paper discusses many solutions, including frameworks and tools to assess the efficiency of systems, such as Hyperledger Fabric, Composer, Docker Container, Hyperledger Caliper, and the Wireshark capture engine, to enhance current shortcomings in healthcare systems using blockchain technology. Furthermore this paper proposes an Access Control Policy Algorithm to improve data accessibility among healthcare providers, helping to simulate environments to enforce the sharing framework of the Hyperledger-based electronic healthcare record (EHR) that uses the idea of a chain code.

Advantages of the Proposed Solution

- Need for a central authority eliminated
- Optimized Performance metrics such as Round Trip Time (RTT) , throughput to produce better results.

Limitations of the Proposed Solution

- No security analysis - security claims of the model rest on the claims of the blockchain technology
- Complex model - hard for an average reader to understand and simulate

Critical Review

In this paper, the authors have recognized all the breakthrough technologies in blockchain development and summarized literature present in that direction. The paper also proposes an electronic healthcare record (EHR) sharing system based on Hyperledger with an access control policy algorithm to increase the accessibility of data between healthcare providers. The authors have done a brilliant job of analyzing the latency, throughput, RTT and other performance metrics of their model. The authors, however, rest the claim of their model on blockchain's claims of immutability. The authors should have included a comprehensive security analysis of their model.

4. Blockchain based searchable encryption for electronic health record sharing

Authors: Lanxiang Chen, Wai-Kong Lee, Chin-Chen Chang, Kim-Kwang Raymond Choo (corresponding author), Nan Zhanga

Journal and Publisher: Future Generation Computer Systems, Elsevier

Clarivate Analytics – 2020 Impact Factor: 6.125

Month and Year of Publication: June, 2019

Novelty

The authors suggest a blockchain-based searchable encryption scheme for EHRs in this paper. The index for EHRs is built and stored in the blockchain via complex logic expressions, so that the expressions can be used by a data user to search for the index. The data owners have complete control of who can see their EHR data, as only the index is transferred to the blockchain to allow propagation.

Advantages of the Proposed Solution

- No central authority
- Data owners have total power of who can see data from their EHRs.
- Using blockchain technology guarantees the transparency, anti-tampering, and traceability of the index of EHRs

Limitations of the Proposed Solution

- No security analysis
- Scalability of the model is in question

Critical Review

In this work, the authors have articulated the severity of data leakage in Electronic Health Records (EHRs). To counter this issue, the authors have proposed a blockchain-based network. Various participating medical institutions and individuals (e.g. medical professionals, hospitals, medical laboratories and insurance companies) in this network will access EHRs with a higher degree of trust held on the blockchain. The authors have evaluated the proposed scheme from two aspects, namely in terms of the overhead for extracting the document IDs from EHRs and the overhead associated with conducting transactions on smart contract in Ethereum. In the future, the authors can try to implement and evaluate the proposed scheme in a real-world environment.

5. Access Control for Electronic Health Records with Hybrid Blockchain-Edge Architecture

Authors: Hao Guo, Wanxin Li, Mark Nejad, Chien-Chung Shen

Conference and Publisher: 2019 IEEE International Conference on Blockchain (Blockchain), IEEE

CORE 2020 Conference Ranking: Unranked

Month and Year of Publication: January, 2020

Novelty

In this work the authors proposed a hybrid architecture to enforce attribute-based access control of EHR data by using both blockchain and edge nodes. A blockchain-based controller handles identity and access control policies within the architecture and acts as a tamper-proof record of access events. In addition, off-chain edge nodes store the EHR data and apply policies specified in Abbreviated Language for Authorization (ALFA) to enforce attribute-based access control on EHR data in collaboration with the blockchain-based access control logs.

Advantages of the Proposed Solution

- Eliminates the need for a central authority
- Privacy-preserving by design
- Performance metrics such as response time, transaction processing time have been evaluated

Limitations of the Proposed Solution

- Lack of in-depth analyses of performance and security claims
- The proposed solution is susceptible to collusions

Critical Review

A Hyperledger Composer Fabric blockchain programmed with smart contracts and ACL policies has been developed by the authors to assess the efficiency of access control by measuring the processing time and response time of transactions against unauthorized recovery attempts. The authors claim that their system provides results in terms of milliseconds making it suitable to be incorporated in real-time and secured EHR data access control frameworks. A brief security and performance analysis is also included in the work. The paper, however, fails to articulate the exact problem that it is solving. Furthermore, the paper fails to address the issue of collusion in such blockchain networks.

6. Blockchain-Based Interoperable Electronic Health Record Sharing Framework

Authors: Gracie Carter, Hossain Shahriar, Sweta Sneha

Conference and Publisher: 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), IEEE

CORE 2020 Conference Ranking: B ranked

Month and Year of Publication: July, 2019

Novelty

In this work the authors put forward a new approach to the use of Amazon Web Services and Ethereum blockchain for a blockchain and cloud storage network to promote semantic level interoperability of Electronic Health Records systems without standardized data types and formatting.

Advantages of the Proposed Solution

- Secure data sharing
- Improved interoperability

Limitations of the Proposed Solution

- Many claims made in the work have not been substantiated with results
- No identity management module in the framework

Critical Review

This work proposes the joining of Amazon Web Services and Ethereum blockchain as a possible solution to interoperability in healthcare. According to the authors, blockchain technology could improve interoperability between points of care by allowing for a complete method of data sharing and increased security. Further improvements of the prototype architecture proposed in this paper might include using faster distributed ledgers technologies instead of Ethereum as they become available and the integration of identity management for patients. Proving this proposed solution with other cloud computing platforms should also be explored to encourage adoption.

7. Ancile: Privacy-preserving Framework for Access Control and Interoperability of Electronic Health Records using Blockchain Technology

Authors: Gaby G. Dagher, Jordan Mohler, Matea Milojkovic, Praneeth Babu Marella

Journal and Publisher: Sustainable Cities and Society, Elsevier

Clarivate Analytics – 2020 Impact Factor: 5.268

Month and Year of Publication: May, 2018

Novelty

In this work, authors have proposed a blockchain-based system for patients, suppliers, and third parties to provide secure, interoperable, and efficient access to medical records while protecting the privacy of confidential information for patients. The proposed platform, called Ancile, utilizes smart contracts for enhanced access control and data obfuscation in an Ethereum-based blockchain, and employs advanced cryptographic techniques for further protection.

Advantages of the Proposed Solution

- Addressed Longstanding privacy
- Secure transfer of records
- Cost and storage effective

Limitations of the Proposed Solution

- Large data cannot be stored effectively
- A collusion of 51% of mining nodes on the system could result in the rewriting of the chain structure

Critical Review

Ancile, the proposed platform, revealed that it would be extremely difficult to fully hide all information and retain an open and interoperable system. However, Ancile still provides substantial privacy preservation and data integrity by using smart contracts to isolate information. However, in the Ethereum community, many of the technologies used in this system, such as smart contracts and authorized blockchains, are still in the early stages of their growth. Hence, the system is heavily dependent on their performance.

8. MedBloc: A Blockchain-based Secure EHR System for Sharing and Accessing Medical Data

Authors: Jack Huang, Yuan Wei Qi, Muhammad Rizwan Asghar, Andrew Meads, Yu-Cheng Tu

Conference and Publisher: 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications (TrustCom)

CORE 2020 Conference Ranking: A ranked

Month and Year of Publication: August, 2019

Novelty

The authors present MedBloc in this work, a stable EHR blockchain-based the framework that allows patients to access and exchange information with healthcare providers, in a practical yet privacy-preserving way. MedBloc uses an encryption scheme to secure medical data.

Advantages of the Proposed Solution

- Secure data sharing
- Token based permission model

Limitations of the Proposed Solution

- Time consuming framework
- Huge data storage required

Critical Review

MedBloc is a fully built, people-powered, collaborative EHR system that enables patients and healthcare professionals via dedicated customer service to access and exchange health records easily. Patients may give and revoke consent at any time with the use of smart contracts and cryptographic techniques. Healthcare providers can request approval and add records that are securely encrypted and stored on the blockchain. MedBloc also preserves the privacy of patients by using a sophisticated encryption system.

9. BHEEM: A Blockchain-based Framework for Securing Electronic Health

Authors: Jayneel Vora, Anand Nayyar, Sudeep Tanwar, Sudhanshu Tyagi, Neeraj Kumar, M. S. Obaidat, Joel J P C Rodrigues

Conference and Publisher: 2018 IEEE Global Communications Conference (GLOBECOM)

CORE 2020 Conference Ranking: B ranked

Month and Year of Publication: December, 2018

Novelty

In this work the authors propose a Blockchain-based system for the efficient storage and management of EHRs that further provides patients, clinicians, and third parties with safe and efficient access to medical data while protecting private patient information. The study analyzes how the proposed structure addresses the needs of patients, providers and third parties, and how privacy and security issues in the healthcare sector are maintained by the framework.

Advantages of the Proposed Solution

- Balance maintained between privacy and accessibility of EHRs.
- Efficient storage of EHRs.

Limitations of the Proposed Solution

- Lack of correlation between the size and noise of blockchain
- Lack of interoperability

Critical Review

This work concludes that it would be very unlikely to fully hide all information and retain an open and interoperable system, but the proposed architecture still provides substantial privacy preservation and data integrity by using smart contracts to isolate information. They should have explored the practicality of the proposed model and should have established the correlation between the size and noise of blockchain.

10. Scalable Architecture for sharing EHR using the Hyperledger Blockchain

Authors: Andressa Fernandes, Vladimir Rocha, Arlindo F. da Conceição, Flavio Horita

Conference and Publisher: 2020 IEEE International Conference on Software Architecture Companion (ICSA-C)

CORE 2020 Conference Ranking: B ranked

Month and Year of Publication: March, 2020

Novelty

In this work the authors put forward a new approach to a scalable architecture using a multi-channel hyperledger blockchain for exchanging electronic health records. The design uses one blockchain to record patient visits and one blockchain to record connections pointing to Electronic Health Records (EHRs) held in external systems for each health institution.

Advantages of the Proposed Solution

- Scalable architecture

Limitations of the Proposed Solution

- Lack of secure data sharing

Critical Review

The proposed framework is proved to be scalable as compared to the other existing architectures. The architecture proposes to divide the entire Blockchain into several smaller ones, using the global and the local Blockchains. However, the paper does not emphasis upon the security of the data being shared.

OCR for Health Records

1. Importance of Multi-modal Approaches to Effectively Identify Cataract Cases from Electronic Health Records

Authors: Peggy L Peissig, Luke V Rasmussen, Richard L Berg, James G Linneman, Catherine A McCarty, Carol Waudby, Lin Chen, Joshua C Denny, Russell A Wilke, Jyotishman Pathak, David Carrell, Abel N Kho, Justin B Starren

Journal and Publisher: Journal of the American Medical Informatics Association, Oxford Press

Clarivate Analytics – 2020 Impact Factor: 4.112

Month and Year of Publication: March, 2012

Novelty

The work describes the development and validation of an algorithm based on EHR to classify subjects with cataracts associated with age. In order to classify cataract subjects and associated cataracts attributes, the model proposes a multi-modal approach consisting of structured database queries, natural language processing on free-text documents and optical character recognition on scanned clinical images.

Advantages of the Proposed Solution

- High PPV (Positive Predictive Values)
- Portable Algorithm

Limitations of the Proposed Solution

- Not Cost effective
- Lack of validation of hand written notes
- Applied to a single clinical domain area.

Critical Review

A critique of this research may be the use of a domain specialist to identify model development characteristics in the CDM, NLP, and OCR approaches. Also the strategy was applied to only one clinical domain area, cataract. It might not work equally efficiently on other domains.

2. Automatic Classification of Scanned Electronic Health Record Documents

Authors: Heath Goodrum, Kirk Roberts, Elmer V. Bernstam

Journal and Publisher: International Journal of Medical Informatics, Elsevier

Clarivate Analytics – 2020 Impact Factor: 3.025

Month and Year of Publication: December, 2020

Novelty

In this paper the authors demonstrate how scanned documents can be classified accurately by text classification systems. It also describes the design and assessment of a system for classifying documents into categories that are clinically relevant and non-clinically relevant, as well as further subclassifications.

Advantages of the Proposed Solution

- High accuracy

Limitations of the Proposed Solution

- Results may not generalize to other institutions
- Depend on the use of already-classified external documents for training and testing

Critical Review

The authors have done a good job as the proposed architecture provides a accuracy of 0.973. Also, the system was evaluated on three different levels of classification using both the individual pages of the document, as well as the entire document as input. Although it has several limitations as well such as scanned documents were evaluated from a single EHR at one institution. Hence, results may not generalize to other institutions. Also, there might be some errors in both the training and test sets due to the use of already-classified external documents.

GAP ANALYSIS

Extensive research has been done in the healthcare industry in context of blockchain. Several research papers are available proposing various strategies to provide a decentralized, scalable, secure, efficient and integrated system to manage the EHRs (Electronic Health Records), providing controlled access and ensuring safe sharing of data.

Since the idea was coined, the interest in Blockchain technology has increased. Most research focuses on revealing and enhancing Blockchain's constraints from the perspective of privacy and protection, but many of the solutions suggested do not have a clear assessment of their effectiveness. Many other problems linked to Blockchain scalability, including throughput and latency, have been left unstudied. The study, however is niche and needs a more practical approach in this field.

It is undeniable that the efforts in the health sector of using blockchain technology are rising exponentially. There are areas that could possibly be strongly impacted by blockchain technology within the health domain.

SIGNIFICANCE OF WORK

In this work, I have studied and analyzed various research papers from elite publications to summarize the current research on Blockchain in managing EHRs and also the current scenario of research done in Optical character recognition (OCR). I have also implemented a framework to provide a direction of practicality to the research.

REFERENCES

- [1] Cao, S., Zhang, G., Liu, P., Zhang, X., & Neri, F. (2019). Cloud-assisted secure eHealth systems for tamper-proofing EHR via blockchain. *Information Sciences*, 485, 427-440.
- [2] Nguyen, D. C., Pathirana, P. N., Ding, M., & Seneviratne, A. (2019). Blockchain for secure ehRs sharing of mobile cloud based e-health systems. *IEEE access*, 7, 66792-66806.
- [3] Tanwar, S., Parekh, K., & Evans, R. (2020). Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *Journal of Information Security and Applications*, 50, 102407.
- [4] Chen, L., Lee, W. K., Chang, C. C., Choo, K. K. R., & Zhang, N. (2019). Blockchain based searchable encryption for electronic health record sharing. *Future Generation Computer Systems*, 95, 420-429.
- [5] Guo, H., Li, W., Nejad, M., & Shen, C. C. (2019, July). Access control for electronic health records with hybrid blockchain-edge architecture. In *2019 IEEE International Conference on Blockchain (Blockchain)* (pp. 44-51). IEEE.
- [6] Carter, G., Shahriar, H., & Sneha, S. (2019, July). Blockchain-Based Interoperable Electronic Health Record Sharing Framework. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 2, pp. 452-457). IEEE.
- [7] Dagher, G. G., Mohler, J., Milojkovic, M., & Marella, P. B. (2018). Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustainable cities and society*, 39, 283-297.
- [8] Huang, J., Qi, Y. W., Asghar, M. R., Meads, A., & Tu, Y. C. (2019, August). MedBloc: A Blockchain-Based Secure EHR System for Sharing and Accessing Medical Data. In *2019 18th IEEE International Conference On Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference On Big Data Science and Engineering (TrustCom/BigDataSE)* (pp. 594-601). IEEE.
- [9] Vora, J., Nayyar, A., Tanwar, S., Tyagi, S., Kumar, N., Obaidat, M. S., & Rodrigues, J. J. (2018, December). BHEEM: A blockchain-based framework for securing electronic health records. In *2018 IEEE Globecom Workshops (GC Wkshps)* (pp. 1-6). IEEE.
- [10] Fernandes, A., Rocha, V., da Conceição, A. F., & Horita, F. (2020, March). Scalable Architecture for sharing EHR using the Hyperledger Blockchain. In *2020 IEEE International Conference on Software Architecture Companion (ICSA-C)* (pp. 130-138). IEEE.
- [11] Peissig, P. L., Rasmussen, L. V., Berg, R. L., Linneman, J. G., McCarty, C. A., Waudby, C., ... & Carrell, D. (2012). Importance of multi-modal approaches to effectively identify cataract cases from electronic health records. *Journal of the American Medical Informatics Association*, 19(2), 225-234.
- [12] Goodrum, H., Roberts, K., & Bernstam, E. V. (2020). Automatic classification of scanned electronic health record documents. *International Journal of Medical Informatics*, 144, 104302.