Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 27 (2024), No. 2-A, pp. 237–248
DOI : 10.47974/JDMSC-1878

Merkle-Damgård hash functions and blockchains : Securing electronic health records

Hayder M. A. Ghanimi [#]
Department of Information Technology
College of Science
University of Warith Al-Anbiyaa
Karbala
Iraq

and

Department of Computer Science
College of Computer Science and Information Technology
University of Kerbala
Karbala
Iraq

Romel P. Melgarejo-Bolivar [†]
Postgraduate Unit of Informatics
Faculty of Statistical and Computer Engineering
Universidad Nacional del Altiplano de Puno
P. O. Box 291
Puno-Peru

Alfredo Tumi-Figueroa [§]
Faculty of Human Medicine
Universidad Nacional del Altiplano de Puno
P. O. Box 291
Puno-Peru

[#] *E-mail:* hayder.alghanami@uowa.edu.iq
[†] *E-mail:* permelro@unap.edu.pe
[§] *E-mail:* alfredo2891@yahoo.es

Samrat Ray ‡
Department of Management Studies
IIMS
Pune 411033
Maharashtra
India

Pankaj Dadheech @
Department of Computer Science and Engineering
Swami Keshvanand Institute of Technology
Management & Gramothan (SKIT)
Jaipur 302017
Rajasthan
India

Sudhakar Sengan *
Department of Computer Science and Engineering
PSN College of Engineering and Technology
Tirunelveli 627152
Tamil Nadu
India

Abstract

The integration of Blockchain (BC) technology with the Merkle-Damgård Structure (M-DS) presents a practical model for securing Electronic Health Records (EHR). This research study examines how these cryptographic mechanisms can be functional to address confidentiality, integrity, and availability concerns when working with specific clinical data. Secure EHR is developed based on the M-DS, which is notable for being capable of generating Collision-Resistant Hash Functions (CRHF). The summary of BC-based technology in the EHR environment enhances access control, allows it to be fixed, and simplifies decentralized access. A decentralized BC system stores the values that have been hashed, resulting in patient EHR using the M-DC, as per the proposed method. As an outcome of smart contracts' ability to manage access, only persons who have been authorized to access can analyze and alter clinical information. The strategy of the model's cryptographic mechanism highlights openness and accountability by securing data confidentiality by encryption and by regularly generating a verification record of every transaction. This work discusses an exhaustive method to secure EHR in a clinical system that develops more computer-aided and data-driven over performing a complete assessment of the mathematical models that are enabling MDC and BC Technology within the overall system of EHR security.

Subject Classification: *68N30.*

Keywords: *Merkle-Damgård structure, Deep learning, Classification, Blockchain, Security, Hash function, Healthcare.*

‡ *E-mail:* **s.ray@iimspune.edu.in**
@ *E-mail:* **pankajdadheech777@gmail.com**
* *E-mail:* **sudhasengan@gmail.com (Corresponding Author)**

1. Introduction

Several hospitals are implementing Electronic Health Records (EHR) systems that enable rapid access to patient information as a result of the increasing requirement for periodic monitoring of health due to the expanding prevalence of illnesses and injuries [1]. As clinical data is generated in high volume, practical handling, and storage are becoming increasingly vital. While there are several performance merits to cloud computing for healthcare providers, there are also some problems regarding the security and confidentiality of data. This is because EHRs include personal information about patients that can be exploited in digital attacks [2]. The problems stated prior have been explained by blockchain (BC) technology, agreeing for secure and efficient EHR and control. The BC methodologies offer privacy by emerging a chain of blocks, each with a correct date, time, and hash, in order to avoid unauthorized use and tampering with the core clinical data [3]. A cost-effective, better service is obtainable by using this method [4]. Even though the core principle of BC has been contemporary since 1991, it was not included until 2008, when Satoshi Nakamoto made it notable by introducing it for Bitcoin.

Blocks in a decentralized, shared ledger system like BC are connected securely by employing hash values in order to form a chain. Transactions using Bitcoin involve the indefinite storage of EHR of transactions that peers check for authenticity [5]. When attackers attempt to manipulate data, the hash algorithms change, making certain portions of the data invalid. As an outcome of BC's method of authentication and cryptography-signed functions, the security of data improves with each authentication [6]. Public and private EHRs can be securely stored in a cloud computing environment because of Elliptic Curve Cryptography (ECC) approaches that create and validate digitally signed documents for transactions. This prevents unauthorized individuals from fraud, modifying, or accessing personally identifiable information [7]. Backups of information on the initial blocks belong to BC's data security model, enabling the easy recovery of data in the case of an emergency and verifying and removing data matching modifications presented by the network's memory [8]. The above method validates BC's cloud-based security metrics that secure patients' private information and avoid unauthorized gain of access to their historical data in a flash [9]. A comprehensive study of prior methods for securing HER from several risks in the digital platform motivated the present research investigation [10]. Optimizing conventional methods for

encryption by selecting random primary keys to generate decryption keys is the primary objective of the research analysis, which aims to improve EHR security versus potential vulnerabilities [11-2]. The economic and privacy implications of sharing confidential patient information boost this action sequence [13-15]. The article's organization is ordered in the following manner: Section 2 is the proposed methodology. Section 3 defines the Simulation Outcome of the research, and Section 4 presents the Conclusion and Future Work of the paper.

2. Proposed Methodology

The recommended encryption method for an authorization BC network is listed more thoroughly below. Each of the five phases of the proposed framework is outlined as follows: Setup, Key Generation (KG), Key Exchange (KE), and Cryptographic. Figure 1 proves the layout of the data encryption system that has been defined.

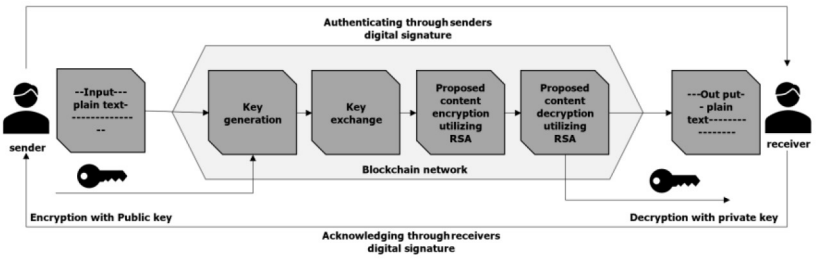

Figure 1

A Method for Secured Authenticating

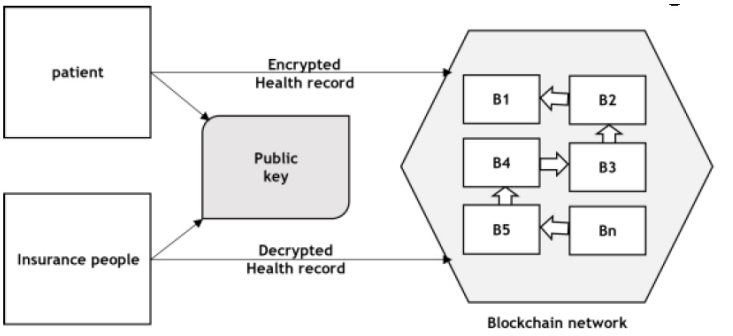

Figure 2

Proposed Encryption Model for EHR

2.1 *Initialization*

The use of symmetric and asymmetric keys provides effective encryption and decryption in the intended privacy-preserving content encryption method. During the setup process, significant data like receivers' and senders' keys, file destinations, information, and the encryption code have been determined. These informational components are part of the BC. Improved security and privacy of transferred data is the outcome of the method's design of protocol implementation. Within this recommendation access BC network, the nodes that are accountable for representing users *(P)* are $\{P_1, P_2, P_3, \ldots P_n\}$, while the nodes presenting blocks *(B)* are $\{B_1, B_2, B_3, \ldots B_n\}$. Each block contains file contents *(D)* represented as $\{D_1, D_2, D_3, \ldots D_n\}$. Participants, blocks, and file content are symbolized as $\{P_i, B_i, D_i\}$, where i belongs to a set of integers. File Path (FP) directs local or IPFS storage, and asymmetric key pairs *(PrK, PuK)* facilitate encryption. For secure communication, the sender and receiver assign keys *(SSK, SPuK, RSK, RPuK).* The system hypothesizes and shares protocol parameters with authorized BC network users before encryption and decryption. Key setups are explained, followed by the KG stage, advancing the system to its subsequent stage Table 1.

Table 1

Elements in Proposed Technique

S. No	Elements	Parameters	Description
1	Users	P = { P1, P2, P3, …Pn}	Sender and Receiver
2	File Path	FP = File Location	Local File Storage, Inter Planetary File System (IPFS)
3	Block	B = { B1, B2,B3,…Bn}	BC Blocks
4	File Content	D = { D1, D2, D3,… Dn }	Original Text / PlainText
5	Cipher Text	C	Encrypted Text
6	Symmetric Key	S_Key	Symmetric Key for Content Encryption
7	Asymmetric Key	PrK, PuK	PuK Cryptography
8	Asymmetric Key	SSK, SPK	Sender Initiated
9	Asymmetric Key	RSK, RPK	Receiver Initiated
10	Digital Signature	Digi_Sign	Authentication and Acknowledgement

2.2 *Key Generation*

This subsection presents the KG method of a proposed hybrid encryption system. The system generates keys for each node after initializing and identifying BC users. The privacy-preserving system performs double encryption and decryption to enhance security and privacy. The protocol design generates three keys for each network participant: one symmetric key and a pair of asymmetric keys with PrK and PuK. This process is consistent for all network participants.

Algorithm 1. PuK and PrK Generation
Step 1: Let K and Q be large primes.
Step 2: Multiply K and Q to get N.
Step 3 : Calculate ij(N) = ((K í 1)(Q í 1))
Step 4 : Generate public exponent e GCD(e, ij(N)) == 1,
Step 5 : Generate private exponent d = (2 + ij (N)) / e
Step 6: Obtain PuK {e, N}
Step 7: Obtain PrK {d, N}

Asymmetric KG using Algorithm 1 and key exchange are described.

2.3 *Key Exchange*

A request to share keys allows multiple users in the proposed privacy-preserving system to exchange keys. Symmetric and asymmetric key pairs let senders and receivers double encrypt/decrypt. The receiver starts the two-stage KE.

Stage 1: Senders exchange Public Keys (RPuKs) and request content.
Stage 2: An encrypted symmetric key from the sender is used to decrypt a PuKey.

The following section explains encryption.

2.4 *Encryption and Decryption of EHR*

EQU (1) compresses EHR into M-sized compression blocks.

$$M \rightarrow M \| 0 \| 0^k \| len(M) \tag{1}$$

Here, a binary representation of the original message length is shown *len(M)*, with '1' representing a single '1' bit and 0^k representing *'k'* zeros, and || denotes concatenation.

Let *'f'* be a Hash Collision Resistance Function (HCRF) explicitly designed for EHR data. The Compression Function (CF) takes the current chaining variable C_{i-1} and the current block of EHR data M_i as input and produces the new chaining variable C_i, EQU (2)

$$C_i = f(C_{i-1} \| M_i) \tag{2}$$

The CF should be designed to handle the unique features of EHR, ensuring confidentiality and integrity.

Divide the padded EHR data into blocks $\{M_1, M_2, \ldots, M_t\}$ of size r bits each. Initialize an *n-bit* chaining variable 'C_0' to a fixed initial value. Apply the CF iteratively, EQU (3)

$$C_i = f(C_{i-1} \| M_i) \tag{3}$$

The final output of the HER-Hash Function (HF) is the last chaining variable, C_t, EQU (4)

$$H(M) = Ct \tag{4}$$

The HCRF of the CF *'f'* ensures that finding two distinct EHR datasets, $\{M_1, M_2\}$, such that $H(M_1)=H(M_2)$ is computationally infeasible. Using a secure HF in the EHR context helps maintain data integrity, ensuring any tampering with the EHR will be detectable. The method of the encryption system for Content Encryption in HER systems is illustrated in Figure 2.

Every HER-related transaction is recorded on the BC, creating an immutable audit trail for accountability. Logs are maintained to track every access to health records, providing transparency and accountability.

2.5 *Merkle-Damgård Construction (M-DC) Theorem*

The M-DC theorem states that if you have a HCRF, $f:\{0,1\}^{n+r} \rightarrow \{0,1\}^n$, where *'n'* is the output length, and *'r'* is the input block size, then you can construct a HCRF function $H:\{0,1\}^* \rightarrow \{0,1\}^n$ by iteratively applying the CF.

The construction works as follows:

Padding: If the message M is not a multiple of *'r'*, pad it with a single *'1'* bit followed by *'0'* bits and append the binary representation of the original message length, EQU (5)

$$M \rightarrow M \| 0 \| 0^k \| len(M) \tag{5}$$

Iterative Application: Divide the padded message into blocks $M_1, M_2, \ldots, M_t$ of size *'r'* bits each. Initialize an n-bit chaining variable 'C_0' to a fixed initial value. Then, iterate the following CF for each block, EQU (6)

$$C_i = f(C_{i-1} \| M_i) \tag{6}$$

Output: The result of the HF is the last chaining variable, 'C_t'

Proof: The proof shows that the resulting HF is also collision-resistant if you have a HCRF. It typically relies on the properties of the CF, such as the difficulty of finding two different inputs that produce the same output. The confidentiality of the data can be assured by providing an approach for developing HCRF, and the M-DC enables that.

3. Simulation Outcome

A computerized simulation to secure EHRs using BC innovation and M-DS can be implemented in Python language. The architecture uses the '*cryptography*' module in Python to model encryption and decryption, and it incorporates the use of Ethereum's smart contract functionality for access control. The model simulation runs for the objectives of testing the model's settings, generating hypothetical patient EHR, and measuring computational overhead and execution time for authentication and authorization. A comprehensive model for examining the recommended EHR security method has been presented by Python data assessment tools, which measure performance indicators.

3.1 *HCRF*

HCRF is vital for ensuring the security of the M-DC. It measures the likelihood of two distinct inputs producing the same hash value. The ideal HF minimizes the probability of collisions, making it computationally infeasible for an attacker to find different inputs with the same hash. The birthday paradox EQU (7) can quantify the HCRF Figure 3.

$$\text{Probalility of Collision} \approx 1 - e^{\frac{n^2}{2.2^b}} \tag{7}$$

where distinct inputs are given as '*n*', and hash output bits are presented as '*b*'. The performance assessment is shown in Table 2.

Table 2

Performance Evaluation

	HCR	(TPS) (s)	SCET (s)	TDT (hr)
Low Probability	0.000000000001	50	2	1
Moderate Probability	0.000000001	30	5	3
High Probability	0.000001	10	10	6

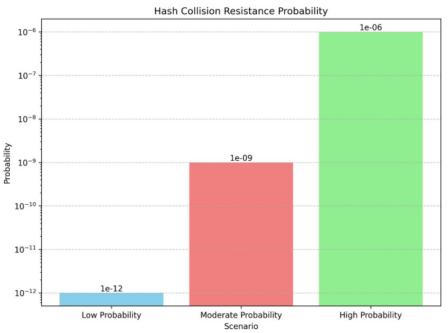

Figure 3

Comparison of HCR

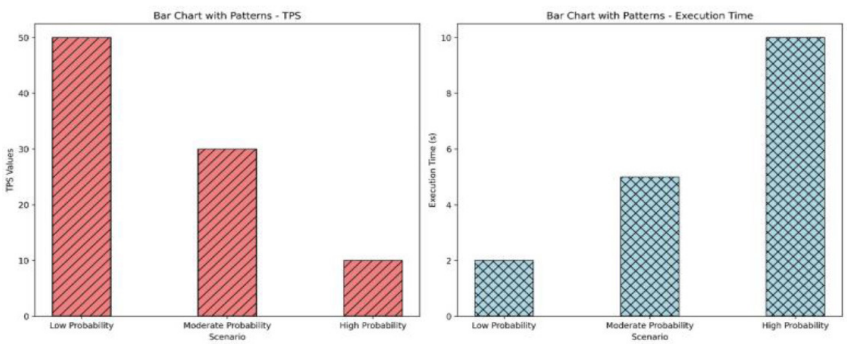

Figure 4

Comparison of TPS and SCET

3.2 *Transaction Per Second (TPS)*

Transaction throughput, often measured in TPS, assesses the rate at which the BC network can process transactions Figure 4. It reflects the system's capacity to handle a high volume of transactions efficiently, EQU (8).

$$TPS = \frac{\text{Number of Transaction}}{\text{Total Time}} \tag{8}$$

The number of transactions is the total number of transactions processed, and the entire time is the duration over which the transactions are measured.

3.3 *Smart Contract Execution Time (SCET)*

The time essential to analyze and perform a smart contract on the BC network is evaluated by SCET. Appropriate access control within an EHR security system requires the efficiency of smart contract processing. The execution time can be represented as:

$$SCET = \text{End Time} - \text{Start Time} \tag{9}$$

where End Time is the time when the smart contract execution is completed, and Start Time is the time when the smart contract execution begins.

3.4 *Tamper Detection Time (TDT)*

TDT is the total length of time it requires for the technology to detect and respond to any attempt at manipulating the medical information that is stored Figure 5. The security and permanence of the EHR relies on it. The TDT is expressed as EQU (10)

$$TDT = \text{Detection Time} - \text{Tamper Attempt Time} \tag{10}$$

where Detection Time is when the system detects the tampering, and Tamper Attempt Time (TAT). TAT is when the tampering attempt occurs.

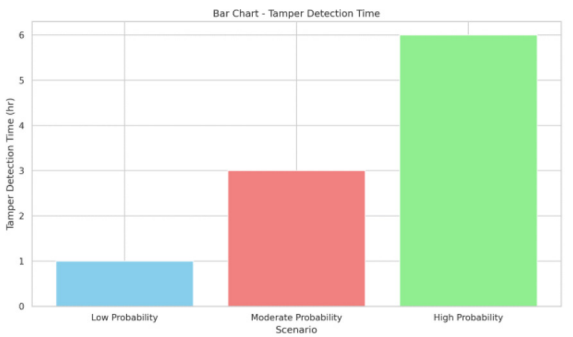

Figure 5

Comparison of TDT

The research examines how BC's invention can be used to expand the EHR security of EHRs, primarily concentrating on HCRF. A mathematical model is predicated on the Birthday Paradox Equation (BPE), and it analyzes the collision probability with different conditions. The predicted result of similar hash values expanding rapidly with lower HCR is made clear by the results. The investigation highlights the implication of a HCR function in evading unauthorized access to clinical data. Investigating how HCR impacts fundamental performance analysis proves that HCR has a low correlation with TPS. While concerning TDT, the subject of analysis dives into finding a balance between the accuracy of access control and the security of HCR. The conclusions of this study can assist with health data privacy decisions by signifying an ideal balance between efficiency metrics and HCR in EHR systems.

4. Conclusion and Future Work

The research aims at the real-time application of Blockchain (BC) technological developments and the Merkle-Damgård Structure (M-DS) to improve EHR security of EHRs. The M-DS outcomes in Hash Collision Resistance Function (HCRF) secure the privacy linked with clinical data. The recommended system minimizes the risk of HCR while improving decentralization, stability, and access control through the EHR environment by monitoring who can view and modify what EHR. Smart contracts make sure that only those with authorization have access to private data. The cryptographic system improves the security of data communication by encrypting confidential data. The model's better use of cryptography, static analysis, and scalability provide a proper match for data-driven, networked EHR systems. Still, it could use a limited improvement in other research fields.

References

[1] D. K. Murala et al., Securing electronic health record system in cloud environment using blockchain technology, In *Recent advances in blockchain technology: real-world applications*, Cham: Springer International Publishing, pp. 89-116 (2023).

[2] K. T. Akhter Md Hasib et al., Electronic health record monitoring system and data security using blockchain technology, *Security and Communication Networks*, pp. 1-15 (2022).

[3] H. B. Mahajan et al., Integration of Healthcare 4.0 and Blockchain into secure cloud-based electronic health records systems, *Applied Nanoscience*, vol. *13, no.* 3, pp. 2329-2342 (2023).

[4] B. K. Rai, "PcBEHR: patient-controlled blockchain-enabled electronic health records for healthcare 4.0," *Health Services and Outcomes Research Methodology*, vol. *23, no.* 1, pp. 80-102 (2023).

[5] H. B. Mahajan, "Emergence of healthcare 4.0 and blockchain into secure cloud-based electronic health records systems: solutions, challenges, and future roadmap," *Wireless Personal Communications*, vol. *126, no.* 3, pp. 2425-2446 (2022).

[6] F. A. Reegu et al., Systematic assessment of the interoperability requirements and challenges of secure blockchain-based electronic health records, *Security and Communication Networks* (2022).

[7] U. Chelladurai and S. Pandian, A novel blockchain-based electronic health record automation system for healthcare. *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-11 (2022).

[8] G. Zhang, Z. Yang, and W. Liu, "Blockchain-based privacy preserving e-health system for healthcare data in the cloud," *Computer Networks*, *203*, 108586 (2022).

[9] M. Alsayegh et al., Towards secure, searchable electronic health records using consortium blockchain, *Network*, vol. *2, no.* 2, pp. 239-256 (2022).

[10] A. Amanat et al., Blockchain and cloud computing-based secure electronic healthcare records storage and sharing, *Frontiers in Public Health*, vol. *10*, no. 938707 (2022).

[11] L. Soltanisehat et al., Technical, Temporal, and Spatial Research Challenges and Opportunities in Blockchain-Based Healthcare: A Systematic Literature Review, *IEEE Transactions on Engineering Management*, vol. 70, no. 1, pp. 353-368 (2023).

[12] W. Liu et al., A Blockchain-Based System for Anti-Fraud of Healthcare Insurance, IEEE 5th International Conference on Computer and Communications, Chengdu, China, pp. 1264-1268 (2019).