INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Preserving the Privacy of the Healthcare, Clinical and Personal Data using Blockchain

P Arul[1], **S Renuka**[2]*

1 Research Supervisor, Assistant Professor, Department of Computer Science, Government Arts College (Affiliated to Bharathidasan University, Trichy-24), Tiruchirappalli, 620 022, India
2 Assistant Professor, Department of Computer Science, Government Arts College (Affiliated to Bharathidasan University, Trichy-24), Tiruchirappalli, 620022, India

OPEN ACCESS

Received: 26-09-2022

Accepted: 27-11-2022

Published: 03-01-2023

Citation: Arul P, Renuka S (2023) Preserving the Privacy of the Healthcare, Clinical and Personal Data using Blockchain. Indian Journal of Science and Technology 16(1): 23-31. https://doi.org/10.17485/IJST/v16i1.1842

* **Corresponding author**.

spkumarrenu@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objective: The healthcare sector produces more data in every day. The security and privacy of this data are prone to misuse. This work aims to use blockchain technology to create healthcare systems that provide users with privacy, control, and data security. **Methods:** This study envisages a new method to secure private healthcare data and allows only the patients, doctors, and a few medically equipped persons to access the precious data by incorporating a recognition method, authorization of the user, user access monitoring, verification using digital signature and storing the confidential data in the blockchain. **Finding:** An elliptical curve (EC) method with a digital signature is used to generate recognition. If permission is granted the hash key will be held off-chain to reduce the burden of checking with the main blockchain. The blockchain 3.0 revolution has conveyed hopes for the healthcare sector. In blockchain 3.0, the multi-party computation protocol provides the healthcare data with the highest level of security and privacy to handle the ownership of the user data to the users and allow them to grant permission to the required personnel. **Novelty:** The decentralization of the data is done and by using blockchain technology in the healthcare sector, the users will have ownership and the privacy of the precious data can be maintained properly. Also, the user can decide who can access their private data. This proposed work can perform non-monetary transactions on the blockchain, which is extensively used in various sectors like supply chain, video streaming, the internet of things, and non-fungible tokens.

Keywords: Healthcare; Blockchain; Privacy; Multiparty Computation (MPC) Protocol; Elliptical curve (EC) method

1 Introduction

Many big corporate collects the user's data and the ownership of that data is limited to them alone, when that data got breached the user will be affected and this and particularly if these breaches occur in the medical and healthcare sector might be fatal.

The advent of big data [1] has transformed the entire data analytic scenario by a large scale and the data is perpetually acquired, processed, and archived. "Data is the new gold" many corporate hunts the data available in every source and procure them to build their business empire.

Mostly the hospital data are stored in centralized servers and the patients, as well as the user will not have any control over the data and the corporate will have the entire control of it [2]. The data-driven society is beneficial for large firms and companies but for the common people, they don't provide the much-needed security, privacy, and ownership of data. To overcome this problem, this study proposes a new method to safeguard precious private healthcare data and permit only the patients, doctors, and a few medically equipped persons to access the confidential data, thereby securing the privacy of the patient data and the patients will have the ownership and the privacy of their precious data. One more important thing is, the patient/user can decide who can access their private data and who cannot. The large number of security breaches, misuse of private healthcare data, and most importantly lapses in the privacy [3] of the data has occurred recently and this leads to develop and foolproof system using the blockchain to store the data in distributed ledgers.

The recent development in the blockchain is now, it can perform non-monetary transactions on the blocks [4] and it is called blockchain 3.0. This technology is extensively used in various sectors like supply chain (logistics), healthcare, video, and audio streaming, Internet of things, and NFT. There are lots of challenges and issues connected with the usage of blockchain in the healthcare sector and this study addresses all these to ensure that the private data of the patients (textual and image) is secured and most importantly the privacy of the data is maintained without any tampering and misusage. Simple blockchain architecture is shown in the followingFigure 1.

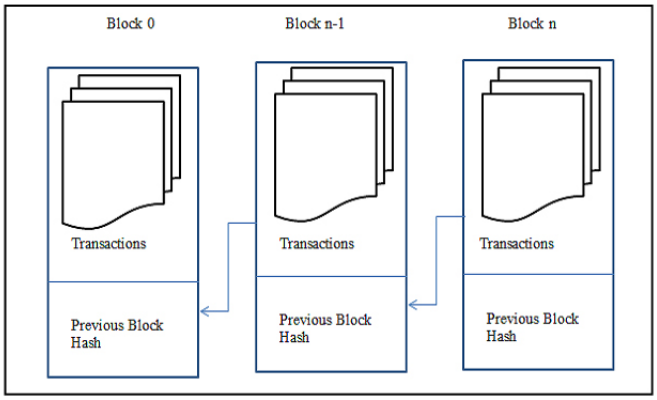

Fig 1. Simple block chain architecture

1.1 Problem Statement

The main problem or the issue to be addressed in this paper is the privacy of the user and their data. It is quite common that corporate hospitals collect the large number of personal and health-related data for the user and the user will not have any control over them. There are ample chances for their private confidential data to be shared with the research communities and thereby breach or violate the very privacy of the user. The proposed system safeguards the following

1.2 Ownership of data

The data of the user will be controlled and owned them and only the people with proper access or right can access the private data. The user or the patient will have all the rights and only if permission is issued, the data can be accessed by others.

1.3 Transparency of data

The patient knows what data is being collected, what data is being archived and how this data is being accessed, and who will access the private data.

1.4 Complete control

The user or the patient will have complete control over the data and they can access permission to some selected personals. The permission can be altered at any time if the user feels their privacy is being disrupted.

The better approach to handle this is to alleviate the permission of the users to access the data in its crude form and instead few computations are done in the network, dividing the data into pieces usually using Shamir secret sharing and the use multiparty computation to process the data.

Modern data sharing and accessing is carried out using the OAuth –an open standard that permits access to the users, allows data accessing grant to the users, allowing the applications to get accessed by the users without disclosing their identity or password introduced in the year 2012 as OAuth2.0 authorization framework. This is often used by the smart contract where some predetermined operations with certain probabilistic situations are presented to get the automated outcomes without any intervention of the third party or users [3].

The encryption is done using asymmetric algorithms and a plethora of methods is being used in the blockchain industry [5]. The types of asymmetric algorithms are named below,

1. RSA
2. ECS – Elliptical Curve
3. Code based
4. Diffie – Hellman

Among these four asymmetric schemes, it is always skeptical which particular method will suit the healthcare industry when blockchain is being utilized and scalability and the speed matter when it comes to accessing the permission blockchain. Also, the size of the data plays a major role in determining the overall performance of the data storage and retrieval in a blockchain based system. This work scrutinizes the speed of storage and retrieval based on the number of nodes and varying data sizes.

2 Methodology

2.1 Proposed Work

The proposed system comprises two important transactions in the blockchain and they are

Ta – Transaction access – Access control

Td – Transaction data – Data archived

A simplified diagram of the proposed work is shown in Figure 2.

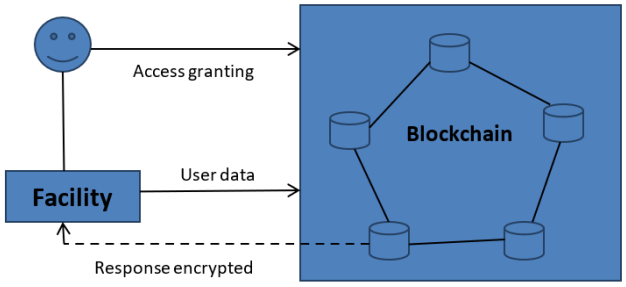

Fig 2. Transaction in the Blockchain

The user data along with the identity, records, and access data be encrypted using a shared key and sent to the blockchain for storage. This data will be routed to off-chain storage [6] and the pointer value or the index value will be present in the public ledger where the index value here is the SHA256 [7] hash of the data itself.

Here in Figure 2 the user and the facility can use the query to access the Td – data archived using the index or the hash associated with that particular data. The blockchain scrutinizes the access permissions of the facility and the user initially, and then the response is gathered in an encrypted format. Here the patients can alter the access permission and block any facility/service from accessing their data and also, they can set new permissions to the data on the blockchain. The major building block of the proposed system is illustrated clearly in this section, and they are,

2.2 Recognize

Usually, blockchain technology utilizes the quasi-identifier purposefully a public key and the patient/user can create any number of them according to their requirement. The recognition part is comprised of two categories and they are single recognition where the owner alone will be present and the second one is multiple recognition where along with the owner/user many numbers of entities are used to access the data.

In multiple recognition (public), the patient "p" and the facility "f" is represented as follows,

$Multiple^{public}_{(p,f)} = (PKey^{p,f}_{(sign)}, PKey^{f,p}_{(sign)})$

Multiple $_{(p,f)}$=$(PKey^{p,f}_{(sign)}, PKey^{f,p}_{(sign)}, FKey^{p,f}_{(sign)}, FKey^{f,p}_{(sign)}, FKey^{p,f}_{(encrypt)})$

The following procedure is used to create recognition as shown in Figure 3 .

Procedure Recognize (Patient *p*, Facility *f*)

Let us assume p and f are secure

Begin Patient:

Produce signature →$(PKey^{p,f}_{(sign)}, PKey^{f,p}_{(sign)})$

Produce Encryption →$FKey^{p,f}_{(encrypt)}$

p shares $(PKey^{p,f}_{(sign)}, PKey^{f,p}_{(sign)})$ with facility F

Begin Facility:

Produce signature →$(FKey^{p,f}_{(sign)}, FKey^{f,p}_{(sign)})$

f shares the$FKey^{f,p}_{(sign)}$with patient *p*

Return $PKey^{p,f}_{(sign)}, FKey^{f,p}_{(sign)}, FKey^{p,f}_{(encrypt)}$

End Procedure

Fig 3. Pseudo code of generating recognition

The user/patient process begins it produces the signature and encryption key and shares it with the facility. Similarly, the facility process begins to produce the signature and shares it with the users. When all the signatures and keys are generated, the computed keys are returned for usage.

2.3 Check permissions

This part checks the permission granted for the facility and the permission denied for the facility as the user will have to grant permission using several attributes like ID, contact, and based on location. The following procedure checks the permissions granted.

Procedure Check Permit (Patient *p*, Permission *Gp*)

Begin

Initialize a FLAG ← empty

Check all the permissions in Store →All_{permit}

If (All_{permit}≠ EMPTY) *do begin*

CHECK the keys and signatures

IF $(PKey^{original}_{(sign)}=PKey^{p,f}_{(sign)}$ ***OR*** $PKey^{original}_{(sign)}=PKey^{f,p}_{(sign)}$

IF ($Gp \in All_{permit}$) do *begin*

Replace the FLAG ← 1

CLOSE IF

CLOSE IF

CLOSE IF

Return FLAG

End Procedure

Fig 4. Pseudo code to check the permissions

Initially, the FLAG value is set to empty to denote that when the user first uses the BC, the permissions granted will be empty and if the user completes their registration, the permissions will be allocated and stored in the transaction data (i.e.) hash key will be stored in the off-chain to ease the burden of checking with the main blockchain.

2.4 Access management

The well-known fact is that the blockchain is tamper-free and the keys utilized in the process are secured very carefully and the user will have complete control over their data. The blockchain security along with the digital signature produced will safeguard

the data from any adverse corruption or attack on the network. The most important thing is the ledger is used to store only the hash value of the data and even if Sybil attack or 51% attack occurs, the data of the users will not tamper.

Procedure MonitorAccess (Patient *p*, Permission *Gp*)
Begin
Initialize a FLAG ← empty
Parse the signatures of user →PAR(*Gp*)
Check the permission FLAG
IF ($PKey_{(sign)}^{original}$=$PKey_{(sign)}^{p.f}$) *do begin*
Load the permission grants →*PKey* (*Gp*)
Replace the FLAG →1
CLOSE IF
Return FLAG
End Procedure

Fig 5. Pseudo code for monitoring the access

2.5 Data archive

The data from the user are stored in the ledger with hash values and the procedure to carry out the storage process is shown in the following Figure 6. Initially, the permissions are checked, if the permissions are ok, the reading and writing grants into the ledger are checked, If the permissions are granted, the hash values are appended to the ledger with the new access grants. The Pseudo code is shown in Figure 6.

Procedure STORE (Patient *p*, D)
Begin
IfCheckPermit (*p*, permissions) = TRUE
Check the Read permission
Check the Write permission
D→current data∪Olddata
Compute the hash H(D) →data(o)
//Hash table
If (data(o) ≠ empty
Return H(D)
Else
Return ϕ
End Procedure

Fig 6. Pseudo code to store the data

2.6 Data processing

To process the data MPC – Multi-Party Computation is used and it provides the much-needed privacy for the users but at the same time preserves the very nature of decentralization [8]. This is elaborated with an example, Let us consider a hospital sends a bunch of clinical data that comprises diagnosis results, and scan reports, details about patients to aggregate the data. The network selects random nodes present and then the aggregated data is securely converted into MPC. Finally, the data (hashed) are stored in the public ledger which is highly impossible to tamper and hack.

The encryption and the decryption parts are shown in the following Figures 7 and 8. The proposed model produces a digital signature on A's content using the private key to ensure its safe from being tampered with the use of B's encryption key the proposed model encrypts the signed Alice data in the smart contract. If "B" needs to access the real content of the "A" data, the corresponding private key of "B" is used to decrypt the data.

The sample data that is used is shown in the following Figure 9 and this data is initially encrypted and then a hash is created to store them in the public ledger.

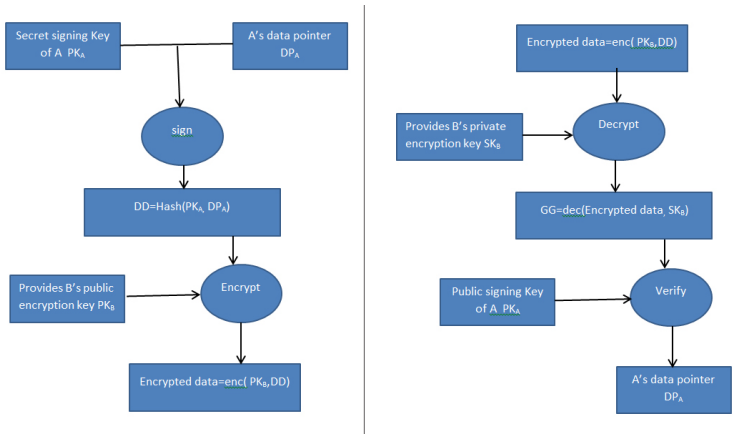

Fig 7. Encryption and decryption process

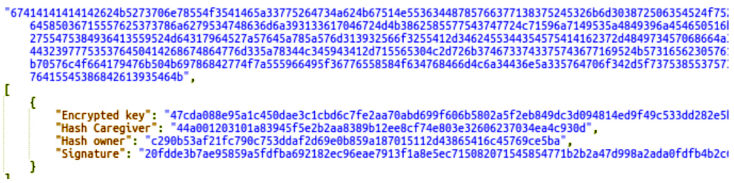

Fig 8. Signing and verification process

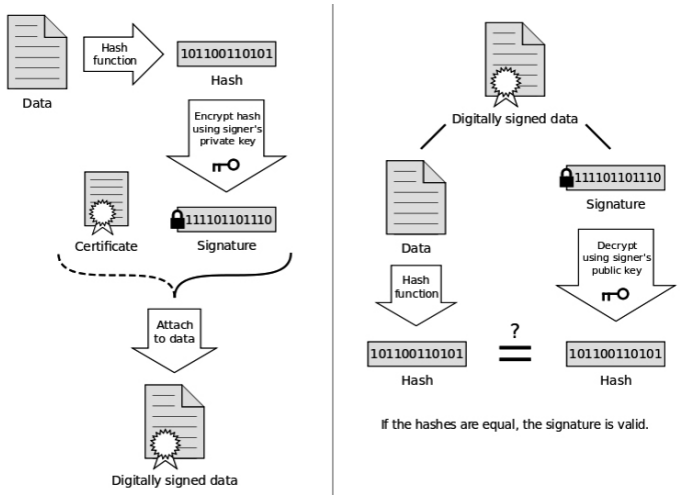

Fig 9. Sample data

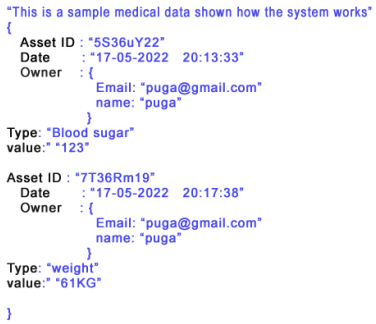

Fig 10. Encryptedsample data

3 Experimental Evaluation

The entire system is evaluated and simulated with 100 nodes and where the encryption algorithms are executed using the Python library tinyec. The evaluation result showcased that the encryption method ECS is better than RSA and DHHHh. The evaluation time is computed for the three methods and compared when executed on dara.cms.gov and the size of the data is varied to find the execution time performance finally, a comparison with respect to the varying number of nodes and varying data size is carried out to test the overall performance of the proposed system.

Table 1. Execution time comparison with respect to varying nodes

Execution time comparison			
Encryption scheme	**Tx in seconds**		
	30 Nodes	**60 Nodes**	**100 Nodes**
RSA	0.126	0.179	0.276
EC	0.102	0.116	0.128
DH	0.135	0.168	0.247

Table 2. Execution time comparison with respect to varying data size

Tx time comparison with 100 nodes			
Encryption scheme	**Data size**		
	1000	**1500**	**2000**
RSA	0.256	0.387	0.493
EC	0.187	0.236	0.289
DH	0.227	0.337	0.427

From Figures 11 and 12, it is quite clear that the elliptical curve method performed quite well when compared with the other two methods RSA and Dillie- Hellman, and consumed very less time to perform the conversion of raw data to cipher data.

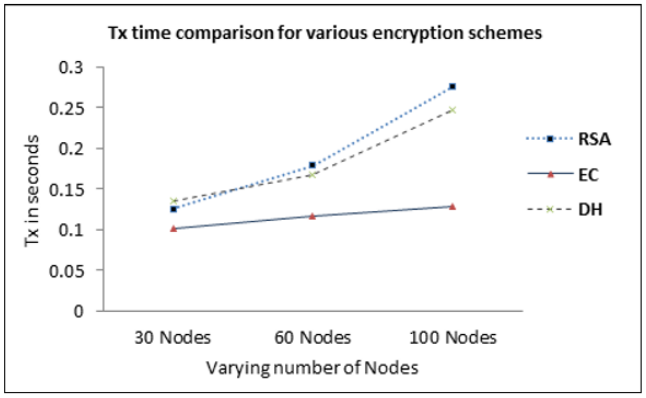

Fig 11. Tx time comparisonwith respect to nodes

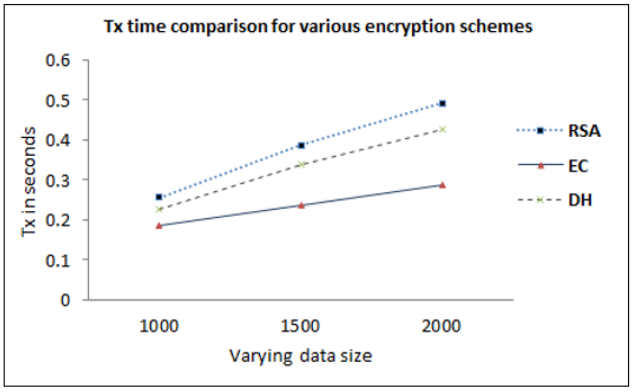

Fig 12. Tx time comparison with respect to data size

4 Conclusion

This study provides a blockchain-based new model that aids the healthcare industry with the utmost security and privacy using the multi-party computation protocol and handling the ownership of the data to the users and allowing them to grant permission to the required personals and more importantly what data is being captured, who all are accessing the data known to the users/patients.

References

1) Leiyongguo H, Li Y. Data Encryption based Blockchain and Privacy-Preserving Mechanisms towards Big Data. *Journal of Visual Communication and Image Representation*. 2019;70:102741. Available from: https://doi.org/10.1016/j.jvcir.2019.102741.
2) Ghosh A, Gupta S, Amitdua, Kumar N. Security of Cryptocurrencies in blockchain technology: State-of-art, challenges and future prospects. Security of Cryptocurrencies in. 2020. Available from: https://doi.org/10.1016/j.jnca.2020.102635.
3) Fu J, Wang N, Cai Y. Privacy-Preserving in Healthcare Blockchain Systems Based on Lightweight Message Sharing. *Sensors*. 2020;20(7). Available from: https://doi.org/10.3390/s20071898.
4) Domadiya N, Rao UP. Privacy Preserving Distributed Association Rule Mining Approach on Vertically Partitioned Healthcare Data. *Procedia Computer Science*. 2019;148:303–312. Available from: https://doi.org/10.1016/j.procs.2019.01.023.
5) Shepingzhai Y, Yang J, Li C, Qiu J, Zhao. Research on the Application of Cryptography on the Blockchain. *Journal of Physics: Conference Series*. 2019. Available from: https://doi.org/10.1088/1742-6596/1168/3/032077.
6) Kumar R, Marchang N, Tripathi R. Distributed Off-Chain Storage of Patient Diagnostic Reports in Healthcare System Using IPFS and Blockchain. *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*. 2020;p. 1–5. Available from: https://doi.org/10.1109/COMSNETS48256.2020.9027313.
7) Haque R, Sarwar H, Kabir SR, Forhat R, Sadeq MJ, Akhtaruzzaman M, et al. Blockchain-based information security of electronic medical records (EMR) in a healthcare communication system. In: Intelligent Computing and Innovation on Data Science . Springer. 2020;p. 641–650. Available from: https://doi.org/DOI:10.1007/978-981-15-3284-9_73.

8) Zhou J, Feng Y, Wang Z, Guo D. Using Secure Multi-Party Computation to Protect Privacy on a Permissioned Blockchain. *Sensors*. 2021;21(4):1540. Available from: https://doi.org/10.3390/s21041540.