HASIL CEK_Developing Data Integrity in an Electronic Health Record System using Blockchain and InterPlanetary File System (Case Study: COVID-19 Data)

by Imam Riad Ahmad, Sarno, Purwono, Ma'arif

Submission date: 23-Apr-2022 10:44AM (UTC+0700)
Submission ID: 1817877211
File name: ain_and_InterPlanetary_File_System_Case_Study_COVID-19_Data.pdf (1.56M)
Word count: 8808
Character count: 46944

Available online at www.ijournalse.org

Emerging Science Journal

Vol. 4, Special Issue "IoT, IoV, and Blockchain", (2020, 2021)

Developing Data Integrity in an Electronic Health Record System using Blockchain and InterPlanetary File System (Case Study: COVID-19 Data)

Imam Riadi [1*], Tohari Ahmad [2], Riyanarto Sarno [2], Purwono Purwono [3], Alfian Ma'arif [4]

[1] *Department of Information System, Universitas Ahmad Dahlan, Yogyakarta 55164, Indonesia.*

[2] *Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia.*

[3] *Department of Informatics, Universitas Harapan Bangsa, Purwokerto, 53182, Indonesia.*

[4] *Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, 55191, Indonesia*

Abstract

The misuse of health data stored in the Electronic Health Record (EHR) system can be uncontrolled. For example, mishandling of privacy and data security related to Corona Virus Disease-19 (COVID-19), containing patient diagnosis and vaccine certificate in Indonesia. We propose a system framework design by utilizing the InterPlanetary File System (IPFS) and Blockchain technology to overcome this problem. The IPFS environment supports a large data storage with a distributed network powered by Ethereum blockchain. The combination of this technology allows data stored in the EHR to be secure and available at any time. All data are secured with a blockchain cryptographic algorithm and can only be accessed using a user's private key. System testing evaluates the mechanism and process of storing and accessing data from 346 computers connected to the IPFS network and Blockchain by considering several parameters, such as gas unit, CPU load, network latency, and bandwidth used. The obtained results show that 135205 gas units are used in each transaction based on the tests. The average execution speed ranges from 12.98 to 14.08 GHz, 26 KB/s is used for incoming, and 4 KB/s is for outgoing bandwidth. Our contribution is in designing a blockchain-based decentralized EHR system by maximizing the use of private keys as an access right to maintain the integrity of COVID-19 diagnosis and certificate data. We also provide alternative storage using a distributed IPFS to maintain data availability at all times as a solution to the problem of traditional cloud storage, which often ignores data availability.

Keywords:

COVID-19;
EHR;
Blockchain;
IPFS;
Network Security.

Article History:

Received:	04	September	2021
Revised:	18	January	2022
Accepted:	29	January	2022
Published:	01	February	2022

1- Introduction

The development of information systems triggers an increase in the data size [1], including those in the health sector. Data security is one of the crucial factors that must be prioritized in the cyber age [2]. Almost two years of the COVID-19 pandemic worldwide, there are still various data integrity and security problems. In 2021, Reuters [3] reported leakage of COVID-19 vaccine certification data of the President of Indonesia from a government-owned application. This indicates a severe problem concerning the weak protection of user data security. Personal data, such as identity card information, which should be kept secretly, can even be easily spread to the public, which irresponsible individuals potentially misuse. Another problem is the synchronization of the fatality number between that of the national and local

* **CONTACT**: imam.riadi@is.uad.ac.id

DOI: http://dx.doi.org/10.28991/esj-2021-SP1-013

governments. Reporting from the BBC [4], there is a considerable difference in the data of fatalities caused by COVID-19, which is about 19,000 cases. This data uncertainty of COVID-19 raises an accusation that the government is trying to "tamper with" the definition of COVID-19 fatalities, which implies the data reports displayed.

The data related to the diagnosis and vaccination of COVID-19 are generally processed by authorized hospitals using Electronic Health Record (EHR). Manipulation of COVID-19 data in the EHR in a hospital may occur due to the actions of irresponsible individuals [5]. Data manipulation is included in the form of fraud that can result in substantial losses [6]. This violation tripled in 2019 from previous years due to minimal security measures or even no security measures in the EHR [7]. Another problem with the traditional EHR is hospitals' control and hoarding of patients' health data. This condition has caused delays in data consolidation between the EHR systems. Furthermore, it results in untimely health care to patients [8].

In this paper, a system that can maintain the privacy of patient data and the security of medical record reports related to COVID-19 is proposed. This system seeks to maintain the integrity of the data generated to protect the privacy of patient data and statistical data of cases displayed on the official website of the Indonesian government, i.e., www.kawalcovid19.id, as an effort to display data according to the facts in Indonesia. To meet this need, we utilize blockchain technology, which is a sophisticated and robust cryptographic mechanism that allows the exchange of data across healthcare providers with patients' consent [7]. We also integrate the Blockchain with the InterPlanetary File System (IPFS), which aims to allow users to process a large amount of data so that the data do not need to be placed on a chain that not only saves network bandwidth but also effectively protects them [9]. The security of the resulting data from this system is shown in the encrypted data stored in IPFS and the distributed files. It is the solution to problems created by cloud service providers who usually delete some data to save storage space. This is why we use Blockchain technology and IPFS to solve this problem. Blockchain as a data-sharing platform ensures that the processing of personal health data (e.g., COVID-19 test results or vaccine certificates) can only be accessed by authorized parties, and the data will always be immutable, so it cannot be changed [10]. Blockchain allows patients to place themselves in an ecosystem environment while increasing their data security confidentiality and interoperability [11].

MetaMask is used as a storage medium for patients' and doctors' private keys. It is a cryptocurrency wallet that can be used in various browsers such as Chrome, Firefox, and Edge [12]. The goal of MetaMask is to connect to the Ethereum blockchain network easily. Additionally, MetaMask is used as a bridge that connects regular browsers with the Ethereum blockchain to ensure that only doctors or healthcare services and patients who have private keys can access the application. A health care provider or doctor can only access patient's medical record data if the patient permits access to the data. Each medical record is stored on a peer-to-peer node ledger. We tested the proposed framework on a Ganache truffle, which is a development environment, a test framework, and an asset channel for blockchain-based on Ethereum Virtual Machine (EVM) [13]. The ganache will allow us to create and compile smart contracts on a private Ethereum blockchain network and check the status and control how the chain operates.

Several related studies that utilize blockchain technology and IPFS have been conducted by Ramos et al. (2021) [14], namely by utilizing this technology to develop a digital vaccination certificate. The research conducted by Hasan et al. (2020) [18] utilizes blockchain technology and IPFS to implement digital passports that are integrated with Covid-19 digital vaccination certification. The research conducted by Kumar et al. (2021) [16] utilizes blockchain technology and IPFS as an effort to overcome health data security problems in cloud computing service providers from ransomware and Distributed Denial of Service (DDoS) attacks during COVID-19, which can cause emergency services to be stopped.

Based on the above statement, this paper proposes a secure storage media that can maintain COVID-19 patients' data privacy in Indonesia with blockchain technology and IPFS. We scheme to utilize smart contracts on the Ethereum blockchain and IPFS to implement a system that can maintain the privacy of patients' data and data security. In general, the main contributions and innovations of this paper are:

- Utilizing the smart contract of the Ethereum blockchain that can identify private keys and public keys as the control of access right in the EHR system. Patients' data can only be stored and shared if they give their private key to other users. This method is useful because it can protect patient's data, and the stored data will be intact to maintain data integrity. If an unauthorized person distributes the data, the system can easily track its distribution.
- Implementing a distributed file system using point-to-point IPFS to encrypt and store patients' medical record data. IPFS ensures that the EHR data also remains available. This system answers a problem that often arises from traditional cloud service providers, which are allowed to delete some data to save storage space.

The remainder of this paper is organized as follows. Section (2) Research Materials and Methodologies summarizes related works on previous literature and research methodologies, including explaining the general concepts of blockchain technology covering work technologies, architectures, Ethereum, smart contracts, and IPFS used in developing the system framework. Section (3) Results and Discussion describes this study's results. Section (4) Discussion that discusses the results obtained in the form of advantages and weaknesses of the system. Section (5) Conclusion that concludes this paper.

2- Material and Methods

This section explains the general concept of blockchain technology which includes the working, architecture, Ethereum, smart contract, and IPFS technology used in developing the system framework. This study uses a quantitative approach with survey and observation strategies based on the development of previous research studies. Figure 1 presents the steps of the research methodology.

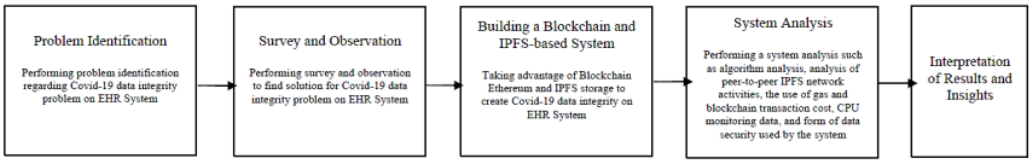

Figure 1. Research methodology steps

2-1-Related Works

Distributed ledger technologies, particularly the blockchain, can be represented in terms of different layers which include the infrastructure, data, network, and application. A layered view of blockchain is presented in Figure 2 and this structure is used to describe different layers as well as the components within. In order to describe these layers and components, an agnostic language of technologies have been applied within a specific platform, for example, the ethereum, hyperledger, and multichain [17].

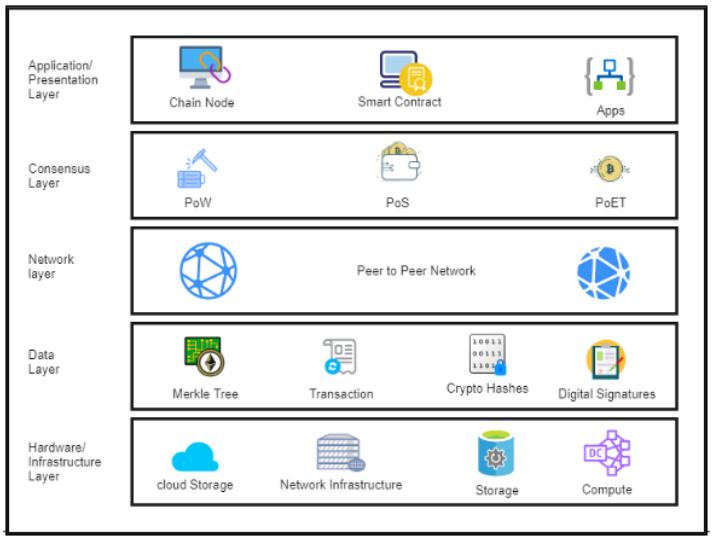

Figure 2. A layered view of blockchain

The IPFS is applied for off-chain documents stored in a decentralized method. As the documents related to the COVID-19 testings, identification, and travel will be too costly to be stored on-chain, those documents are required to be stored in a decentralized and secure method. The IPFS storage is circulated and published to everyone; therefore, the information saved on the IPFS must be encrypted and only the authorized entities have the permission to read the plaintext content.

The detailed steps included in proxy re-encryption are presented in Figure 3. The COVID-19 test-takers use one symmetrical key in order to encrypt the medical records and the PII (Personal Identifiable Information). After that, the encrypted data and the symmetric key are stored on the IPFS. By only storing the hash of the encrypted information on-chain in the smart contract, only the data owners are able to identify the symmetric key as it is without encryption. The data owner will recreate a new key using the receiver's public key and his/her private key whenever a receiver needs to retrieve some data stored in the IPFS. Then, the newly generated key is sent to the proxy network to be re-encrypted. As shown in Figure 2, the COVID-19 test-takers store the encrypted le,EncK (F) and the key on the IPFS. If any interested party (academic institutions, travel agents, or transportation facilities) wants to access the IPFS le content, it will communicate with the re-encryption proxy network. The COVID-19 test-takers will later send a key to the re-encryption proxies in order to generate a new key, which can be used by the one who needs it.

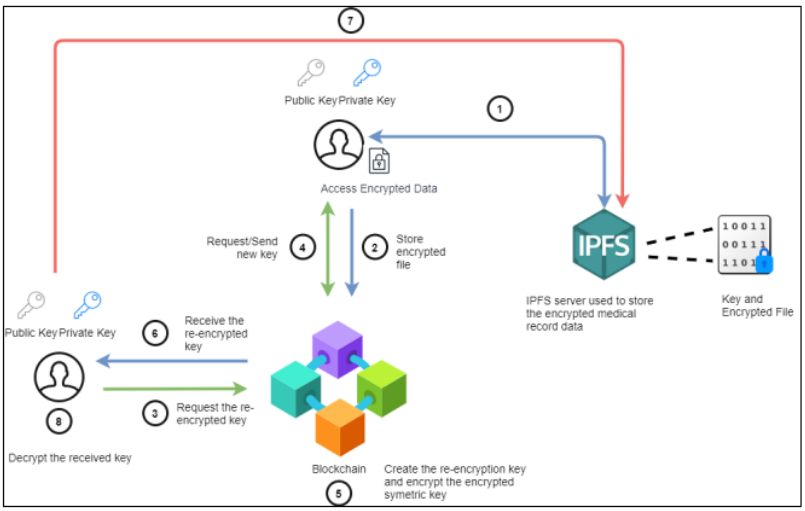

Figure 3. Re-encryption Proxy

In Figure 3, the block above numbers 2 and 4 is the COVID-19 test taker, the block above number 5 is the re-encryption proxy network, and the block above number 8 is the interested parties. Number 1 contains a store or access processing for encrypted data of K key. Number 2 contains a Store encrypted file and Pub_O process, namely the owner's public key. Number 3 contains a request for the re-encrypted key. Number 4 contains a request/sending of a new key. Number 5 contains the creation of the encryption of D key and encryption of the encrypted symmetric key, $Enc\_D\ (Pub\_O\ (K))$ namely K as the encrypted using D key. Number 6 contains a receiving of the re-encrypted key. Number 7 contains an access to the encrypted data. Number 8 contains decryption of the received D key using Prv_R to obtain K and decryption of the file using K.

2-2-Blockchain

Blockchain is distributed using a consensus scheme that allows transaction data to be stored securely on the blockchain network after verification and validation processes without third-party interventions [19]. All transaction data will be recorded on all blockchain networks whose procedure is not centralized on a single party. This technology will record all transactions on each node so that it is difficult for irresponsible individuals to modify [20]. Each block has a block header and a transaction block. The block header stores the hash code of the previous block, and the transaction block contains the data stored in that block. The first blockchain block is called the genesis block, which has no parent block [21]. The logic diagram of a blockchain is presented in Figure 4.

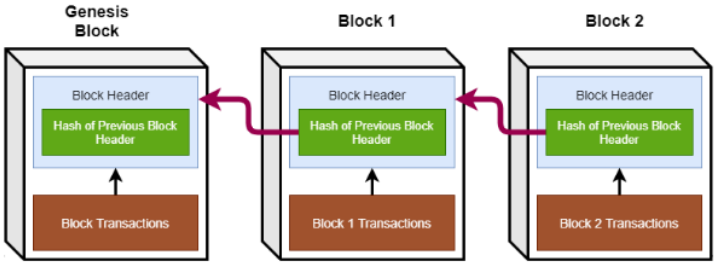

Figure 4. The Blokchain's Logic Diagram

Based on Figure 3, the concept of blockchain security can be seen from how each block stores the hash code of the previous block header. When a hash code changes, the entire chain on the blockchain is broken. This will make it difficult for a security hacker to destroy the data because changing a single transaction data need hacking of all existing data. Blockchain uses the concept of decentralization, which means that every computer connected to the network can form a peer-to-peer network that allows easier tracking of data. If there is an error in one of the computers, the data will be backed up, and the problematic computer will be removed from the network [8]. Blockchain also uses the concept of distribution, namely the stored transaction data will be copied and distributed to all computers connected to the network [21]. An illustration of the distribution of transaction data on a blockchain network is shown in Figure 5.

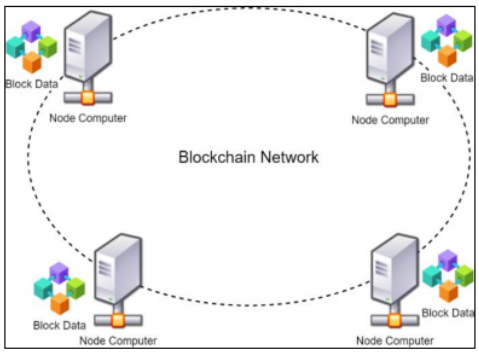

Figure 5. Distributed Data Transaction in the Blockchain Network

The addition of new blocks to the blockchain uses a consensus algorithm. This algorithm is a protocol used to agree on a single data value. The consensus algorithm is mechanistic in nature and automatically synchronizes all transaction data on the blockchain [22]. The mechanism of this algorithm first requires certainty of network status and determination of nodes that can validate transactions. One of the available consensus algorithms is the proof-of-work (PoW) algorithm which requires solving complex math in cryptography through nodes in the network to run together and a random process that provides answers to basic experiments and errors [23].

The proof-of-work algorithm can also be called a mining process because modern computers can quickly make transactions when we experience difficulties solving hash blocks [24]. It can easily be concluded that this algorithm produces a majority decision in a group. The decision of this majority must be accepted by all its members. Members who disagree with the majority decision are considered no longer members. The real implementation in synchronizing transaction data on the blockchain network is when one of the transaction data is different from most of the network data, which is considered invalid data. This creates excellent security in protecting the data of all members.

Computers that can add new blocks are called mining nodes. Each mining node earns a reward as a block of transaction verification called cryptocurrency. One of the most famous cryptocurrencies is Bitcoin (BTC). Several factors should be considered in the valuation of cryptocurrencies, including electricity, internet fees, and ambient temperature to cool mining machines. This cost calculation can be seen in Equation 1 [25]:

$$Price\ of\ Energyday = (\$/kWh \times 24/day \times \in)(\rho/(1000W/kW)) \quad (1)$$

By showing the mining machine's Hashing power in Hash/second, the energy efficiency in Watts/Hash depends on the currency being mined. In addition, the power of mining (Hashing) has a significant value in the mining with more powerful engines, which will give faster results and put a higher initial cost to the investment. The reward in BTC in terms of the mining power is displayed in the following equation [25]:

$$BTC, LTC, (BCH)\_day = (\beta \times \rho \times 86400\ sec/day)/(\delta \times 2^{32}) \quad (2)$$

Here, β is the block reward value in BTC, LTC (for Litecoin), or BCH (for bitcoin cash) according to the blockchain network that is mined, while δ is a difficulty constant of the network. In brief, the ratio between Energy costs and coins received during mining results in the bitcoin production value.

$$V^* = (Energy\ Cost)/(Coins\ Received) \quad (3)$$

If A wants to transfer some amount of money to B, A will use an electronic wallet (a user interface, which can be an application or WebApps) to create transaction records regarding the amount and purpose of the transfer. Furthermore, transactions represent as blocks by the wallet, which later the block transactions are deployed across the Bitcoin network. Several parties (called miners) are competing to validate the block. The proven to be valid blocks will be connected to a "Blockchain" (called Blockchain), in which it has also been validated previously. When the transaction block proposed by A is in the Blockchain, B receives money from A, then the Blockchain gets longer by one block in the network.

2-3-Ethereum

Ethereum is one network that implements a distributed blockchain technology that utilizes smart contracts and is open-source and programmable. It functions as a global network of computers that work together to form a supercomputer, which can create, manage and run decentralized digital applications known as "dapps" [7]. Smart contracts allow members of the ethereum blockchain network to enter agreements and to carry out transactions directly, without third-party intermediaries. Ethereum has its particular cryptocurrency that is called Ethers [26]. This cryptocurrency can be used to share between accounts connected to the Ethereum blockchain network [27]. Ethereum

has its own programming language specifically for creating and deploying smart contracts, namely Solidity. This programming language is a type of high-level language with a contract-oriented characteristic. Solidity is influenced by the C++, Python, and Javascript language paradigms and is designed specifically for the Ethereum blockchain [28].

2-4-Smart Contract

A smart contract contains executable codes that facilitate, implement, and enforce the conditions contained in the agreement to untrusted parties [29]. A smart contract can run automatically when the conditions stated in the contract are met. A smart contract enforces strict rules between the parties in the Ethereum blockchain network without anyone's intervention [16]. Smart contracts extend blockchain to build peer-to-peer collaboration protocols [30]. There is a public interface in every smart contract that can handle the relevant events. This interface is invoked by transactions with proper payload data, and all valid transactions are recorded in the blockchain [31].

2-5-IPFS

IPFS is a peer-to-peer distributed file system that can connect all computing devices to the same file system. Currently, the size of the data stored in a block on the blockchain is only about 1MB, so it is not easy to store large amounts of data [32]. The data uploaded to IPFS is encrypted with a symmetric key, which is encrypted with a public key [33]. IPFS is not only a type of peer-to-peer file sharing system but can also be used for other needs. One of them is the InterPlanetary Name Server (IPNS), which is a distributed alternative to a centralized DNS system. In a website hosted on IPFS, a user can access it via hash codes that are returned by the IPFS network. Hash is used to visiting any websites hosted with the ipfs.io gateway [16]. IPFS is directly supported by Ethereum blockchain as a decentralized form of big data storage that supports smart contracts [34]. Content identifiers, also known as CIDs, are labels to designate content on IPFS. Although it does not show where the content is stored, it forms an address based on the content itself. IPFS uses the SHA-256 algorithm by default, supporting other types of hash algorithms. This research still uses the SHA-256 algorithm in data uploading on IPFS [35].

3- Results

Personal health data are vital and provide high value for developing of a better health care system [36]. The system model to be developed is inspired by research conducted by [37]. It is a system that uses trusted parties to generate and distribute public parameters and private keys to users. The private keys can be used by users such as patients to control access rights related to the use of medical record data diagnosed by doctors. This system uses a cloud server service as a health data storage medium. The framework proposed earlier by [37] is displayed in Figure 6.

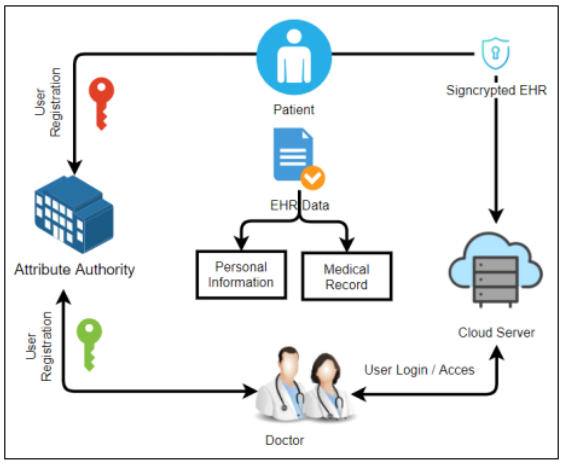

Figure 6. Earlier EHR System Framework

As shown in Figure 6, an EHR system has been developed, including a trusted party, i.e., AA (Attribute Authority), who will provide an excellent private key, and this system involves a signcrypted EHR data process. This research makes an improvement by proposing a system framework using blockchain and IPFS technologies. This framework will answer a problem when an authorized AA manipulates or distributes data without a patient's permission. The offered solution will also overcome the problem of data security stored on cloud servers from attacks such as DDoS. Therefore, to maintain data integrity, we design a new framework focusing on protecting patients' data, especially those related to COVID-19, such as diagnostics or vaccine certificates. Figure 7 shows the development of the previous system, namely [32], by adding the technologies of blockchain and IPFS.

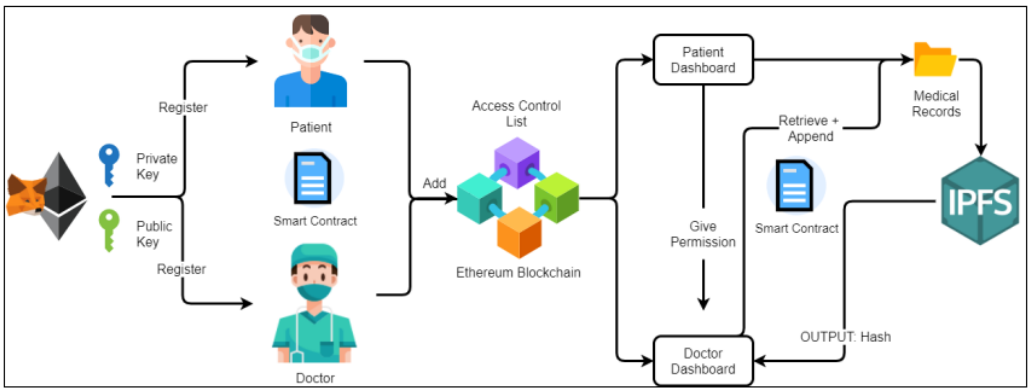

Figure 7. The Proposed EHR Framework

Based on the Figure 7, the proposed EHR framework system has the following workflows:

- System users, both doctors and patients, must first have private keys and public keys from the Ethereum blockchain network. These private keys will be used to enter the EHR system. The access right can be referred to as digital signature. This signature is done cryptographically using a private key and a public key [25], which the two keys are related systematically. This algorithm is called EDSA (Elliptic Curve Digital Signature Algorithm). The following are the steps in using EDSA in the digital signature process.

  a) Take k random integers from the set $\{1, \dots, n-1\}$ where n is the order of the subgroups.
  b) Calculates the point value $P = kG$ where G is the base point of the subgroup.
  c) Calculate the number $r = xP \ mod \ n$ where xP is the x-coordinate of P.
  d) If $r = 0$, choose another k and repeat again.
  e) Calculate $s = k - 1(z + rdA) \ mod \ n$ where dA is the private key and $k - 1$ is the multiplication inverse of $mod \ n$.
  f) If $s = 0$, repeat again by selecting another k.

- The private key must be registered in the MetaMask wallet. Once registered, a user can use the public address of the Ethereum network using the public key.
- To log in to the EHR system, a user must first be registered on the Ethereum blockchain network through metaskating on a browser, such as Chrome or Firefox. The user can choose to register as a patient or a doctor.
- Once registered, the user can login through re-metasking. The system will recognize the type of user, either as a patient or a doctor.
- A patient has the right to give access or not to the doctor who will add his/her medical record data. Granting an access right requires a private key on each medical record data. The patient can also revoke the doctor's access right at any time.
- The doctor who has been granted his/her access rights can add his/her patient's medical record data.
- Medical record data are stored on the distributed IPFS system network.
- IPFS will provide a hash code to the user who wants to view the data, so if the user wants to open the data, he/she must provide the private key again.

The proposed EHR concept is the implementation of research [14] and the improvement to the system made by previous researchers [13], where they created a system only focusing on verifying the covid-19 vaccination certificate. The proposed system stores not only vaccine certificate data but also all activities of the medical record. The system also uses IPFS as distributed storage to maintain data availability as proposed in [15]. The created smart contract concept includes user login authentication, granting permission to access medical record data, and storing medical record data on the blockchain. Functions of the created smart contract can be seen in Figure 8.

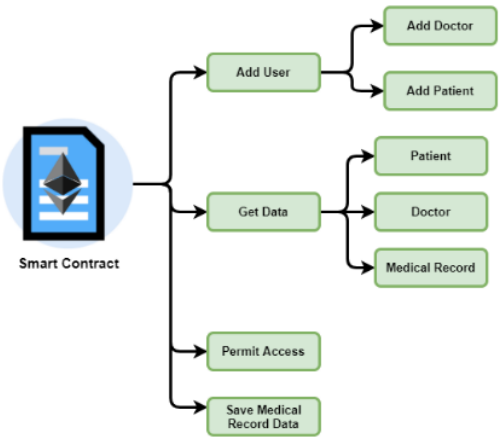

Figure 8. The Smart Contract Concept

Based on Figure 8, a smart contract is programmed for various functional needs such as adding users (doctors or patients), retrieving patients' and doctors' data records and medical records from the blockchain, creating a function to add permission for patients if the data are seen and added by doctors as well as a function to add medical record data to the blockchain.

3-1-EHR System Analysis

We use the operating system of Windows 10 64 Bit with 16 RAM and 512 SSD. The smart contract is built using the Solidity programming language. From the backend side, the researchers use Python programming language to connect with IPFS. From the frontend, we use the Javascript library, namely *React*. The smart contract is then deployed with truffle and Ethereum test using ganache. Based on Figure 6, the first stage in this system is that users, both patients and doctors, must have private and public keys to enter the EHR system. MetaMask is used in the authentication concept, in which MetaMask will ask for private and public keys to be registered in the authentication medium. Figure 9 illustrates how MetaMask authenticates a patient's or doctor's Ethereum account.

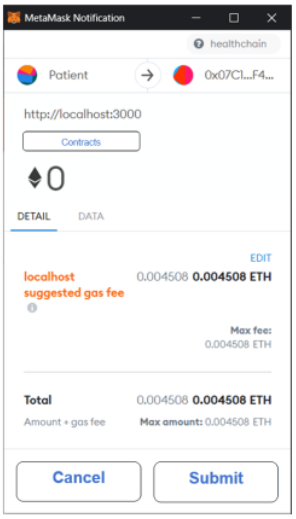

Figure 9. Access to the EHR System using MetaMask

To log in to the EHR system, a user must have the private and public keys to be registered in the MetaMask. Users who are not registered in the MetaMask will not be able to access the EHR system. When the login process is executed, MetaMask will provide a notification in the form of interaction to the user whether to confirm or deny this interaction. If the confirmation button is selected, a gas fee will be charged from Ethereum and if its balance is sufficient, the user can enter the system.

Doctors who have obtained permission from patients to provide medical record data can then upload medical record data which will be stored on the network of blockchain and IPFS. Permission can be granted when a patient adds a doctor's public key to the EHR system. Patients have a right to add or remove their access rights from their respective doctors. The flow of adding medical record data can be seen in Figure 10.

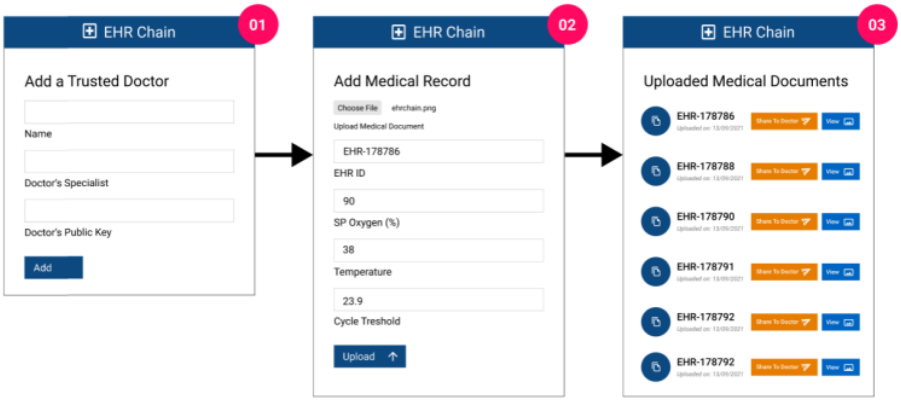

Figure 10. Medical Record Input

The following is the pseudocode for smart contract functions that are used in the EHR system. Algorithm 1 is a function to add trusted doctors to access and add medical record data. Algorithm 2 is a function to store medical record data on the Ethereum blockchain.

Algorithm 1: Add Doctor to Access Patient Data

```
Input: doc_name, doc_specialist, doc_public_key
1   if Patient is registered in smart contract then
2   |  Add Doctor with doc_name, doc_specialist and doc_public_key
    |  Doctor has acces Patient data
3   end
4   else
5   Show an Error Message in System
6   end
```

Algorithm 2: Save Medical Records

```
Input: ehr_id, sp_oxygen, temperature,ct_value, ipfs_hash
1   if Patient is registered in smart contract then
2     |  if   Doctor has access then
3     |   |   Add medical records with ehr_id, sp_oxygen, temperature, ct_value, ipfs_hash
4     |   |   Data saved to blockchain
5     | end
6     | else
7     |    | Show an Error Message in System
8   end
9   else
10        |  Show an Error Message in System
11  end
```

Algorithm 1 is a function to add doctors by patients. This algorithm explains how doctors are given access rights to patients' medical record data. First of all, to use this function, a patient must first be registered on the smart contract. If a patient has been registered, the patient can add permission for his medical record data to a doctor. Algorithm 2 is a function to store medical record data created by doctors on the Ethereum blockchain. This algorithm will check whether or not a patient has been registered on the MetaMask. If a patient is already registered, the system will re-examine it to determine whether or not a doctor has obtained permission to access a patient's medical record data. A doctor who has obtained the permission can then add medical record data stored on the Ethereum blockchain.

Medical record data added by doctors are stored on the Ethereum blockchain and IPFS distribution network. The data are stored in a hash code using the SHA-256 algorithm as the standard encryption of IPFS, distributed on a peer-to-peer network. Figure 11 [38] shows a peer-to-peer network in the current system.

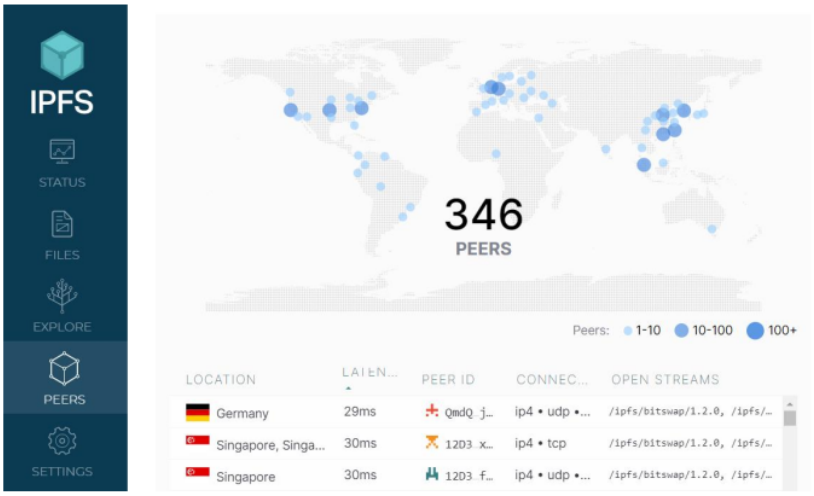

Figure 11. Peer to Peer Network of IPFS

3-2-IPFS Network Data during the Process of Data Addition in the EHR System

When the EHR system is running, there are several commands to add patients' data, doctors' data, medical record data, and all data that have been uploaded to IPFS. By utilizing the system monitoring tool found on the computer used for testing the system, we obtain information regarding data packets in sending and retrieving those data packets. Based on the IP address connected to the IPFS network when the system is running, the network activity data are acquired and can be seen in Table 1. The data obtained from the computer monitor are used for the EHR system testing. From 346 computers connected to the IPFS network, the samples of network activity data are taken from the best 18 computers. The process of data distribution when the system is used can be seen clearly from the existence of data transmission (send) and data reception (receive) from each computer connected to the IPFS network in bytes per second. The highest data transmission activity (send) is carried out by the computer with the IP Address of 117.174.25.133, whose value is 225 B/sec; while the highest receiving activity is done by the computers with the IP Address of 183.100.178.46 and 220.246.207.162, whose value is 179 B/sec. This activity has proven the existence of a data distribution process on the IPFS network. Table 1 shows that there are fewer bandwidth usages than previous research conducted by [13] in the research of system making to verify the authenticity of Covid-19 vaccine certificates using Blockchain technology.

Figure 11 is a graphical visualization of when the EHR system is run based on data packets sent and received by each IP Address. Based on the graph, it can be seen that there is a process of sending and receiving data packets that run dynamically from every computer connected to the IPFS network when the EHR system is running.

In the network, network latency is one of the main parameters to consider when designing and implementing remote monitoring for security and system events [39]. Therefore, latency is a parameter used to measure the delay that occurs. It is usually measured as the round-trip delay, the time it takes for information to reach its destination and return. When transactions are made on the EHR system, we encounter latency data when the computer used for testing is connected to multiple computers on the IPFS peer-to-peer distribution network. We can see that the local address (the computer used as a test) is connected to the remote address (the computer in the IPFS network) and provides latency data in milliseconds. The Transmission Control Protocol (TCP) Connection tool from the local address of the computer is applied. From the top 10 latency data that have been sorted, it is found that the highest latency occurs at the remote IP Address 68,183.72.119 with a latency value of 4,294,918,865 ms. The data can be seen in Table 2.

Table 1. IPFS Network Activity

IP Address IPFS	Send (B/sec)	Receive (B/sec)
117.174.25.133	225	0
183.100.178.46	45	179
220.246.207.162	44	179

10.244.5.98	221	0
97.85.175.122	62	157
111.19.254.170	197	22
10.0.66.159	218	0
10.0.69.131	218	0
61.85.208.120	45	172
183.31.18.13	214	0
179.119.153.77	45	170
220.87.199.88	45	169
121.172.209.22	44	167
51.195.117.245	102	102
119.196.28.199	39	164
139.178.89.157	201	0
190.203.145.193	41	155
159.65.108.245	194	0
60.244.159.238	39	154
218.0.186.175	193	0
192.168.1.5	193	0
135.181.155.219	71	114

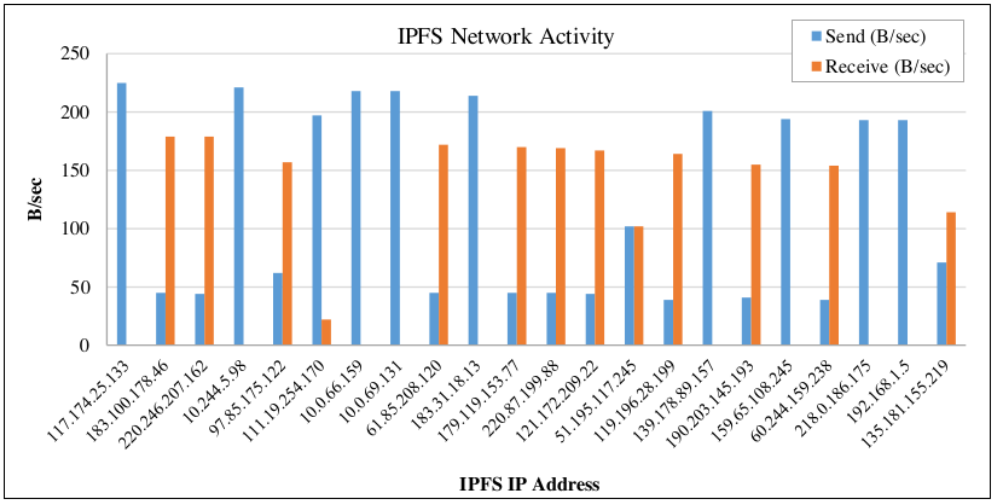

Figure 12. IPFS Network Activity

Table 2. Latency Monitor

Local Address	Remote Address IPFS	Latency (ms)
192.168.3.62	68.183.72.119	4.294.918.865
192.168.3.62	211.184.4.246	2.147.483.584
192.168.3.62	111.33.31.67	2.147.482.692
192.168.3.62	109.194.35.249	1.073.740.811
192.168.3.62	42.2.69.32	159.072.565
192.168.3.62	136.244.110.105	113.021.078
192.168.3.62	101.99.164.67	104.755.104
192.168.3.62	219.77.107.12	45.210.135
192.168.3.62	1.65.152.63	39.768.159
192.168.3.62	83.136.84.248	5.457.344

3-3-Ethereum Gas Usage Data for Every Data Addition

Every data transaction, whether data addition or data viewing on the Ethereum blockchain network, will consume the transaction gas. This gas is a calculation that shows the cost of an action performed by the Ethereum blockchain. This fee can be considered a form of reward for using the computing capabilities of the Ethereum blockchain. Table 3 shows the information of hash transactions generated and the gas used when ten transactions are performed in the EHR system.

Table 3. Gas Usage Data

Block Number	Hash Transaction	Gas Used	Event
1	TX 0xd877d63ab5b5551d5e694995a68a4b92bc337f059fc80c372790c2495cf49abe	164391	Contract Creation
2	TX 0x7f61ce578c5130817125af9609f5728edbf6f5221515ec0a89f404518e376bf1	42341	Migration Contract
3	TX 0x9cfe200bbdd9c9a6f87c4dc231be58e00f690d58249963b438fb61f09e84a1d0	1094493	Contract Creation
4	TX 0x5579bf304af8875e4944f4c9eaeb055dc01f83fdeaf1de9e21ab4e67ce9d6ff4	135133	Call Contract Authentication
5	TX 0x5579bf304af8875e4944f4c9eaeb055dc01f83fdeaf1de9e21ab4e67ce9d6ff4	135133	Call Contract Save Data
6	TX 0x606778bc263ad3fb1b8a266eb54df7e6a1c462e82420f52793244e336625a6e0	135205	Call Contract Save Data
7	TX 0x80e3803f71d36c116b39fd28ad040fff2a327e11fa3e7ddda9cbdefd5eb0a8e3	135205	Call Contract Save Data
8	TX 0xbf8d21687f1aa52e96fffa9a3a50a1dfe2e636fc5bdd45d8aa96cc3e3e35eba1	135205	Call Contract Save Data
9	TX 0x259480ea07d877f9c240cf2dbb969d45758c656c1b76d26f0e7e87e230e76eac	135205	Call Contract Save Data
10	TX 0x8d053a42122af6b38e936c42316f799063e43c40a3a61bdac9c3244e11cd7e00	135205	Call Contract Save Data

Based on Figure 13, among the data of the ten transactions carried out, the largest gas is 164391 and 1094493, respectively, when the smart contracts are created. The gas released during the smart contract migration is 42341. The use of gas is 135205 to add or view the data when the contract is called. The gas used is on the Wei scale or with a scale of 1 ETH, which is equal to 108 Wei. The highest gas usage in the design of this system is 164391 Wei so that if it is converted into Ether, it is around 0.00000000000164391 ETH. The conversion to USD is 0.0000000006 USD. When compared with [14], it can be concluded that the transaction costs incurred are less because research [14] found the average result of using transactions in the blockchain used is 0.01 – 1 USD.

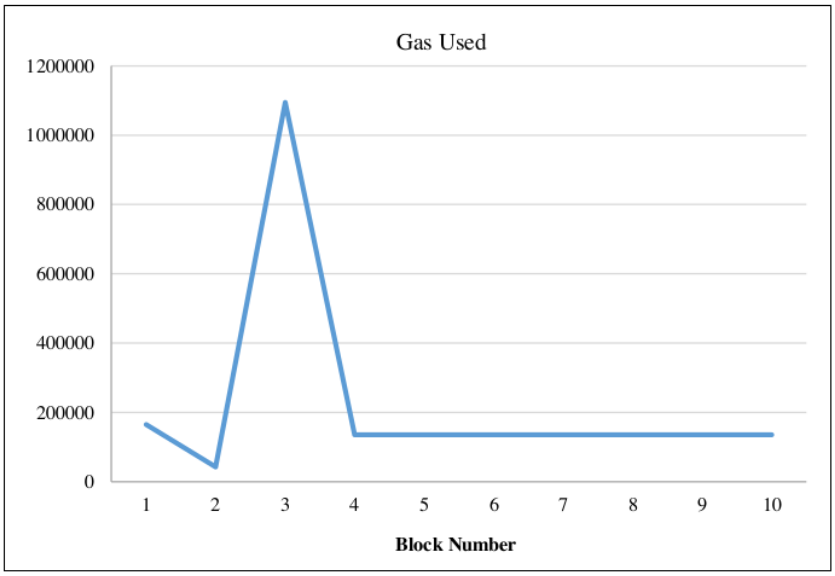

Figure 13. Gas Used

3-4-Table of Medical Record Hashing Data on the IPFS

The stored data on the IPFS distributed network have been hashed with the SHA256 algorithm. From 10 trials of storing medical record data on the IPFS network, we found the result, namely the IPFS CID as the hash of each added data. IPFS CID data from each stored medical record data can be seen in Table 4.

Table 4. Results of Medical Record Hashing on IPFS

No	Medical Record ID	IPFS CID
1	EHR-178786	QmeG9MV9sthg6ApneXtgTg4mSDckQtEhCRLb8Q2sAD258f
2	EHR-178788	QmS9DN2nLKeNGP4npuFSDxNc7uNJZQoX2VJtn5yQH8LBSh
3	EHR-178790	QmXaaWof6TrGzMYZyezGNyHRGVLkbWWWC111BiZwcd96TY
4	EHR-178791	QmSfu6AEnvCsVpxL9A52b7DUFYDbXmAGmAn8gWn8i1ec7k
5	EHR-178792	QmYeJg8yCjrxcoaFZTAh77YBk4jksHi1QSwbuRNE6pU63h
6	EHR-178791	QmYeJg8yCjrxcoaFZTAh77YBk4jksHi1QSwbuRNE6pU63h
7	EHR-178795	QmPp6nBsdbxW7NXpfrwdn6k4RrxMDbuAhwzbMTt7vCfYEM
8	EHR-178796	QmSEq6HadVjRbush7B1E9EbemqKX7R6kWAyN1Rct1pMoMu
9	EHR-178798	QmdobPg2nP21eUnra4nqhzX4jtA3CExyzAUjnNkag6tFkj
10	EHR-178799	QmSEq6HadVjRbush7B1E9EbemqKX7R6kWAyN1Rct1pMoMu

3-5-CPU Monitor During the EHR System is Running

CPU performance test results are shown in Table 5 and Figure 14. Several parameters observed, namely threads, CPU Load, and Average CPU. Threads function as a connecting mechanism, helper, and data network to be processed or entered in every stage of the process that occurs in the CPU, namely fetch or waiting for execution, decode, execute, and writeback commands. The CPU load is the number of processes currently executed by the CPU or waiting to be executed by the CPU. Average CPU is the number of average processes currently being executed or waiting to be executed during the last 1, 5, and 15 minutes. The results show that the CPU execution speed ranges from 12.98 GHz to 14.08 GHz. The performance results show that running the proposed method requires a CPU with a fast data execution speed so that the programming takes a shorter time.

Table 5. CPU Monitor

Threads	Load CPU	Average CPU (Ghz)
46	19	14.08
47	16	14.24
47	11	14.25
47	15	14.47
46	14	11.47
46	16	11.34
47	13	11.76
47	12	12.06
47	13	12.32
47	12	12.32
46	14	12.56
46	13	13.84

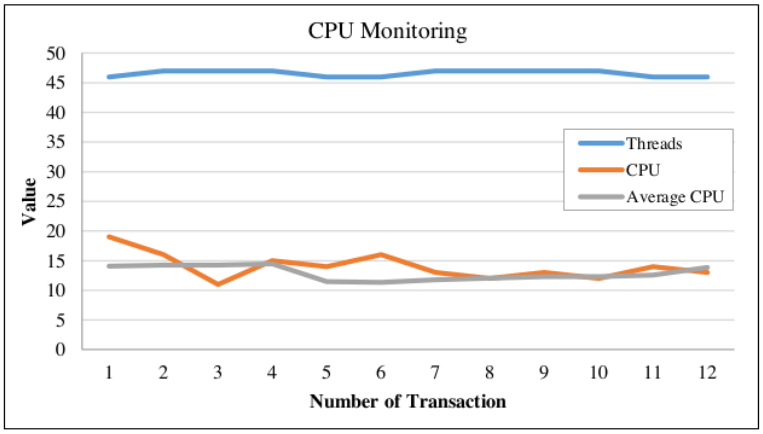

Figure 14. CPU Monitoring During Data Transaction

3-6-Bandwidth Overtime during the EHR Data Storage

In the IPFS network monitoring application called IPFS Desktop which can be accessed by first activating the IPFS daemon command, we create a graph of bandwidth information when the HER system is used to send the transaction data to the blockchain and IPFS network. The data can be monitored in real-time, based on the sending of transactions. The blue indicates the incoming bandwidth and the orange indicates the outgoing bandwidth, as seen in Figure 15.

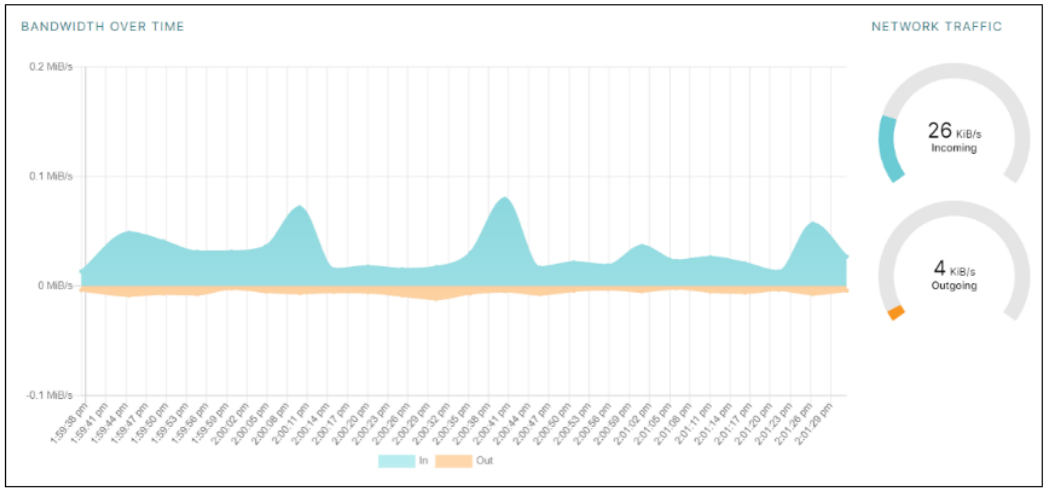

Figure 15. Bandwidth during Data Transaction

4- Discussion

We have successfully designed a Blockchain and IPFS-based EHR system from the experimental data and the proposed system architecture. This system utilizes the Ethereum blockchain as the main technology for a decentralized network of medical record data to keep the COVID-19 diagnostic data safe from the attempts of manipulating or destroying the data. As a solution for maintaining data availability, we also apply the technology of distributed data storage, namely IPFS, as a solution to the problem of traditional cloud storage, which tends to have the potential to lose data at any time.

The obtained results show that several main highlights distinguish it from previous studies, such as the refinement of the system by Sestrem et al. (2021) [13], who only focused on the integrity of vaccine certification. Simultaneously, we improve all series of the medical record system related to COVID-19 diagnosis. The system is also designed by applying the role of the researchers [15]. The results also show more efficient bandwidth use when this system is compared to Sestrem et al. (2021) [13]. The transaction costs used in the framework of this system are more efficient than those of the research conducted by Hernández-Ramos et al. (2021) [14].

Advantages of the system: Medical record data can be stored securely on the Ethereum and IPFS blockchain network. Data can only be stored if there is an agreement between the patient and the doctor through authentication in the form of a private key. Data is well distributed across all computers connected to the Ethereum and Blockchain IPFS so that data availability will always be maintained. This data availability certainly overcomes a common problem that often occurs in traditional cloud storage which often loses data due to a provider's authority to delete some data to save storage space. The data that has been entered also cannot be manipulated because blockchain is immutable (cannot be deleted or changed) in the storage process.

Weaknesses of the system: This system is very costly in its implementation because it has to take advantage of Blockchain architecture. Any additional transactions will incur gas fees from the Ethereum blockchain. This gas cost can also continue to increase over time, so it needs special consideration in utilizing this technology. The maintenance of this system also needs a special expertise in Blockchain knowledge. There are only a few developers in blockchain technology, so the cost to pay people with this expertise will be very high. The stored data also cannot be changed or deleted when it enters the blockchain network, therefore the data input process must be done conscientiously because the stored data will be the final data.

5- Conclusion

The implementation of smart contracts on blockchain technology and the IPFS network environment is critical to improving the security of the medical record system. Blockchain can run well without special needs. It only needs personal computers that support IPFS networks to connect to the Internet to implement the blockchain technology. Based on the results of the tests that have been carried out, it can be concluded that the blockchain technology architecture on the IPFS network can be implemented to improve the performance of the medical record system.

For further research, it is suggested that the development can be carried out not only for the COVID-19 diagnosis data and Vaccine Certificates but also for various health data that are private and require better data security. This system framework can also be applied or tested in several fields, such as drug distribution in the pharmacy department and various distributions of the other medical equipment needs (supply chain). Some developments can also be carried out on a web-based blockchain network, such as making special programs to combine mobile interfaces to ease the development, so users can utilize blockchain technology more easily. To make it easy for users to do a medical check-up, a barcode system can synchronize the blockchain network. Artificial intelligence technology can also be integrated with the blockchain network to be more accessible.

6- Declarations

6-1-Author Contributions

Conceptualization, I.R., T.A., R.S., P.W. and A.M.; methodology, I.R., T.A., and R.S.; software, P.W. and A.M.; validation, I.R., T.A., R.S., P.W. and A.M.; formal analysis, I.R., T.A., and R.S.; investigation, I.R., T.A., and R.S.; resources, I P.W. and A.M.; data curation, P.W. and A.M.; writing—original draft preparation, I.R., T.A., R.S., P.W. and A.M.; writing—review and editing, I.R., T.A., R.S., P.W. and A.M.; project administration, I.R., T.A., R.S., P.W. and A.M.; funding acquisition, I.R. All authors have read and agreed to the published version of the manuscript.

6-2-Data Availability Statement

The data presented in this study are available in article.

6-3-Funding

This Research is funded by the Indonesian Ministry of Education, Culture Research and Technology and Universitas Ahmad Dahlan, Yogyakarta, Indonesia.

6-4-Acknowledgements

Authors would like to express appreciation and gratitude to Universitas Ahmad Dahlan for funding this research and World Class Professor (WCP) Program 2021 managed by the Indonesian Ministry of Education, Culture Research and Technology (Research Grant No. 2817/E4.1/KK.04.05/2021).

6-5-Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Sarno, R., Sungkono, K. R., Taufiqulsa'di, M., Darmawan, H., Fahmi, A., & Triyana, K. (2021). Improving efficiency for discovering business processes containing invisible tasks in non-free choice. In Journal of Big Data (Vol. 8, Issue 1). doi:10.1186/s40537-021-00487-x.

[2] Ahmad, T., & Samudra, Y. (2020). Reversible data hiding with segmented secrets and smoothed samples in various audio genres. Journal of Big Data, 7(1), 1-19. doi:10.1186/s40537-020-00360-3.

[3] Reuters, (2021). Privacy alarm in Indonesia over president's leaked vaccine certificate. Available online: https://www.reuters.com/world/asia-pacific/privacy-alarm-indonesia-over-presidents-leaked-vaccine-certificate-2021-09-03/ (accessed on September 2021).

[4] BBC. (2021). Covid-19 death rate: Central and regional government data difference reaches 19,000 cases, "green outside red inside. Available online: https://www.bbc.com/indonesia/indonesia-57971840 (accessed on September 2021).

[5] Annaka, S. (2020). Political Regime and Suspected COVID-19 Death Data Manipulation. doi:10.33774/apsa-2020-1xfvz.

[6] Sarno, R., Sinaga, F., & Sungkono, K. R. (2020). Anomaly detection in business processes using process mining and fuzzy association rule learning. Journal of Big Data, 7(1). doi:10.1186/s40537-019-0277-1.

[7] Fatokun, T., Nag, A., & Sharma, S. (2021). Towards a blockchain assisted patient owned system for electronic health records. Electronics (Switzerland), 10(5), 1–14. doi:10.3390/electronics10050580.

[8] De Aguiar, E. J., Faiçal, B. S., Krishnamachari, B., & Ueyama, J. (2020). A Survey of Blockchain-Based Strategies for Healthcare. ACM Computing Surveys, 53(2), 1–27. doi:10.1145/3376915.

[9] Sun, J., Ren, L., Wang, S., & Yao, X. (2020). A blockchain-based framework for electronic medical records sharing with fine-grained access control. PLoS ONE, 15(10 October), 239946. doi:10.1371/journal.pone.0239946.

[10] Balistri, E., Casellato, F., Giannelli, C., & Stefanelli, C. (2021). BlockHealth: Blockchain-based secure and peer-to-peer health information sharing with data protection and right to be forgotten. ICT Express, 7(3), 308–315. doi:10.1016/j.icte.2021.08.006.

[11] Qu, J. (2022). Blockchain in medical informatics. Journal of Industrial Information Integration, 25, 100258. doi:10.1016/j.jii.2021.100258.

[12] Liao, C. H., Lin, H. E., & Yuan, S. M. (2020). Blockchain-Enabled Integrated Market Platform for Contract Production. IEEE Access, 8, 211007–211027. doi:10.1109/ACCESS.2020.3039620.

[13] Sestrem Ochôa, I., Reis Quietinho Leithardt, V., Calbusch, L., De Paz Santana, J. F., Delcio Parreira, W., Oriel Seman, L., & Zeferino, C. A. (2021). Performance and Security Evaluation on a Blockchain Architecture for License Plate Recognition Systems. Applied Sciences, 11(3), 1255. https://doi.org/10.3390/app11031255.

[14] Hernández-Ramos, J. L., Karopoulos, G., Geneiatakis, D., Martin, T., Kambourakis, G., & Fovino, I. N. (2021). Sharing Pandemic Vaccination Certificates through Blockchain: Case Study and Performance Evaluation. In W. Li (Ed.), Wireless Communications and Mobile Computing (Vol. 2021, pp. 1–12). doi:10.1155/2021/2427896.

[15] Kalla, A., Hewa, T., Mishra, R. A., Ylianttila, M., & Liyanage, M. (2020). The Role of Blockchain to Fight against COVID-19. IEEE Engineering Management Review, 48(3), 85–96. doi:10.1109/EMR.2020.3014052.

[16] Kumar, S., Bharti, A. K., & Amin, R. (2021). Decentralized secure storage of medical records using Blockchain and IPFS : A comparative analysis with future directions . Security and Privacy, 4(5). doi:10.1002/spy2.162.

[17] Nasir, M. H., Arshad, J., Khan, M. M., Fatima, M., Salah, K., & Jayaraman, R. (2022). Scalable blockchains — A systematic review. Future Generation Computer Systems, 126, 136–162. doi:10.1016/j.future.2021.07.035.

[18] Hasan, H. R., Salah, K., Jayaraman, R., Arshad, J., Yaqoob, I., Omar, M., & Ellahham, S. (2020). Blockchain-Based Solution for COVID-19 Digital Medical Passports and Immunity Certificates. In IEEE Access 8, 222093–222108. doi:10.1109/ACCESS.2020.3043350.

[19] Sarmah, S. S. (2018). Understanding Blockchain Technology. Computer Science and Engineering, 8(2), 23–29. doi:10.5923/j.computer.20180802.02.

[20] Satapathy, U., Mohanta, B. K., Panda, S. S., Sobhanayak, S., & Jena, D. (2019). A Secure Framework for Communication in Internet of Things Application using Hyperledger based Blockchain. 2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019. doi:10.1109/ICCCNT45670.2019.8944811.

[21] Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. In Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017 (pp. 557–564). doi:10.1109/BigDataCongress.2017.85.

[22] dos Santos, R. P. (2019). Consensus Algorithms: A Matter of Complexity? Between Science and Economics, 147–170. doi:10.1142/9781786346391_0008.

[23] Lucas, B., & Paez, R. V. (2019). Consensus algorithm for a private blockchain. ICEIEC 2019 - Proceedings of 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication, July, 264–271. doi:10.1109/ICEIEC.2019.8784500.

[24] Agung, A. A. G., Dillak, R. G., Suchendra, D. R., & Robbi, H. (2019). Proof of work: Energy inefficiency and profitability. Journal of Theoretical and Applied Information Technology, 97(5), 1623–1633.

[25] Muchtadi-Alamsyah, I., Imdad, M. T., & Sutikno, S. (2020). Group signature based ethereum transaction. International Journal on Electrical Engineering and Informatics, 12(1), 19–32. doi:10.15676/ijeei.2020.12.1.2.

[26] Gupta, S., & Sadoghi, M. (2019). Blockchain Transaction Processing. Encyclopedia of Big Data Technologies, 366–376. doi:10.1007/978-3-319-77525-8_333.

[27] Shahnaz, A., Qamar, U., & Khalid, A. (2019). Using Blockchain for Electronic Health Records. IEEE Access, 7, 147782–147795. doi:10.1109/ACCESS.2019.2946373.

[28] Hegedűs, P. (2019). Towards Analyzing the Complexity Landscape of Solidity Based Ethereum Smart Contracts. Technologies, 7(1), 6. doi:10.3390/technologies7010006.

[29] Alharby, M., Aldweesh, A., & Van Moorsel, A. (2018). Blockchain-based Smart Contracts: A Systematic Mapping Study of Academic Research (2018). International Conference on Cloud Computing, Big Data and Blockchain, ICCBB 2018, 1–6. doi:10.1109/ICCBB.2018.8756390.

[30] Khatoon, A. (2020). A Blockchain-Based Smart Contract System for Healthcare Management. Electronics, 9(1), 94. doi:10.3390/electronics9010094.

[31] Hu, B., Zhang, Z., Liu, J., Liu, Y., Yin, J., Lu, R., & Lin, X. (2021). A comprehensive survey on smart contract construction and execution: paradigms, tools, and systems. Patterns, 2(2), 100179. doi:10.1016/j.patter.2020.100179.

[32] Ye, H., & Park, S. (2021). Reliable vehicle data storage using blockchain and ipfs. Electronics (Switzerland), 10(10), 2021. doi:10.3390/electronics10101130.

[33] Battah, A. A., Madine, M. M., Alzaabi, H., Yaqoob, I., Salah, K., & Jayaraman, R. (2020). Blockchain-based multi-party authorization for accessing iPFS encrypted data. In IEEE Access 8, 196813–196825. doi:10.1109/ACCESS.2020.3034260.

[34] Ethereum. "Decentralized Storage". Available online: https://ethereum.org/en/developers/docs/storage/ (accessed on December 2021).

[35] IPFS. (2021). Content addressing and CIDs IPFS. Available online: https://docs.ipfs.io/concepts/content-addressing (accessed on December 2021).

[36] Liang, X., Zhao, J., Shetty, S., Liu, J., & Li, D. (2017, October). Integrating blockchain for data sharing and collaboration in mobile healthcare applications. In 2017 IEEE 28th annual international symposium on personal, indoor, and mobile radio communications (PIMRC) (pp. 1-5). IEEE. https://doi.org/10.1109/PIMRC.2017.8292361.

[37] Alkhushyni, S. M., Alzaleq, D. M., & Kengne, N. L. G. (2019). Blockchain technology applied to electronic health records. EPiC Series in Computing, 63, 34–42. doi:10.29007/2x3r.

[38] IPFS. (2021). IPFS powers the Distributed Web. Available online: https://docs.ipfs.io/install/ipfs-desktop/ (accessed September 2021).

[39] Kufel, L. (2015). Network latency in systems event monitoring for multiple locations. Scientific Programming (Vol. 2015, pp. 1–6). doi:10.1155/2015/371620.

HASIL CEK_Developing Data Integrity in an Electronic Health Record System using Blockchain and InterPlanetary File System (Case Study: COVID-19 Data)

ORIGINALITY REPORT

5%	7%	0%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

5%

★ www.researchgate.net

Internet Source

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		