A blockchain-based architecture for securing electronic health record systems

Guang Yang1 | Chunlei Li2 | Kjell E. Marstein2

1Department of Electrical Engineering, Western Norway University of Applied Sciences, Bergen, Norway
2Department of Informatics, University of Bergen, Bergen, Norway

Correspondence
Guang Yang, Department of Electrical Engineering, Western Norway University of Applied Sciences, 5063 Bergen, Norway. Email: guang.yang@hvl.no

Present Address
Guang Yang, Western Norway University of Applied Sciences, Postbox 7030, 5020 Bergen, Norway

Funding information
UH-nett Vest in Norway, Grant/Award Number: 720025

Summary
This paper presents a blockchain-based architecture for our current electronic health record (EHR) systems. Being built on top of existing databases maintained by health providers, the architecture implements a blockchain solution to ensure the integrity of data records and improve interoperability of the systems through tracking all events that happen to the data in the databases. In this proposed architecture, we also introduce a new incentive mechanism for the creation of new blocks on the blockchain. The architecture is independent of any specific blockchain platforms and open to further extensions; hence, it potentially fits in with other electronic record systems that require protection against data misuse.

KEYWORDS
electronic health record, incentive mechanism, permissioned blockchain, threshold cryptography

1 | INTRODUCTION

National health professionals operate various healthcare services at different locations. Usually, a user visits more than one health professional, eg, general practitioner, specialists, clinics, pharmacies, etc, for different needs. Figure 1 shows a scenario in the current healthcare systems, where records issued by a health provider are stored locally in the provider’s data system. The provider handles all the management and maintenance of the data, most of which are database related tasks. Example tasks include SQL reporting, the database backup, system recovery, etc. It is not uncommon to utilize the storage server solutions, and some of the providers have to take the physical data storage into their consideration also. Only this provider has the edit right to the records. Provided with access rights to the databases, users can query their records from different providers. Providers can also query records of a common user from other providers when they need, eg, consulting related records for making a diagnosis. Due to the lack of a standardized integration of the system, management of the entire health system is so far a significant challenge. The demand for multiple access from users and health providers have also raised the issues of security, interoperability, and privacy of the data.

1.1 | Electronic health record (EHR)

Elektronisk pasientjournal (EPJ) is the Norwegian Electronic Patient Record format that records health information of users and has been primarily used by health providers in Norway. An EPJ system stores EPJ files in databases and offers an interface for registering, searching, and displaying information from the EPJ files, etc. Such databases are (partly) open to users and other providers with different specified permissions. Currently, this EPJ system, like the electronic health record (EHR) systems in many other countries, is fragmented across health providers and the entire system has yet to be integrated. The health records of a user are maintained in different health providers’ databases; thus, providers cannot have a comprehensive overview of all the records of a single user.

This is an open access article under the terms of the Creative Commons Attribution Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Concurrency and Computation: Practice and Experience published by John Wiley & Sons Ltd.
1.2 Blockchain

In recent years, the blockchain has drawn a lot of attentions from both academic and industrial fields, and its applications have been penetrated from financial domains into in some of the information and communications technology areas. The concept of blockchain technology was first introduced by Satoshi Nakamoto in the well-known paper on the Bitcoin. The blockchain technology is a genuine combination of existing techniques from distributed computing, cryptography, game theory, etc, and it promises to radically transform the way digital assets are exchanged among untrusted participants, securely tracking the ownership of these assets without the control of a central authority.

At the heart of blockchain technology lies a distributed ledger that is collaboratively maintained in an immutable manner. The ledger is a chain of blocks that are time-stamped and linked by cryptographic hash functions, as each new block holds a hash value of the previous block. A block is a data structure consisting of a header part and a data part: the header contains a hash value of the previous block, a time stamp, Merkle root of the data, and some optional fields; the data part contains application-specific data assets, e.g., financial transactions, logic codes, etc. The chain is continuously growing, and new blocks are chronologically appended to the end of the chain. The cryptographic linkage among blocks in the chain contributes to the append-only property of the blockchain. The blockchain is duplicated across the distributed network and maintained by a group of nodes, known as mining nodes in Bitcoin, which take the responsibility to create and append new blocks to the chain according to a predefined consensus mechanism. The consensus mechanism is crucial to the uniqueness of the blockchain throughout the whole network. The Bitcoin protocol introduced a consensus mechanism by voting the CPU power of each node, known as proof-of-work (PoW). However, the PoW consumes a massive amount of CPU power to maintain the Bitcoin network. Therefore, other consensus mechanisms have been proposed in recent years to remedy the drawback of PoW, including proof-of-stake (PoS), delegated PoS (dPoS), proof-of-memory, proof of time-elapse, practical Byzantine fault tolerance (PBFT), etc.

With the broader deployment of blockchain technology, the needs for functional data assets with automatized logics, termed a smart contract, are highlighted. The idea of a smart contract was first described by Nick Szabo, where he introduced the objectives and principles for the design of smart contracts to formalize and secure relationships over computer networks by digital means. The contracts are implemented in distributed ledgers, and the execution processes are automated. The idea has been implemented in many blockchain projects, such as Ethereum and Hyperledger, as a piece of computer code that rules relationships of nodes in the network storing on the blockchain. Since the smart contract operates on the publicly shared ledger in a distributed network, the execution processes are immutable once they have been agreed to run.

1.3 Blockchains for E-healthcare systems

The E-healthcare system is believed to be one of the fields where blockchain has great potential due to its inherent characteristics, especially for the management of electronic health records. Significant research efforts have been made in this direction in the last few years. Azaria et al, in 2016, proposed a decentralized records management system, termed MedRec, which was built on the Ethereum platform and utilized Ethereum’s smart contracts to create an intelligent representation of existing medical records stored within individual nodes on the network. Patients have control over their medical records across providers and treatments sites in this system. While medical stakeholders, such as researchers, public health authorities, etc, are incentivized to participate in the mining of the blockchain. The blockchain ledger keeps an auditable history of medical interactions of patients, providers, and regulators. This solution brings questions about to which levels that patients should own their medical information and to which degrees that the data can be shared. The related regulations have been discussed under various circumstances, and in most cases, are decided by health authorities. For this reason, the available variation of the system should also be taken into the consideration that patients have not full control over their medical records and data are shared at different levels among all medical stakeholders. In the same year, Yue et al proposed a blockchain-based smartphone application architecture, termed Healthcare Data Gateway (HDG), to improve the privacy aspect of sharing private patient data. The proposed architecture consists of three layers: raw data are encrypted and stored in the private blockchain cloud at the storage layer; database management, including data access management is placed at the management layer; and usage layer is where health care data are utilized, e.g., for the medical records system, data analytics, etc. In this architecture, a private blockchain is implemented. Unlike a public blockchain that anyone can join the network, the private blockchain is a permissioned blockchain with restrictions.
Another approach to improving privacy issues when sharing healthcare data between different stakeholders was proposed by Peterson et al.11 The article discusses the challenges of choosing common interoperable data syntaxes and security protocols and approaches to solving these challenges. A new consensus mechanism, termed “Proof of Interoperability,” along with an algorithm describing the process, is proposed. Data must conform to both structural and semantic constraints to be verified to reach the consensus. In this study, the EHR system utilizes Fast Healthcare Interoperability Resources (FHIR) standard data format, and the proposed consensus mechanism correspondingly uses the profile of the FHIR as the interoperability constraints. For different formats that other EHR systems are utilizing, a set of structural and semantic constraints need to be designated beforehand to implement this consensus mechanism.

Subsequently, Kuo et al introduced blockchain to the biomedical/healthcare domains.12 The article detailed the benefits of applying blockchain in biomedicine/healthcare by comparing it with traditional distributed databases. It also discussed the potential challenges and proposed solutions for adopting blockchain technologies in these domains. This work gives a general introduction about blockchain technologies to the biomedical and health care informatics researchers. Liang et al13 proposed a user-centric health data sharing solution by utilizing a decentralized and permissioned blockchain. In this work, a mobile application was deployed to collect health data from personal wearable devices. The collected data were synchronized to the cloud and then shared with healthcare providers and health insurance companies. The blockchain network is deployed for data integrity protection, in addition to which, it also stores access control policies and all access activities of the personal health data. However, the mobile application collects personal health data from either the sensors of the wearable devices or manual input by users. Therefore, the approach has its limitation of implementation in EHR systems where data are primarily accessed by health professionals and recorded in standardized formats. Very recently, Kokoris et al14 presented a scalable distributed ledger, the OmniLedger, where they used a technique known as sharding, to create subsets of nodes for parallel state and transaction processing. By implementing this approach, the processing capacity increases as the network scales out and overcomes the capacity issue of the traditional ledger, where the processing capacity decreases as larger consensus groups likely generate more overheads.

2 A BLOCKCHAIN-BASED ARCHITECTURE

2.1 Overview

To alleviate health providers’ workload of adopting the blockchain solution, we propose an architecture building on top of the existing providers’ databases. All accesses to the health records in these SQL server databases have to go through the blockchain, which then keeps track of all logs of queries, such as select, insert, delete, etc. Ownerships and access rights of records are important metadata added to the chain in addition to the logs. In this way, the history of all accesses is stored on the blockchain that provides a full view of all events that have happened to each record and hence guarantees data integrity and prevents misuse of user records.

Since users do not store records but only read, and each provider is responsible for data management and maintenance; we only let providers be involved in the maintenance of the blockchain. In this design, we also propose a new incentive mechanism, where a concept of significance of each provider is introduced and used as the main factor for selecting a provider who is responsible for creating a new block. Associated with this architecture, we design the smart contracts as Figure 2 shows.

- **Summary contract** (SC) holds a list of references of providers that a user has records with. A user-provider relationship is identified by a user ID and a provider ID. Associated with each user-provider pair is a reference number to the corresponding Record Relationship Contract (RRC) and a timestamp for the last edit of the record.

- **Record relationship contract** (RRC) stores metadata of a record. Such metadata includes ownership, access policies that indicate which entities are authorized to access the record, a status indicates if the updated RRC has been added to the blockchain or not, the current significance of the provider, and hashed logs of all read and write events that have happened to the record since the last update.

In the initialization phase, a significance value will be calculated for each provider according to both the quality and quantity of records that each provider maintains. The significance value will be continuously updated when the provider updates its database and is actively involved in maintaining the blockchain. All logs of events that have happened to the records, such as read/write of existing records or create new records, will
be collected, hashed, and added to the new block. After a pre-agreed update time, the new block will be appended to the blockchain according to a consensus mechanism.

2.2 | Implementation of the blockchain

Initialization
Suppose n healthcare providers (P) agree to join the blockchain network to share their records. In general, they are assumed to agree on:

- The summary contract and the record relationship contract;
- The calculation of significance and the incentive mechanism when a new block is added to the chain;
- Levels of frequency for updating the blockchain (e.g., daily, emergency case);
- The processes of generating, verifying, and appending a new block.

At the initialization stage, each provider \( P_i \), \( i = 1, 2, \ldots, n \), will be associated with a significance \( s_i \) based on the quantity and value of the health records in its own database. We have the quantity and the value of the records defined as: suppose \( P_i \) has \( m_i \) pieces of records in its database, the significance \( s_i \) of provider \( P_i \) is defined as:

\[
s_i = \sum_{t=1}^{m_i} v_t,
\]

where \( v_t \) refers to the value of each record \( R_t \) for a user \( U_t \). The interpretation of the value of a record may vary for different stakeholders. Here, we propose a definition of the value of a record based on two principles: completeness and redundancy. For measuring completeness, all providers agree in advance on a list of \( \Omega \) important items in a record, and for each important item present in a record, it contributes to the total value of that record. For the redundancy, we must realize that the contents of some items could be repeated by multiple providers for a common user’s records. In that case, the contribution from such items will be less valued than those that are unique. By taking completeness and redundancy into consideration, we define the value of a record as

\[
v_t = \frac{1}{\Omega} \sum_{i=1}^{\Omega} X_{t,i} I_{t,i},
\]

(1)

where \( X_{t,i} = 1 \) if the \( \omega \)th item in the record has content and \( X_{t,i} = 0 \) otherwise. \( I_{t,i} \) is the redundancy indicator of the \( w \)th item in a record and is defined as \( 1 \) divided by the number of providers having records with the \( \omega \)th item filled. Thus, the significance of a provider \( P_i \) is given by

\[
s_i = \sum_{t=1}^{m_i} \left( \frac{1}{\Omega} \sum_{i=1}^{\Omega} X_{t,i} I_{t,i} \right),
\]

(2)

If \( X_{t,i} = 0 \) for all \( \Omega \) items, it essentially means that the provider \( P_i \) does not have any value from the \( t \)th record and therefore gets no value of contribution from this record added to its significance. Note that this definition ensures that a record always has value less than or equal to 1.

Here, we give an example to illustrate the initialization of the significance. Suppose there are three providers: \( P_1 \), \( P_2 \), \( P_3 \) that agree on \( \Omega = 32 \) important items of a user’s health record. Assume that provider \( P_1 \) has 20 records and that each record has eight items out of 32 filled with contents that are not filled by any other providers; then, the value of each record is assigned \( \frac{8}{32} = \frac{1}{4} \); and \( P_1 \) has significance \( 20 \times \frac{1}{4} = 5 \); assume that provider \( P_2 \) has 30 records with 16 items filled with content, of which eight records have duplicate information for four items with provider \( P_3 \), and all the other items and records are only maintained by \( P_2 \); then, provider \( P_2 \) will get the significance as

\[
s_2 = 8 \times \frac{1}{32} \left( 4 \times \frac{1}{2} + 12 \right) + 22 \times \frac{16}{32} = 14.5,
\]

where each of the eight records has value \( \frac{14}{32} \) and each of the remaining 22 records has value \( \frac{16}{32} \).

After the initialization, each provider \( P_i \), \( i = 1, 2, \ldots, n \), has an associated significance \( s_i \) as defined in (2). The significance \( s_i \) will be dynamically updated, more specifically, the providers that fill more items to an existing record, create a new record, or create a new block will have their significance increased as follows:

- Creation of new contents to an existing record: the value \( v_t \) of this updated record will be increased by the items with new contents as in the initialization phase;
- Creation of a new record: the significance of the provider that creates the new record is increased by the value of the record as defined in (1);
- Creation of a new block: the provider who was selected to create a new block will be granted a certain amount of significance as an incentive if the block is successfully verified by the other providers.

It is worth noting that \( s_i \) is recalculated in the RRC once new contents or new records have been created, along with a new hashed log adding to the RRC.
Generating a new block

At a pre-agreed update time, all the providers that have updated records and that want to be involved the process of generating the new block will broadcast a tuple (Provider’s ID, role) in the network, where “role” is a two-bit string indicating whether the provider has any updated records and/or wants to create the new block. Each provider in the network will first collect all this information and later broadcast to the network what it has collected. The union of the broadcast collections will be the final status accepted by all providers.

Assume there are \( m \leq n \) providers that have registered in the collecting process for generating the block. One of these providers will be selected to carry out the task of block creation. The selection is based on the resources, i.e., the significance, which each provider owns. We propose a selection method as the incentive mechanism that makes providers with higher significance, indicating that they maintain more health records or records with higher values, less likely be selected, while providers that have less significance are more likely to be selected to carry out the task. For this purpose, we define

\[
p_i = \frac{1}{m-1} - \frac{s_i}{(m-1)\sum_{i=1}^{m} s_i}, \quad 1 \leq i \leq m.
\]

It is easily seen that

\[
\sum_{i=1}^{m} p_i = \frac{m}{m-1} - \frac{\sum_{i=1}^{m} s_i}{(m-1)\sum_{i=1}^{m} s_i} = 1.
\]

The selection process is: the providers \( P_1, P_2, \ldots, P_m \) are assigned to the ranges \( A_1, A_2, \ldots, A_m \), respectively, where \( A_j \) given by

\[
A_j = \begin{cases} 
(0, p_{j-1}), & \text{if } j = 1 \\
\left(\sum_{i=1}^{j-1} p_i, \sum_{i=1}^{j} p_i\right), & \text{if } 2 \leq j \leq m.
\end{cases}
\]

A pre-agreed pseudo-random number generator (PRNG), which takes participating providers’ ID and the pre-agreed update time for the new block as its seed, is used to generate a random number used to select the provider for creating the new block. The random number, in the range of \([0, 1)\), is derived from dividing the last generated \( m \) bits of the PRNG by \( 2^m - 1 \). If the random value falls into a range of \( A_j \), then the provider \( P_j \) is responsible for the task of creating the new block. The selected provider \( P_j \) is to create the new block, and its tasks include the following:

- Collect all RRCs from \( P_i \) who has new hashed log/logs;
- Create a block with all the new hashed logs;
- Broadcast to the network about the new block and call for verification.

Verifying the new block

Verifying the new block has two procedures:

- Each involved provider \( P_i \) verifies its logs in the new block, and sends its signed proof to \( P_j \);
- When \( P_j \) has received all the signed proofs, it updates its \( s_j \) with an incentive \( c \) and broadcasts all providers to append the new block.

Note the selected provider will be granted a certain significance as an incentive, which is denoted by \( c \). We will discuss this incentive value in the Section 3.2.

Appending the new block

After successful verification of the new block, all providers extend their own blockchain by appending the new block, in which the status and significance in each RRC are updated, the SCs will be updated with a new timestamp of last edits as well. Figure 3 shows the entire cycle of the implementation of the blockchain solution in the EHR system.

![FIGURE 3 The cycle of blockchain](image-url)
2.3 | Usage scenarios

A user reads his/her health record from a provider

A user wants to read his/her record from a provider’s database. He/she makes a query request to access the record. This query, together with a return value that is either the requested data or an access denied message, will be a log that is hashed and added to the RRC.

A health provider reads records from another provider

A provider B wants to read a user’s record from provider A’s database. B makes a query request to provider A. This query, together with a return value that is either the requested data or an access denied message, will be a log that is hashed and added to the RRC.

A health provider edits records

A provider wants to read or edit a user’s record from its own database. The provider makes a query to the record. All actions and edited data will be a log that is hashed and added to the RRC.

Figure 4 shows these three usage scenarios. Before a new block is mined, the status in each RRC indicates pending. If more than one hashed log has been added to an RRC within one blockchain update cycle, they will all be added to the RRC. After the new block has been appended to the blockchain, the status in each RRC changes to updated.

3 | DISCUSSION

3.1 | The incentive mechanism and significance

In many blockchain protocols, participating nodes are motivated by incentives to carry out a series of computations. For example, in the Bitcoin system, the incentives are in terms of bitcoins that can be obtained as transaction fees or rewards for mining blocks. In the MedRec, two incentivizing models were proposed: (i) using Ethereum’s inherent incentivizing model where the digital currency is Ether. Health providers need Ether to pay for their activities, and users need Ether to share information; (ii) bring the medical researchers and healthcare authorities to mine new blocks in order to get benefit from the network. However, these incentive mechanisms are not suitable for our application since the current healthcare systems are part of a national social welfare system that does not intend to involve monetary value in any shape or form at this level. Hence, data access requests from researchers or authorities are ruled by different protocols. Moreover, the blockchain we propose in this paper is indeed a permissioned blockchain that only authorized providers can participate in. For such permissioned blockchains, incentive mechanisms based on system performance and operation factors can serve better than classic monetary incentive mechanisms used in publicly open blockchains.
The concept of *significance* proposed in this work depends simply on the efforts that providers make in maintaining data and computing new blocks for the electronic record system. Instead of creating tradable digital currencies, we are aiming to achieve fairness among all providers for the sustainability of the system. However, other mechanisms can be considered for further deployments.

### 3.2 The incentive value $c$

In the proposed model, providers that have higher significance are less likely to be assigned to the computation task. Moreover, as a reward for completing the task, the selected provider will be awarded a certain amount of significance. The exact value of $c$ or the ratio of $c$ over the total amount of significance is open for discussion. We suggest two possible approaches for determining the value of $c$: (i) based on the size of the network and the distribution of the initial significance, define $c$ as a fraction of the average significance; (ii) obtain $c$ by deducting significance from not-selected providers as applying a zero-sum game theory where the total significance is fixed after the initialization. After every appended new block, the provider that executed the task takes a portion of significance from other providers, as taking more from the providers who have higher significance than those providers who have lower.

In this incentive mechanism, the distribution of significance changes by each new block adding to the blockchain (an update). To determine the trend of this distribution changing, we utilized MATLAB to run computing simulations. Each simulation recursively calculates 1000 providers’ significance by every update, and the initial significances were random numbers according to a $\mathcal{N}(10, 1)$ Gaussian distribution. Figure 5 and Figure 6 plot the variances of the distribution along with the times that new blocks were created and appended to the blockchain. Moreover, the two figures present results of the two approaches for determining the value of $c$, respectively.

The first approach we suggest is defining $c$ as a fraction of the average significance. Figure 5 shows the variances of the distributions when three different fractions are applied. With this approach, the total significance increase every time a new block is added to the blockchain, and a smaller fraction results in a lower increase of the overall significance. From Figure 5, we find that applying a small fraction (1%) of the average significance at each update generates the distributions with low variances. The variances of all three fractions reach over $10^6$ after 100 times of updates.

On the other hand, when the second approach is applied, in the process of every update, a selected provider takes a portion of significance from each of the not-selected providers as an incentive, while the total amount significance in the network is fixed. Figure 6 shows the results for four different portions of significance: 0.1%, 1%, 20%, and 30% are applied as incentives, respectively. For 5000 simulations, when fewer portions of significance (0.1% and 1%) are applied, increase the variances of significances distribution along with the number of updates and end on stable trends; when higher portions (20% and 30%) are applied, the variances start on a higher value and decrease along with the number of updates and end with stable trends as well.

The variances of distributions indicate the *difference* of significances among the providers; from the simulation results, we find using fractions of the current average significance as incentives will differentiate significances more than the approach where providers take significance from others as incentives. From the consideration of the fairness and the sustainability of the system, a *good* incentive approach would ideally reduce the difference of significances. In real implementations, the actual number of updates would likely be less than 300 (when update frequency is daily and global reset is yearly), and significance values would also be updated based on the addition of new records and contents.
3.3 | The selection of block creators

Designing an incentive mechanism for a closed blockchain network is a challenge. Our primary concern lies with the fairness among various scales of health providers when they are national-wise integrated as a blockchain network. Hence, the sustainability when the system deploys. We introduce the concept of significance that depends simply on the electronic records that the providers are maintaining, and the selection of block creator depends on the significance that the providers own. There are two advantages of this selection: (1) simplify the deployment of the blockchain since only the significance values are count; (2) the blockchain network is only about the records and not anything else; any profits or monetary benefits are exclusive. However, this simplicity may limit the diversity of this blockchain network, as it may not be feasible for the development of additional applications to the system.

3.4 | Implementation and scalability

The implementation of our design comprises a Hyperledger Fabric blockchain and a supporting Java application. It is a proof of concept for demonstrating the features realized by our proposed smart contracts and incentive mechanism. Essentially, the smart contracts enforce access control and auditing on EHRs stored on the blockchain, while the incentive mechanism is incorporated in the Fabric’s execute-order-validate pattern where it selects peers for the initial execute step and subsequently rewards the effort.

The implementation proves that the proposed framework can be realistically implemented in a real-world blockchain. The parameters of this blockchain system, as concisely presented in the design, are viable to reparameterization with increasing size of participants. A new user that is registered in the system will get a new summary contract and corresponding record relationship contracts, which will be stored on the blockchain; thus, the user is free to leave or rejoin any providers anytime. Providers will increase their significance by creating new records and adding new contents. The system also reserves some spots in the important item list in case new items have to be included in the list. The significance values and the list of important items will be globally reset on a regular time basis, eg, every second year.

The systems for managing EHRs have relatively few transactions compared to, eg, bank systems. Therefore, the frequency of generating new blocks can be configured to be hourly or daily. The system defines two update levels: either regular for the customarily scheduled updates or emergent if any immediate nonscheduled updates are required.

3.5 | Access control

In E-healthcare systems, access control is of utmost importance due to the sensitivity of the data. In our proposed architecture, the access to a record is ruled by an access control list that is maintained by the involved user, provider, and collective authority. By default, we only assign writing rights to the provider who initially issued the record and reading rights to other authorized providers and the user that owns the data.

To enhance the security of user data stored at providers, in this design, we adopt a combination of solid cryptographic mechanisms, including publicly verifiable secret sharing, distributed key generation, and a threshold variation of ElGamal cryptosystem to implement a collective authority that collectively manages the access control, acting a role similar to the miners in the Bitcoin protocol.
Before we introduce our approach to the access control, we need to recall certain cryptographic mechanisms that are used in the access control. Shamir\textsuperscript{18} introduces the \((t, n)\) secret sharing scheme, which refers to a method to distribute a secret \(s\) among a group of \(n\) participants, each of which allocates a share of the secret that is of no use on its own. The secret \(s\) can only be reconstructed when at least \(t\) out of \(n\) shares are combined. The basic scheme by Shamir solves the problem for the case where all players in the share distribution and secret reconstruction protocols are honest. Such a strong assumption is not needed in publicly verifiable secret sharing (PVSS) schemes,\textsuperscript{16} where any party can publicly verify the validity of the shares distributed by the dealer. The PVSS scheme allows for creating a collective private-public key pair without a trusted dealer. Each participant runs their own PVSS scheme in parallel to contribute a secret share to the key pair. After the participants reach a consensus on valid shares, a collective key pair \((sk, pk = g^s)\) will be generated, where \(g\) is a primitive root in a certain cyclic group \(G\). The public key \(pk = g^s\) will be made known to all \(n\) participants, of which \(t\) participants can collaboratively reconstruct the private key \(sk\) with their shared secrets.\textsuperscript{16}

In our proposed framework, we introduce a collective authority (CA) formed by all participating providers. We assume certain trust among providers since they are collaborating to realize a better E-healthcare system, which is their joint goal and their major reason to participate. Nevertheless, we still need to consider malicious attacks and system corruptions. The PVSS protocol\textsuperscript{16} is employed to generate a collective private-public key pair \((sk, pk = g^s)\). We adopt the Byzantine fault tolerance\textsuperscript{17} strategy in the reconstruction of the private key \(sk\), namely, we assume that at least \(t = \left\lfloor \frac{n - 1}{2} \right\rfloor + 1\) out of \(n\) shares \(sk_i\)'s are required to recover \(sk\). The access control in the design can be described on a high level as follows.

- **Distributed key generation**: the distributed key generation protocol in the work of Gennaro et al\textsuperscript{16} is used to generate a pair of collective private and public key \(pk_{secret} = g^{sk_{secret}}\), where \(sk_{secret}\) is the unknown private key. Each provider \(P_i\) holds a share of the private key denoted as \(sk_i\), and all providers know the public counterpart \(pk = g^s\). The collective private key \(k\) can be reconstructed by combining a threshold \(t = \left\lfloor \frac{n - 1}{2} \right\rfloor + 1\) of the individual shares. To be more concretely, a secret sharing polynomial \(s(x) = \sum_{i=0}^{t-1} a_i x^i\) of degree \(t - 1\) is chosen and \(sk_{secret}\) is set as \(sk_{secret} = g^{sk}\); each provider \(P_i, j = 1, 2, \ldots, n\), compute the encrypted share \(s_j = pk_j^{sk_{secret}^{-1}}\) of the shared secret \(sk_{secret}\) and create the polynomial commitments \(b_i = h_{sk^i}^{s_j}\) for \(0 \leq i < t\); by the Lagrange interpolation, the private key \(sk_{secret} = g^{sk}\) can be reconstructed from \(t\) decrypted shares \(s_j = \frac{1}{s_j} b_j = g^{sk}\).

- **Record updating**: the healthcare record \(R_{U_i}\) for a user \(U_i\) is updated after his/her visits to a healthcare provider \(P_j\) and is stored at provider \(P_j\)'s database in encrypted form \(c = AES256(m, K)\) by the symmetric cipher 256-bit AES with a symmetric key \(K\). The key \(K\) is generated from a secure hash algorithm as

\[
K = SHA256\left(U_i's\ credential, P_j's\ credential, \text{ Timestamp}\right).
\]

Furthermore, the symmetric key \(K\) is encrypted into a secret \(s = EIGamal_{Enc}(K, pk_{secret})\) by the preceding collective public key \(pk\). Provider \(P_j\) broadcasts the writing action on the record \(R_{U_i}\), the secret \(s\), and the updated access control list (ACL) of providers who are authorized to read the updated record in its database, if necessary, in the network. All of these information will be registered on the blockchain accordingly.

- **Record accessing**: assume an authorized provider \(P_i\) wants to access the latest updated record in Provider \(P_j\)'s database. Provider \(P_i\) issues a request to the network. Each member will first verify \(P_i\)'s access right and add \(P_i\)'s request to the log field of the record \(R_{U_i}\). Suppose \(P_i\) is authorized to access the record, \(P_i\) will be granted shares \(s_j\)'s from other providers. As long as \(P_i\) gather more than \(\left\lfloor \frac{n-1}{2} \right\rfloor\) secret shares, the private key \(sk_{secret}\) will be recovered, by which the symmetric key \(K\) will be obtained \(K = EIGamal_{Dec}(s, sk_{secret})\). Consequently, the key \(K\) is used to recover the plain form of the updated record in \(P_j\)'s database.

From the above processes of record updating and accessing, we can see that this access control design has the following features: (i) the user's record is in encrypted form, which prevents against malicious attacks or misuse of the record itself; (ii) the symmetric key is generated from the credentials of both the provider and the user, which indicates that the user has control of his/her data at the provider’s database. The processes also reflect the user-centric demand in modern E-healthcare systems (we need to mention that the user normally needs also to sign an agreement allowing for privileged access to the symmetric key for emergency cases); (iii) the timestamps in an hour and minute precision is also an input used for generating the symmetric key, which assures that the symmetric key \(K\) is time-sensitive and can only be used to decrypt the corresponding updated record; (iv) the underlying PVSS scheme\textsuperscript{16} assures that each provider cannot get any information of the private key \(sk\) or symmetric key \(K\); (v) the CA registers all access request on the blockchain as described in Section 2.2.0, which provides sound traceability of all data access events. Note that the access control list was initially made based on agreements of involved users and providers. The participating providers jointly run the above design of access control, so as to enforce publicly verifiable tracking of access to users’ records.

### 4 | CONCLUSION

In this paper, we proposed an architecture that implements blockchain technology in national EHR systems for the integrity of healthcare records and improves the interoperability of the current systems. It is compatible with existing e-healthcare systems and feasible for health providers to apply without compromising on the efficiency of records management. Considering that the system is essentially a multiple access system and health providers individually maintain records, we let providers take primary responsibilities of the maintenance of the blockchain, including the creation, verification, and appending of new blocks. Smart contracts are utilized in this design and can be further developed to other preferences,
eg, adding particular items in records to the blockchain for tracking. This heuristic architecture is independent of any specific blockchain platforms, and its variations can potentially fit other multiple access electronic records systems.

ACKNOWLEDGMENTS

This work was supported by the research project number 720025 from UH-nett Vest in Norway and partially presented at The 10th IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2018) held in Nicosia, Cyprus on December 10-13, 2018.

ORCID

Guang Yang https://orcid.org/0000-0002-3787-9616

REFERENCES

1. Nakamoto S. *Bitcoin: A Peer-To-Peer Electronic Cash System*. Campinas, Brazil: Portal Unicamp; 2008.
2. Tschorsch F, Scheuermann B. Bitcoin and beyond: A technical survey on decentralized digital currencies. *IEEE Commun Surv Tutor*. 2016;18(3):2084-2123.
3. Zheng Z, Xie S, Dai H, Chen X, Wang H. An overview of blockchain technology: architecture, consensus, and future trends. Paper presented at: IEEE International Congress on Big Data; 2017; Boston, MA.
4. Nguyen GT, Kim K. A survey about consensus algorithms used in Blockchain. *J Inf Process Syst*. 2018;14(1):101-128.
5. Szabo N, Kim K. Smart contracts: building blocks for digital markets. *EXTROPY: J Transhumanist Thought*. 1996;18(16).
6. Szabo N, Kim K. Formalizing and securing relationships on public networks. *First Monday*. 1997;2(9).
7. Ethereum. Ethereum Homestead Documentation. 2019. http://www.ethdocs.org/en/latest/
8. Cachin C. Architecture of the hyperledger blockchain fabric. Paper presented at: Workshop on Distributed Cryptocurrencies and Consensus Ledgers; 2016; Chicago, IL.
9. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: using blockchain for medical data access and permission management. Paper presented at: IEEE International Conference on Open and Big Data (OBD); 2016; Washington, DC.
10. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. *J Med Syst*. 2016;40(10):218.
11. Peterson K, Deeduvanu R, Kanjamala P, Boles K. A blockchain-based approach to health information exchange networks. Paper presented at: NIST Workshop Blockchain Healthcare; 2016; Gaithersburg, MD.
12. Kuo TT, Kim HE, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. *J Am Med Inform Assoc*. 2017;24(6):1211-1220.
13. Liang X, Zhao J, Shetty S, Liu J, Li D. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. Paper presented at: IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC); 2017; Montreal, Canada.
14. Kokoris-Kogias E, Jovanovic P, Gasser L, Gailly N, Syta E, Ford B. OmniLedger: a secure, scale-out, decentralized ledger via sharding. Paper presented at: IEEE Symposium on Security and Privacy (SP); 2018; San Francisco, CA.
15. Schoenmakers B. A simple publicly verifiable secret sharing scheme and its application to electronic voting. Paper presented at: Conference on the Theory and Application of Cryptology; 1999; Santa Barbara, CA.
16. Gennaro R, Jarecki S, Krawczyk H, Rabin T. Secure distributed key generation for discrete-log based cryptosystems. *J Cryptol*. 2007;20(1):51-83.
17. Desmedt Y, Franklin M. Threshold cryptosystems. Paper presented at: Conference on the Theory and Application of Cryptology; 1990; Santa Barbara, CA.
18. Shamir A. How to share a secret. *Commun ACM*. 1979;22(11):612-613.
19. Castro M, Liskov B. Practical byzantine fault tolerance. Paper presented at: 3rd Symposium on Operating Systems Design and Implementation; 1999; New Orleans, LA.

How to cite this article: Yang G, Li C, Marstein KE. A blockchain-based architecture for securing electronic health record systems. *Concurrency Computat Pract Exper*. 2021;33:e5479. https://doi.org/10.1002/cpe.5479