Special Collection: Smart Wellness Services for Lifestyles of Health and Sustainability (LOHAS)

INTERNATIONAL JOURNAL OF
ENGINEERING BUSINESS MANAGEMENT

International Journal of Engineering Business Management
Volume 10: 1–12
© The Author(s) 2018
DOI: 10.1177/1847979018790589
journals.sagepub.com/home/enb
SAGE

Blockchain technology for providing an architecture model of decentralized personal health information

Sandi Rahmadika and Kyung-Hyune Rhee

Abstract

The personal health information (PHI) is an activity among the health-care providers and the patients in terms of managing the data which is sensitive to the parties. The PHI data have been maintained by multiple health-care providers, thus resulting in separated data. Moreover, the PHI data are stored in the provider's database, hence the patients have no authority to manage their own information. Therefore, in this article, we propose a conceptual model for managing the PHI data which is derived from several health-care providers by relying on the blockchain technology in the peer-to-peer overlay network. In addition, we elaborate the security analysis that might be occurring in the proposed model. By leveraging on our model, it allows the patients and the providers to collect effectively the PHI data onto a single view as well guarantee of data integrity. The blockchain offers an immutable of the data record without having to trust a third party. The experimental results show that the proposed approach is promising to be developed due to the high success rate in terms of data dissemination.

Keywords

Blockchain, chord-based distribution, overlaying network, personal health information

Date received: 1 December 2017; accepted: 20 February 2018

Introduction

The personal health information (PHI) literally is the activity that relates patients to healthcare providers in organizing, managing, and accessing their PHI data. The essence of the decentralized system is based on a powerful idea to organize, properly structure, and steer the system in order to improve the service of a health care system in general. However, the decentralized system still has some crucial issues in terms of managing the data. Therefore, the issues become essential for the developers to arrange the strategy in advance.[1]

In recent years, the PHI has been used by health organizations in order to provide the updated information of the patient by simply accessing the cloud storage or the provider's database.[2] The medical data literally relate to the information about the patient such as gender, date of birth, the care plan, and many others.[3]

The majority model of PHI data has been compiled and maintained by several health-care providers,[4] thus resulting in separated and disseminated patient data.[5] Mostly, medical data are stored in the provider's database. As a result, the patients have no full access to manage it,[6] and the health-care providers also do not share the data directly to the patient. The providers might send PHI data in different formats because there is no agreed standard format yet. As a result, it is difficult for the patients and the providers to access the data. In this regard, managing the PHI data is

Pukyong National University, Busan, Republic of Korea

Corresponding author:
Kyung-Hyune Rhee, Pukyong National University, A12-1305, Daeyeon Campus, Yongso-ro 45, Nam-gu, 48513 Busan, Republic of Korea.
Email: khrhee@pknu.ac.kr

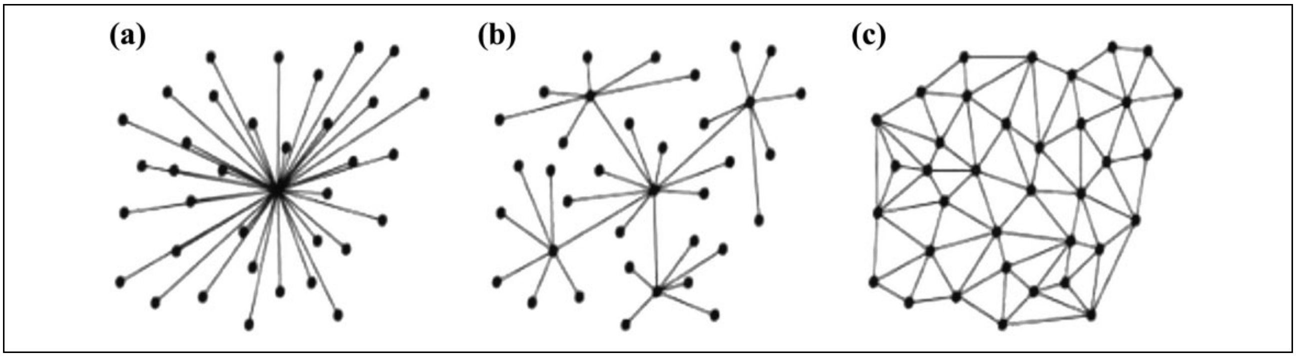


Figure 1. (a) Centralized, (b) decentralized, and (c) distributed networks.

paramount for the industry to develop a well-decentralized system in the health care area.

In this article, we remedy this problem by proposing an architecture model for managing the PHI data using blockchain technology. Data are propagated in the peer-to-peer (P2P) overlay network, while valid data are stored in the blockchain storage. In order to verify the validity of the transaction, miners should solve the proof-of-work (PoW) puzzle before a new block is added to the main chain. The structure of blockchain allows the parties to track the history of PHI data, because every transaction record is stored in the chaining block.

The PHI data come from various health-care providers such as doctors, chiropractors, clinical psychologists, national providers, and laboratories, to name a few. In the proposed model, data are divided into several data blocks making it easier for the parties to collect them.

The structure of the article is organized as follows. The second section presents the overview of blockchain technology, while the third section elaborates prior works. The core system components such as P2P network, PoW, chord protocol, and overlay network are given in the fourth section. Security issues are described in detail in the fifth section. The system model is given in the sixth section. The simulation results, analysis, limitation, challenges, and opportunities are given in seventh section. Finally, the last section presents a conclusion and remarks some future works related to the blockchain technology in the health care area.

Blockchain technology

Blockchain technology is a data structure used to create a decentralized ledger[1] composed in a serialized manner. A block of the blockchain contains a set of transactions, a hash value of the previous block, time stamp, block reward, and block number, to name a few. Every node in the blockchain network holds the same replica of the data. Blockchain can hold the information and set rules on how the information is updated. The major advantages of blockchain are related to the automation system and creation of transparent yet secured application. It allows preventing fraud, corruption, and enables to solve many other problems. Blockchain technology can support a new generation of transactional applications and streamlined business processes by establishing trust among parties, accountability, and transparency that are essential to modern commerce.

The modern commerce and the industry rely on trust with the cryptography protocol module embedded in the system to ensure the credibility of the transaction, mostly, the traditional network architecture; for example, in a centralized network (see Figure 1), the nodes send data first to the central node instead of directly sending the data to the recipient. Literally, the transaction relies on the middleman. As opposed to the centralized system, the decentralized model is worth mentioning as a distributed network of centralized networks system without the central node. A decentralized system is referred to a propagated data onto the entire network without an intermediary party. In this regard, the system does not rely upon a single server and rid of the single point of failure that might have been caused by the middleman. In various sectors, blockchain is used for several purposes such as financial registries, operational registries, and contracts.

(i) *Financial registries*: The cryptocurrencies such as Litecoin, Bitcoin, and Dogecoin can be used as an alternative for the real currencies in the blockchain system. The cost of a cryptocurrency transaction is potentially lower than the cost of a classic transactionand enables micro-payments.

(ii) *Operational registries*: Blockchain allows tracking and certification of specific products or assets, including renting contracts, land registers, and notary deals or votes.

(iii) *Smart contracts* (automated actions on the blockchain): The account holding objects contain several code functions to make decisions, store data, and send the cryptocurrencies to other's network. The smart contract has an ability to execute the code (*self-executing*), and it is more superior to the traditional contract because it provides security with the cryptography protocol embedded. Blockchain can drive and run on autonomous conditions and terms of a predefined contract.

One of the prominent decentralized cryptocurrencies is Bitcoin. Every address in Bitcoin has a unique value from a pair of the public and private key as an identity.[4] A bitcoin address is an identifier of 26 to 35 alphanumeric characters, and the address is computed from an Elliptic Curve Digital Signature Algorithm (ECDSA) public key. The address owner knows the corresponding private key by using a transformation based on a hash function. Since the hash is a one-way function, it is possible to compute an address from a public key, but it is infeasible to retrieve the public key solely from the address.

Related works

The hierarchical distributed electronic health record (EHR) model[4] aims to maintain the patient's data in the health organization and at the same time replicate to other hospitals, ensuring the credibility of data. In this research, the P2P distribution technique is a future proposal. The other discussion such as security, privacy, and data integrity is beyond the research. An architectural model which is called OmniPHR[5] was proposed in order to address some issues in the centralized health care area. In this model, there is no guarantee of security in terms of privacy and data integrity. OmniPHR architecture model faces some challenges regarding PHR contexts, that is, how the health-care providers enable to access the up-to-date data regarding dynamics of the patient's condition and the issue of data replication.

A conceptual model of privacy is applied for the health information system using a wearable device with a distributed mechanism based on cloud server. The CIA and HIPAA protocols are adapted to provide security and privacy to the user.[6] Even though the model privacy framework is applied on a wearable device, there is no information related to the operating system, that is, how the system generates the programs and the way they communicate with each other. Similarly, the concept and the application of a ubiquitous PHR system[7] was proposed to make it easier to access the medical information for a new patient. The model is based on several literature reviews. The information related to the security and privacy issue is beyond the scope of the article.

In Table 1, the models concentrate on health data of allpatients in single and multiple servers, following a centralized client–server (CS: client–server, DO: distributed object, DC: distributed component, DE: distributed event-based). Avancha et al.[8] refers to Microsoft Health Vault (MHV) and Google health as two well-known PHR services. However, the Google health was permanently discontinued because of the lack of widespread adoption. The MHV is a cloud-based platform that guarantees some cryptographic properties which are used to manage the data from providers. The future work is to develop an innovative concept for wellness management system between the health-care provider and the patient.

Table 1. Architecture model of related work.[5]

Related work	Model	Security
HDEHR	DE, P2P	—
m-Health	DE	—
uPHR	DE	—
CF	CS, DO	CIA, HIPAA
HealthVault	CS	Authentication
healthTicket	CS	CP-ABE
DEPR	DC	—
My HealtheVet	DE	Security policies
SNOW	DC	Privacy policies

HDEHR: hierarchical distributed electronic health record; uPHR: ubiquitous personal health record; CIA: Confidentiality, Integrity and Availability; DEHR: Doversity and Equity in Health Reform; CS: client–server; DO: distributed object; DC: distributed component; HIPAA: Health Insurance Portability and Accountability; CP-ABE: Ciphertext-policy Attribute-based Encryption; P2P: peer-to-peer.

Core system components

In this section, we present components of the system such as the general information about the P2P network as a distributed application architecture, PoW as a consensus mechanism, chord-based distributed system, and the overlay network.

P2P networks

P2P network is an architecture network that makes workload partitions between the nodes in the network. A node is also called as peers that refer to the way they connect to each other and doing some multiple tasks together in the network.[9] The P2P network enables the peer to provide a broadcasting information to the entire node in the same network. A peer provides incoming Transmission Control Protocol (TCP) connections (8333 for Bitcoin) to refer to other peers. The connections between the peers are generated automatically through IP addresses in the same network.[10] In the Bitcoin, when a peer receives a list of full Bitcoin node IP addresses, the peer maintains up to 8 outgoing connections with the other peers.

The P2P transaction, for example, Bitcoin transaction, is executed by the owner of the coins by transferring some of the cryptocurrencies to the next owner by signing (digital signing) a hash value from the previous transaction as shown in Figure 2. The hash value is generated by executing the cryptographic hash function such as SHA-256 and RIPEMD160.[11] A hash function is a cryptographic protocol, and it is easy to produce a hash for several inputs, but it is extremely difficult to get the original value of the input.[12] A unique message digest from the hash value is used to keep the credibility of the data.[13] The properties of hash values such as *preimage resistance* and 2*nd preimage resistance* guarantee the original value of the message. Whenever an attacker tries to tamper a data block in the blockchain network, every peer easily detects the attacker's activity. More precisely, the attacker needs to change the whole data block in the network which is impossible.

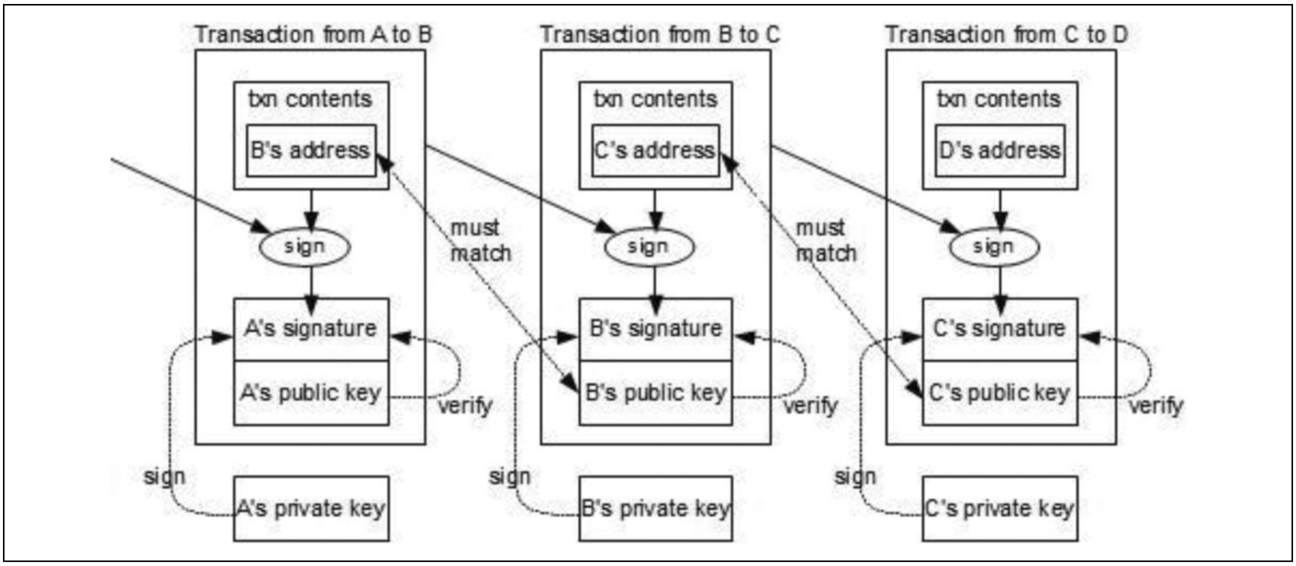


Figure 2. Peer-to-peer transaction.

Simple payment verification client, in particular, do not provide direct services to the P2P overlay network and only receive information relevant to the address. Define t_{f} as time to finish the service once the connection between the nodes is generated. For the request is the node, let's assume the request is random from several nodes. Then the service time s is defined as follows[14]

$$s = \begin{cases} t_{\mathrm{f}}, & \text{if the peer is connected} \\ t_{\mathrm{f}} + \delta, & \text{if the peer is not connected} \end{cases} \tag{1}$$

Whenever a demand arrives, the node can be connected with the probability p_{c}. We can define the value of average service time as follows

$$E[s] = (1 - p_{\mathrm{c}})t_{\mathrm{f}} + p_{\mathrm{c}}(t_{\mathrm{f}} + \delta) = t_{\mathrm{f}} + p_{\mathrm{c}}\delta \tag{2}$$

Let $\rho = \lambda/\mu$ be the operation of the node and the demand rate from several sources, and μ is defined as an average rate; the fraction can be defined as follows

$$u_s = \rho\left(\frac{1 - p_{\mathrm{c}}}{E[s]}\right) \tag{3}$$

$$u_{\mathrm{s}} = (1 - \rho)\rho_{\mathrm{c}} \tag{4}$$

$$u = 1 - (u_{\mathrm{s}} + u_{\mathrm{i}}) \tag{5}$$

PoW

A PoW is a method to reach the consensus among the miners in Bitcoin blockchain system. In the PoW, the miners have to solve the cryptographic puzzle by finding 256 bits of the target value. The miner is relying on the computer power to solve the puzzle. The blocks are created by a procedure called mining. The lowest a target value, the more time it takes to solve the puzzle and vice versa. A probability of finding *nNonce* of proof H for given target T is

$$P(H \le T) = \frac{T}{2^{256}} \tag{6}$$

Table 2. Blockheader format.

Field	Description	Size (bytes)
Version	Block vers. numb	4
hash-prev block	Hash of prev.header	32
Merkle root hash	Tx Merkle root hash	32
Time	Unix time stamp	4
bBits	Current difficulty	4
Nonce	Allows miners search	4

The drawback of PoW is related to efficiency, as it wastes too many computational resources to find the target value. The hash in PoW begins with the number of zero-bit hashes (SHA-256) and involves the scanning for the value when hashing a data. The block header format presented in Table 2 contains the information for one block such as the version of the block, hash of the previous block, Merkle root hash, current difficulty value, and the nonce (the number of attempts to find the target value). PoW is described as the race between the miners, and the fastest miner who solves the puzzle is the winner and gets some cryptocurrencies as reward. The more the computational power of the miner, the more chances to solve the puzzle.

Chord-based distributed system

Chord protocol in the overlay network allows the node to edit the data keys, insert the object, and so on.[15] Chord algorithm can be used as an intelligent option to decrease latency and decrease the message cost. Chord-based

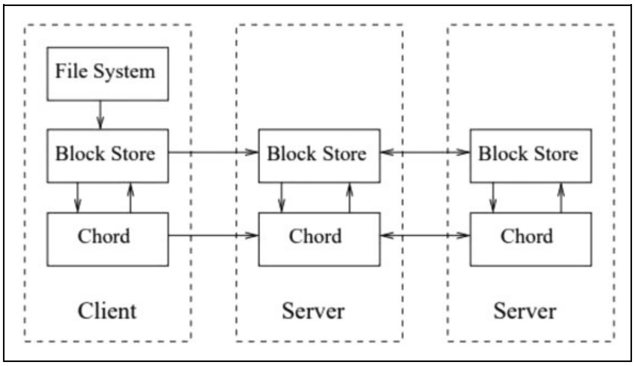


Figure 3. Chord algorithm (peer-to-peer).

distributed system consists of the hash value which is referred to the nodes as follows:

1. *SHA-1* (IPaddress, port) 160 bit
2. Cutt-off to *m* bits
3. Referred to *peerid* (*0* and $2^m - 1$)
4. Own specific character for every value
5. Map nodes *2^m*

Figure 3 shows the basic structure of chord algorithm from the client to the servers. On the client's side, there are some components such as file system, block store, and the chord. However, from the server, there are block store and the chord that connects to another server. The main usage of the chord protocol as a query value from a client is to find a successor (k) that referred to $O(N)$ query time, and then N is referred to the number of machines (ring).[14] The distributed hash table is a similar data structure, except it runs in a distributed system rather in a single process, and the objects are files that have the unique keys. The load balancing is another issue in the distributed hash table. More precisely, it is referred to each host to have about the same number of objects stored, because the system does not want some hosts to be overloaded with objects while others have fewer objects.

Figure 4 shows the basic information of a chord protocol, which consists of three identifier nodes 0, 1, and 3, followed by the set of the keys (1, 2, 6) as a key identifier (the keys are the input of the three nodes). The successor of key 1 is derived from the node 1, and the successor of key 2 is obtained from the value 2. In order to maintain the consistency of hashing value, whenever a node n joins the network, the successor of n's keys should be assigned to n. In case the node n leaves the network, all of the assigned keys are reassigned back to n's successor.[16]

Overlay network

An overlay network is a virtual network that is built on the top of another network in a computer system. The node in the overlay network can be described as virtual[17] or logical network that corresponds to a path (see Table 3). The central idea behind overlay networks is tunneling, for example, the packets in the network are captured and encapsulated to be forwarded to their actual destinations. The common form of an overlay network is used to distribute the key value stores to exchange the topology between agents such as *Zookeeper*, *Etcd*, and *Consul.* Such protocols also introduce the extended virtual network IDs, for example, a 24-bit VXLAN Network ID (VNI) and Virtual eXtensible LAN (VXLAN) support over 16 million virtual networks. The prominent examples are as follows:

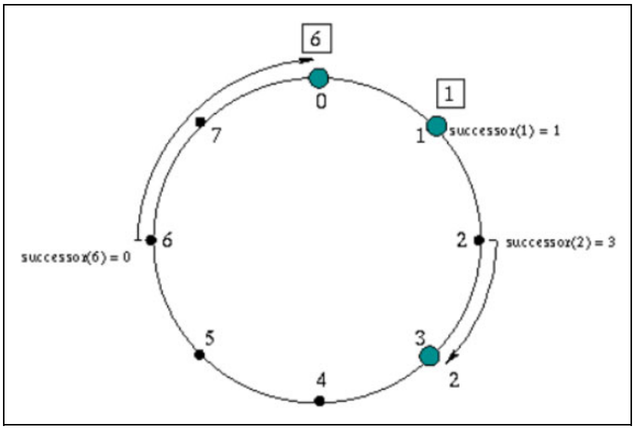


Figure 4. Chord network consisting of three nodes (0, 1, and 3).

Table 3. Physical path (Figure 6).

Link overlay network	Physical path of the network
A-B	A-B
A-C	A-a-b-C
A-E	A-a-c-d-E
B-C	B-a-b-C
C-D	C-D
C-E	C-e-E
D-E	D-C-e-E

1. VXLAN, encapsulates L2 into User Datagram Protocol (UDP), and the tunneling using L3 (no specialized hardware is required and could be built purely in software).
2. Network Virtualization using Generic Routing Encapsulation (NVGRE), encapsulates L2 into Generic Routing Encapsulation (GRE). The essential of NVGRE standard include identifying a 24-bit Tenant Network Identifier (TNI) using GRE to create an isolated virtual network.

The overlay network has the ability to self-organize. For example, when a node fails to form any specific reason, the overlay network protocol has some preventive action and gives solutions to solve the problem, such as recreating a proper network for the system.[18]

1. Retain existing connectivity model (internal addressing, network services, and security model).
2. Isolate virtual segments and network services.
3. Per-tenant QoS (Quality of Service) which is implemented to support differentiated service level agreements with tenants.

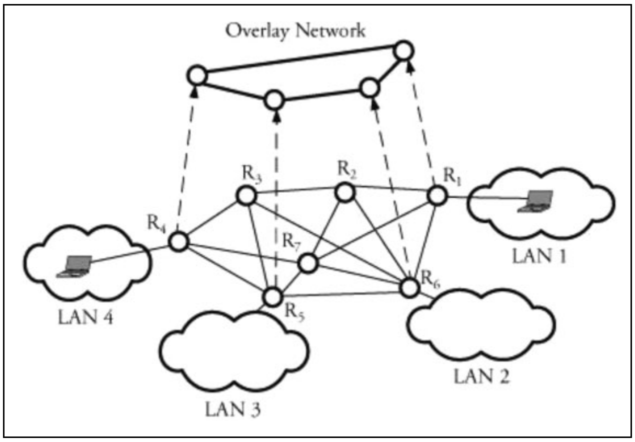


Figure 5. An overlay network for the connections. LAN: local area network.

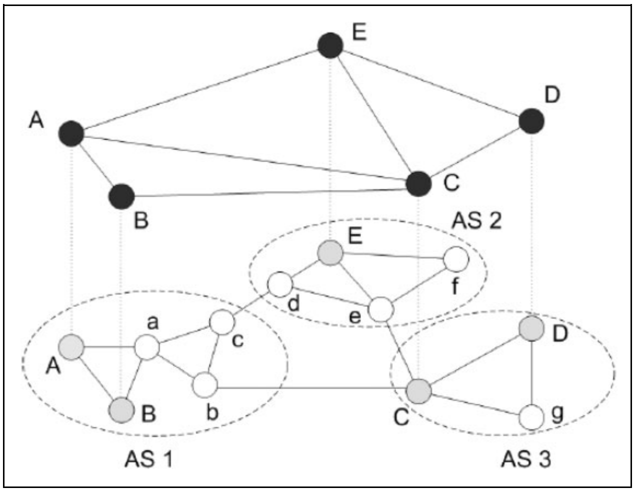


Figure 6. Overlay and physical network.

Figure 5 describes a virtual network for the connections between two local area networks. Content delivering network (CDN) is an example for the overlay network. CDN distributes the static content of websites, for example, pictures and videos, and puts them in a specific location. If a regular visitor visits a page without CDN, it goes directly to the Web host.

System model and architecture

In this section, we present the model overview of the system, the components of the network, and the network architecture to distribute the PHI data to the entire network.

Model overview

In the model overview, we focus on the way the parties distribute and collect the PHR data from several health-care providers. In the proposed model, the patient is enabled to receive the latest version of his/her PHR data from the provider in a single view in the blockchain storage. The PHR of the patient is distributed in the overlay network with a chord protocol embedded. The PHR is divided into several blocks.

The version of PHI data is dynamic (can be changed anytime), as health-care providers have the ability to input the serial data. The proposed system allows the patient to get a single view of their PHI data with the guarantee of data integrity.

The patient and the health-care provider is connected to each other in the blockchain network. The provider is categorized as hospitals, laboratories, chiropractors, clinical psychologists, national providers, and health and social care professional (see Figure 7). In this sense, the health-care providers and the patients allow accessing the latest version of PHI data. The health-care providers also act as miners that solve the PoW puzzle, but the winner will not get any cryptocurrency as reward. The miner who solved the first puzzle of PoW propagates the result to other nodes through the P2P network. The rest of the nodes then validate the result of PoW which is easy to confirm. Furthermore, a new block will be added to the main chain, after which every node in the network keeps the replica of the new block. In the Bitcoin protocol, the data block is valid when it reaches six confirmations from other nodes.

In order to support the communication between the nodes in the network, a publish–subscribe system is used as the messaging pattern in the overlay network. It has the ability to receive and update PHI data blocks and allows the node to maintain a system user register and maintain access permissions to PHI.

The network architecture

The network architecture is designed based on virtual P2P network,and it is built on the top of another network as shown in Figure 8. The peers in the blockchain network come from the various health-care providers. The PHI data can be updated anytime and anywhere by the provider, after which the data will be stored in the blockchain storage right after the block is confirmed by the nodes. The overlay network possesses some advantages such as scalability. The overlay network aims to decentralize the data and locate the nodes on the network. By leveraging this model, the positive goals could be reached such as providing distribution, data replication, security, and data integrity.

In the proposed scheme, the PHI data are divided into small pieces in the storage. Once the data are stored, it will be encrypted using the private key. Meanwhile, the patients use the public key to decrypt the data. We assume that a pair of keys is generated automatically by the system and it is known in advance. The key pairs are simply referred by consecutive natural numbers which are defined as an integer from a given range of numbers.

The components of the network consist of the router, local computer, laptop, switch, and wireless router. A router literally can be defined as a device that determines a packet to be forwarded to the destination point which

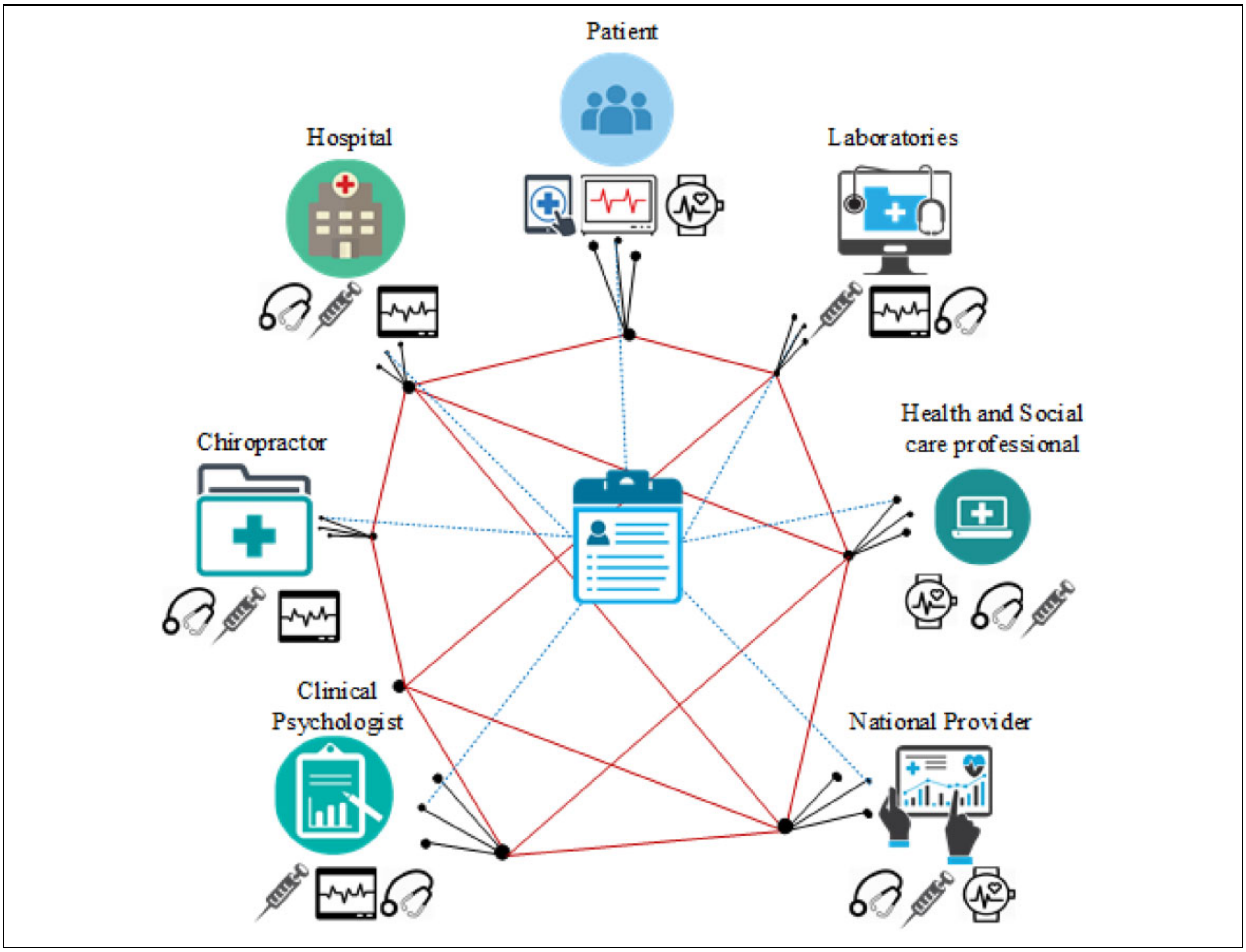


Figure 7. The system model.

connects to the networks. Whenever there are requests to access the main page from a different computer, the router will check first the IP address and use it as a hash key to direct to the destination. More precisely, the router transmits the packet data of PHI from source to another computer in the blockchain network. Every health-care provider at least installed one router to make a connection with the different nodes in the same blockchain network.

Healthcare providers and patients in the blockchain system are described as a local computer (16 computers) which is connected to the routers (see Table 4) and wireless routers. The wireless routers use Wi-Fi Protected Access (WPA-PSK) and Pre-Shared Key (PSK) passphrases to connect with other devices and every message is encrypted using Advanced Encryption Standard (AES). The principle of AES is based on a substitution–permutation network and a combination of both substitution and permutation technique. AES has several fixed lengths of block size, for example, 92 bits, 128 bits, and the maximum key size is 256 bits. The very early step of the cipher is to put the data into an array, after which the cipher transformations are repeated over a number of encryption rounds.

Results and analysis

In this section, we present the detailed information about the components. We elaborate the system model and analyze the simulation results. Yet we highlight the security analysis that might be occurring in our system. The limitation, challenges, and the opportunities are drawn in this section.

System model analysis

The components of the system model perform a key role, because it is essential for the input and output gateway of the data blocks. In charge of distributing the data blocks on the network, the component requires knowledge of data blocks location as well as ability to fetch data blocks in the appropriate node that contains the requested data. In general, the data blocks are stored on the computer where it was created, and some copies are distributed on the routing overlay network. In this sense, the original data reported by a health-care provider remain stored in the health organization with copies of data blocks distributed over the network.

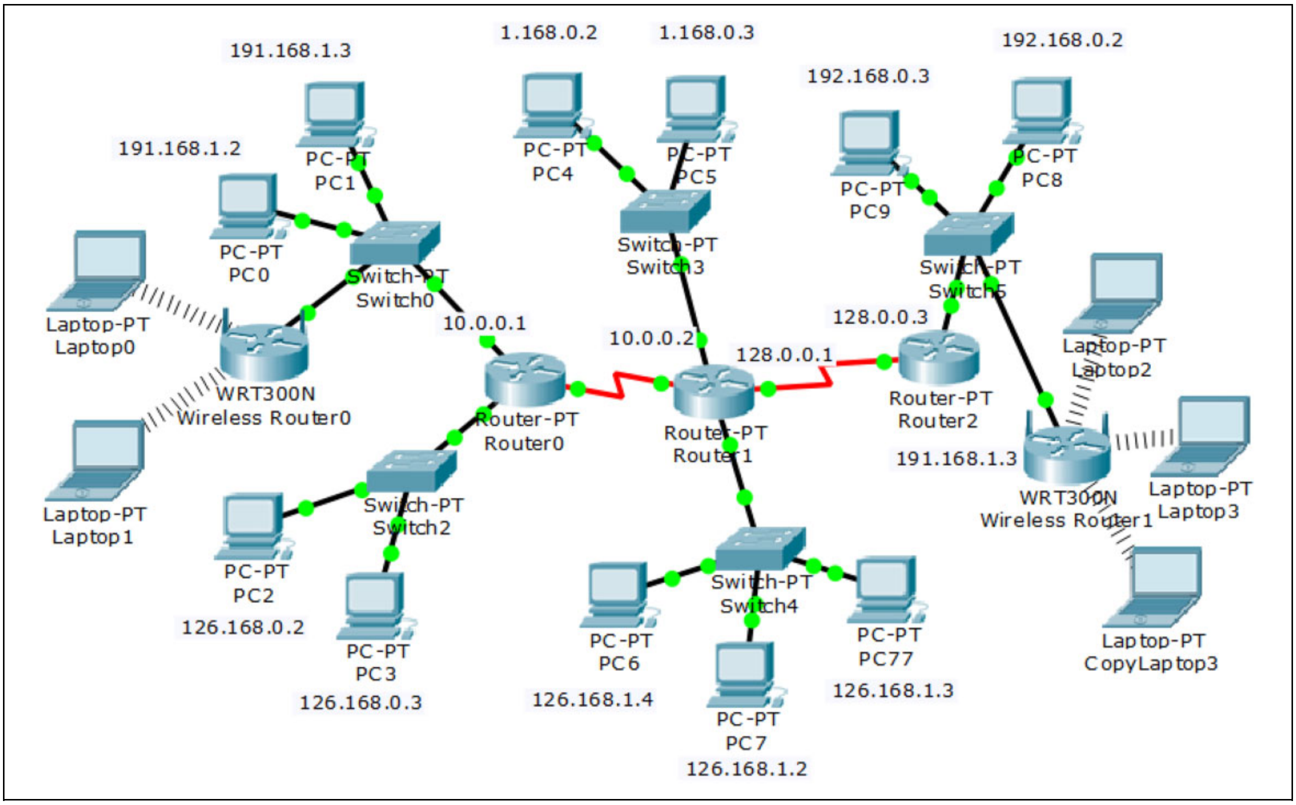


Figure 8. The network architecture of the system.

Table 4. Components setting.

Components	Router0	Router1	Router2
ID config	10.0.0.1	10.0.0.2	128.0.0.3
FEtrnt0/0	1.168.0.1	192.168.0.1	191.168.1.1
FEtrnt1/0	128.168.0.1	126.168.1.1	—
FEtrnt3/0	—	128.0.0.1	—
Tx ring	10	10	10
Clock rate	2000000	2000000	2000000

The PoW puzzle is the mathematical problem that the miner's computer needs to solve before adding a new block to the main chain of the blockchain network. Figure 9 describes the communication system among the health-care providers and the update of PHI data is propagated in the overlay network. The detailed information of communication data can be seen in Figure 10.

In this model, we use an open cache solution, which aims to achieve better performance and to make sure the node keeps following the chord protocol. The system has a validator that is assigned to validate every data in the blockchain storage. More precisely, it checks the integrity of new blocks, ensures the consistency of the network, and corrects the sequencing of data blocks. To control the I/O node in the network and promoting scalability and load-balancing capabilities, a node manager[5] is assigned to remedy the problem. Whenever a new node requests to join the network, a node manager generates a new identifier ID for the new node. A digital signer component responsible for keeps the integrity of the data blocks on the transmission. The system provides a digital signature which then is used to assign the data respectively.

The result of data communication is shown in Figure 11. It shows that the success rates reach 100% (there are no losses data) for the communication and data exchange among health-care providers. However, the average time is 1:18-millisecond for 100 bytes of Internet Control Message Protocol Echos. The data source and destination are recorded randomly at *n* period time and the color of the table represented message data. The results show that using the proposed approach for managing the PHI data is remarkably promising as it facilitates the access, control and manages the data of PHI.

Security analysis

In this section, we highlight several attacks that might occur in the model, especially the attacks in the P2P network. There are some conditions for a malicious agent attempts to attack the blockchain network and gain an unfair advantage for the personal purpose, either by tampering the PHI data or by creating fake transactions, to name a few.

Eclipse attack—security threats in the blockchain may come from P2P networking. A prominent example is the

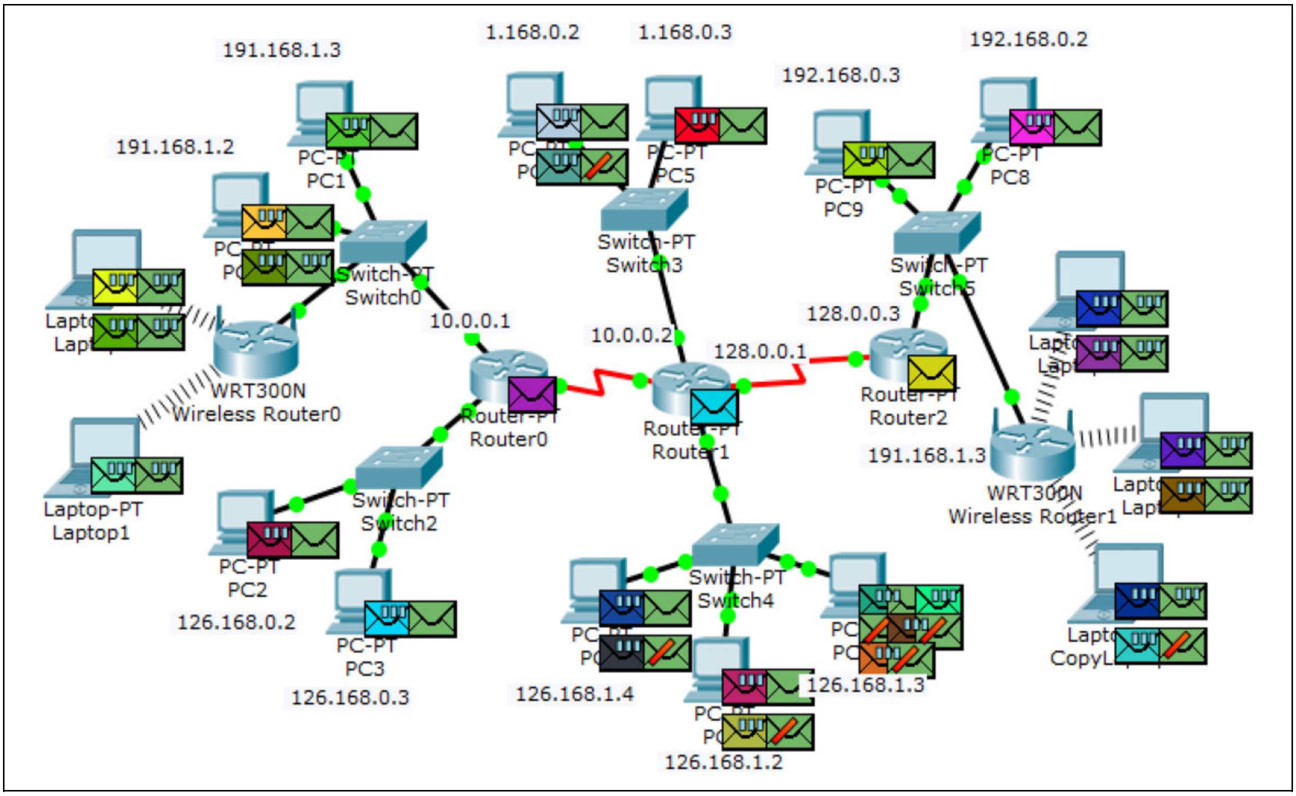
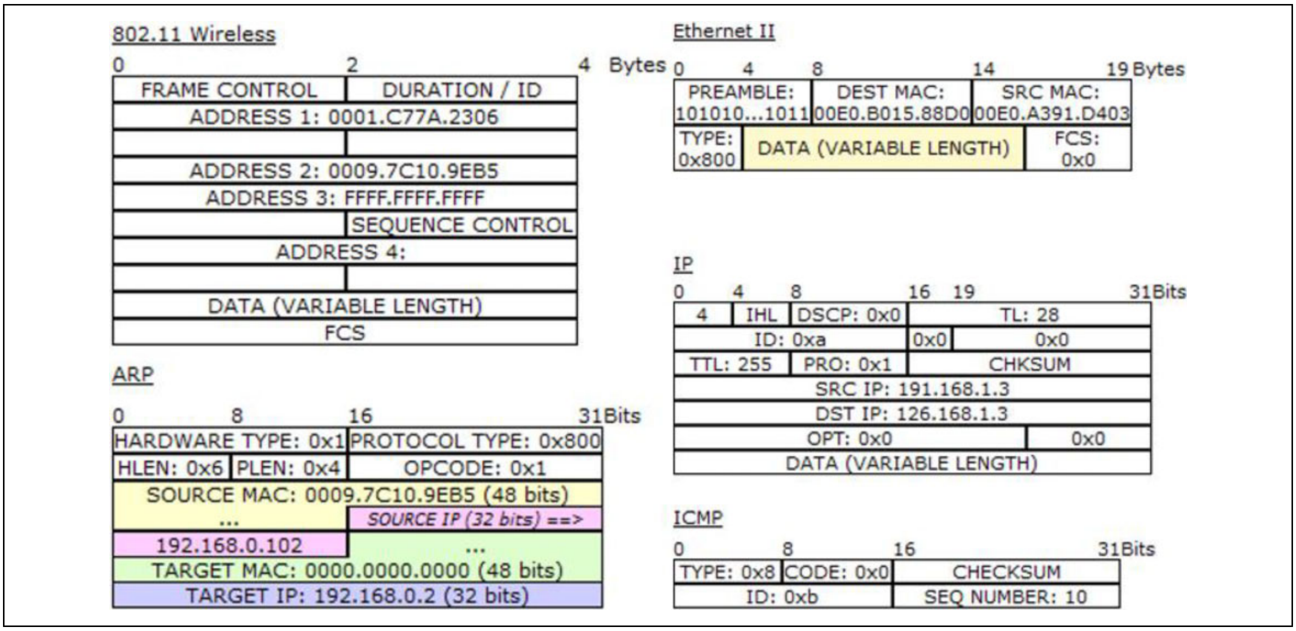


Figure 9. Propagating the PHI data. PHI: personal health information.



802.11 Wireless

0 2 4 Bytes

FRAME CONTROL	DURATION / ID
ADDRESS 1: 0001.C77A.2306	
ADDRESS 2: 0009.7C10.9EB5	
ADDRESS 3: FFFF.FFFF.FFFF	
	SEQUENCE CONTROL
ADDRESS 4:	
DATA (VARIABLE LENGTH)	
FCS	

ARP

0 8 16 31Bits

HARDWARE TYPE: 0x1		PROTOCOL TYPE: 0x800
HLEN: 0x6	PLEN: 0x4	OPCODE: 0x1
SOURCE MAC: 0009.7C10.9EB5 (48 bits)		
...		SOURCE IP (32 bits) ==>
192.168.0.102		...
TARGET MAC: 0000.0000.0000 (48 bits)		
TARGET IP: 192.168.0.2 (32 bits)		

Ethernet II

0 4 8 14 19 Bytes

PREAMBLE: 101010...1011	DEST MAC: 00E0.B015.88D0	SRC MAC: 00E0.A391.D403
TYPE: 0x800	DATA (VARIABLE LENGTH)	FCS: 0x0

IP

0 4 8 16 19 31Bits

4	IHL	DSCP: 0x0	TL: 28
ID: 0xa		0x0	0x0
TTL: 255		PRO: 0x1	CHKSUM
SRC IP: 191.168.1.3			
DST IP: 126.168.1.3			
OPT: 0x0			0x0
DATA (VARIABLE LENGTH)			

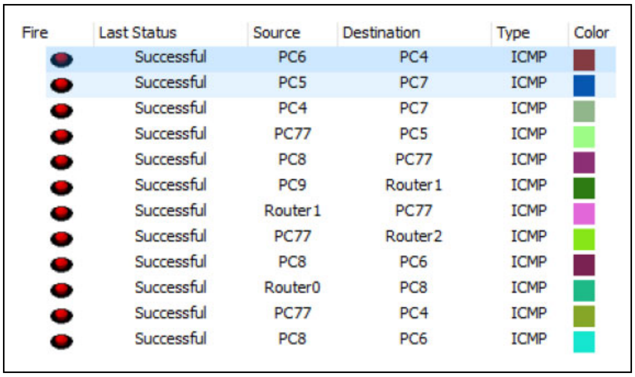
ICMP

0 8 16 31Bits

TYPE: 0x8	CODE: 0x0	CHECKSUM
ID: 0xb		SEQ NUMBER: 10

Figure 10. Information of communication data.

eclipse attack,[19] in which all neighbors of an agent are under the control of the malicious agent. The eclipse is an attack that gains control over a network access to information in P2P network, and with proper manipulation of P2P, the adversary can eclipse the target, as a result of which the victim is able to communicate with the malicious node. To do so, an attacker can manipulate the node so that all of its outgoing connections head to the attacker IPs. More precisely, the adversary fills the node's peer tables with attacker IPs. Then, the node restarts and loses its current outgoing connections. Finally, the node makes new connections only to the attacker IPs.



Fire	Last Status	Source	Destination	Type	Color
●	Successful	PC6	PC4	ICMP	🟥
●	Successful	PC5	PC7	ICMP	🟦
●	Successful	PC4	PC7	ICMP	🟩
●	Successful	PC77	PC5	ICMP	🟩
●	Successful	PC8	PC77	ICMP	🟪
●	Successful	PC9	Router1	ICMP	🟩
●	Successful	Router1	PC77	ICMP	🟪
●	Successful	PC77	Router2	ICMP	🟩
●	Successful	PC8	PC6	ICMP	🟪
●	Successful	Router0	PC8	ICMP	🟩
●	Successful	PC77	PC4	ICMP	🟩
●	Successful	PC8	PC6	ICMP	🟦

Figure 11. The status of communication data.

Recently, Heilman et al.[20] showed how to attack the Bitcoin blockchain system by manipulating the node in the blockchain network. In a nutshell, the attacker could possess a large number of IP addresses at his/her disposal and control a large number of machines (e.g. a botnet). Alternatively, the adversary might be an Internet service provider or a nation–state adversary.

The intuition behind eclipse attack is straightforward. Eclipsing entails blinding the view of the victim from the blockchain and requires that the adversary is able to isolate the honest node out of the network by monopolizing all of the victim's outgoing and incoming connections. It is achieved by exploiting the way in which Bitcoin clients store the IP addresses that are advertised in the network. Namely, in Bitcoin, peers exchange *addr* messages that contain IP addresses and their time stamps. These messages are used by nodes to obtain network information from peers. Public IPs are stored at each node in two tables: tried and new tables. The tried table consists of 64 buckets, and each can store 64 unique addresses from peers with whom the node has established communication before. The node also keeps the time stamps of each tried IP address. When the node connects to a new peer, his/her address and the time stamp are put in the tried storage. If the storage is full, the current address will be inserted at random location in the bucket and will replace the address that was stored there before.

The PoW algorithm applied makes the system immutable. A valid PoW means that the miners are proving that they do a certain amount of work (solve the cryptographic puzzles) to produce a block. If the attackers want to change a block, they must perform the same amount of work that went into creating the block, then do the same work for every single block that follows it because those blocks would be invalid since a block in their history has changed. This mechanism prevents Sybil attack activity and makes the block become immutable.

Double spending is also a concern in blockchain technology. Double spending is a failure in the digital payment system. It can be described as the payee spends the same token twice to purchase some stuff literally in the digital payment system, and one amount of token only can be used for one time.[21] In the energy sector, the double spending can be extended to the ability of the prosumer to sell the ownership of the energy twice with same times tamp and tokens. The race attacks assume that the malicious will be able to send two or more conflicting transactions in the network. The victims who accept a payment immediately upon seeing "*0*" = *unconfirmed* are exposed to this attack. Finney attack is also a kind of double spending attack:

- It only works if the payee received the unconfirmed transactions.
- The condition still works if the payee waits a few moments to verify that miners in the network agree the payer was paid.
- It requires the malicious to be mining and controlling the block, in particular less than 50% of the network hash rate. The system is secure as long as the majority of the nodes in the network are honest nodes.

In the blockchain network in this system, we assume there are malicious node and the honest node which always solves the PoW puzzles to find the target value before adding to a new valid block (from the honest node) in the blockchain. The system is like a race between the honest node and the malicious node. In fact, if the attacker wants to control the node then the attacker is supposed to have a very powerful computing power to control the node. The probability of malicious for catching up the new block is as follows:

$$y = \begin{cases} 1, & \text{if } r \leq s \\ \left(\dfrac{q}{p}\right)^{z}, & \text{if } r > s \end{cases} \tag{7}$$

where r = probability for an honest node to find the block; s = probability of the attacker.

Define r is the mining power and probability from the honest node of the miner to add the new block. Then, define s as the attacker's probability to find the up block using their computer and relying on the computational power. Let $r + s = 1$; in this case, we assume the race between the malicious node and the honest node and that only one of them will win a block for every round. The malicious node's progress will be described as a *Poisson distribution* as follows

$$\lambda = z\frac{s}{r} \tag{8}$$

Sybil attack—the Sybil attack may happen in the decentralized overlay network.[22] This attack is described as a single attacker (see Figure 12) controlling several nodes on a network for some purposes. It is very risky for the new user because the user does not realize that she or he is in the malicious node and somehow the malicious node can

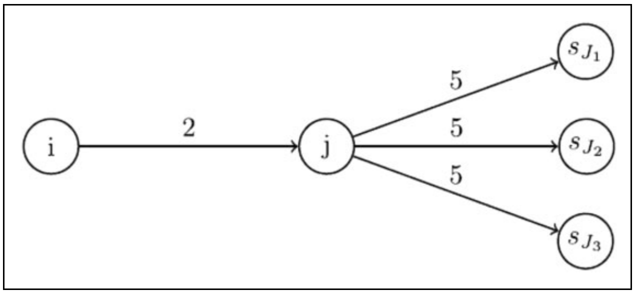


Figure 12. Sybil attack performed by agent *j*.

control multiple computers, virtual machines, and IP addresses. The malicious node is able to make several accounts from different IDs such as name and e-mail, and they pretend that they really exist in different countries of the world.

Preimage attack—It is also called as pre-mined block propagation delay attack to overcome the case when the victims are waiting for n confirmation before they are accepting a transaction through the pre-mining and on purposes for delaying propagation of $n + 1$ created blocks to generate a fork and make attacker's blockchain become longer than before. This attack on cryptographic hash function is also referred to as an activity of the malicious node to find the original value of the message.[23] The hash function must be resisted for this attack because it has some properties such as:

- *Preimage resistance* must be very difficult to find original value of the data, that is, given t, it is extremely difficult to find $h(d) = t$.
- *2nd preimage resistance*, it should be computationally difficult, that is given d, it is difficult to find a second preimage $y(d) = y(d)$.

The digital signatures in the blockchain systems guarantee the integrity of the PHI data, authentication, and non-repudiation of the transactions. The systems are implemented using elliptic curve cryptography (ECC) because of the length of keys. With the same size of security length, for example, 80 bits, ECC needs only 160 bits key size instead of Rivest-Shamir-Adleman (RSA) 1024 bits, and for the maximum size of ECC 521 bits, it provides security length of 256 bits while RSA needs 15,360 key size to provide it. ECC performance (Table 5) evaluation shows that it is more efficient (keyless) for the signature generation compared to RSA algorithm. *secp256k1* applied in ECC refers to the parameters keys of the ECDSA curve used in Bitcoin and some other platform. Furthermore, secp256k1 has identity p, which is defined over the prime field Z.

Table 5. Comparison of key-length.

Security strength (bit)	ECC (bit)	RSA/DSA/DH (bit)
80	160	1024
112	224	2048
128	256	3072
192	384	7680
256	521	15360

ECC: elliptic curve cryptography; DH: Diffie Helman; DSA: digital signature algorithm; RSA: Rivest-Shamir-Adleman.

Limitations

In terms of medical record, there are various formats of PHI data, such as health information system, electronic medical record, PHR, EHR, and it is essential to clearly know the exact characteristics of those data in practice. Hence, we bounded that the PHI data that is used in this system is the standard format that has been agreed by the health-care provider.

The limitation is also related to the number of health-care providers who join in the same blockchain network. We assume the identity of the health-care provider is verified by the system before joining in the private blockchain. Another limitation is related to the number of data blocks which is composed of many data from several providers with some attachment (up to 40 MB). The data sizes are usually up to several tens of megabytes.[5]

Challenges and opportunities

To ensure security and privacy in practice, the system needs more additional components such as authenticator, digital signer, and distributor to support the system. Authenticator component[5] ensures authorized access and proper attribution profile as well as preventing unauthorized access, blocking and providing lost access recovery mechanisms. When entering on the network, the user must have an ID generated for health records identification. The ID creation follows the OpenID code, which is used to identify users[24] to avoid duplication of the user ID.[25]

In relation to the architecture model, the system needs a service module (translator component). This component aims to convert and equalize the communications[26] with the heterogeneous health care system. Regardless of all that, the proposed system seems remarkably promising to be developed. A complexity is also inherent in the system because blockchain technology involves an entirely new vocabulary for the network size.

Conclusion

In this article, we propose an architectural model to manage the PHI data using blockchain technology and several protocols embedded. The PHI data are derived from several health-care providers in the same blockchain network. By leveraging on our model, the parties enable to collect and manage effectively the PHI data in a single view and also guarantee for the data integrity. The proposed model is remarkably promising to be developed to create a better decentralized health-care system. For the future work, the model needs to be evaluated, especially the strategy to prevent several attacks in the P2P network.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of following financial support for the research, authorship, and/or publication of this article: This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government(MSIT) (No.2017-0-00156, The Development of a Secure Framework and Evaluation Method for Blockchain) and partially supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2017-2015-0-00403) supervised by the IITP(Institute for Information & communications Technology Promotion).

ORCID iD

Sandi Rahmadika http://orcid.org/0000-0002-7848-6579

References

1. Krawiec RJ, Housman D, White M, et al. (2016). Blockchain: Opportunities for health care. Proc. NIST Workshop Blockchain Healthcare, (August), 1–16. Retrieved from https://www.healthit.gov/sites/default/files/4-37-hhs_blockchain_challenge_deloitte_consulting_llp.pdf
2. Sikorski JJ, Haughton J, and Kraft M. Blockchain technology in the chemical industry: machine-to-machine electricity market. *Appl Energy* 2017; 195: 234–246.
3. Tarau AN, De Schutter B, and Hellendoorn J. Centralized, decentralized, and distributed model predictive control for route choice in automated baggage handling systems. *Control Eng Appl Inform* 2009; 11(3): 24–31.
4. Xia C and Song S. Resource allocation in hierarchical distributed EHR system based on improved poly-particle swarm. In: *2012 5th international conference on biomedical engineering and informatics, BMEI 2012*, Chongqing, China, 16–18 October 2012, pp. 1112–1116. IEEE.
5. Roehrs A, da Costa CA, and da Rosa Righi R. OmniPHR: a distributed architecture model to integrate personal health records. *J Biomed Inform* 2017; 71: 70–81.
6. Safavi S and Shukur Z. Conceptual privacy framework for health information on wearable device. *PLoS One* 2014; 9: 12.
7. Simon SK, Anbananthen KSM, and Seldon L. A ubiquitous personal health record (uPHR) framework. In: *Proceedings of the 2013 international conference on advanced computer science and electronics information (ICACSEI 2013)* (ed Fangli Zheng), Beijing, China, 25 July 2013, vol. 41, pp. 423–427. Atlantis Press.
8. Avancha S, Baxi A and Kotz D. Privacy in mobile technology for personal healthcare. *ACM Computing Surveys* 2012; 45(1): 1–54.
9. Schollmeier RA. Definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In: *Proceedings—1st international conference on peer-to-peer computing, P2P 2001*, 2001, pp. 101–102.
10. Bandara HMND and Jayasumana AP. Collaborative applications over peer-to-peer systems-challenges and solutions. *Peer Peer Netw Appl* 2013; 6(3): 257–276.
11. Karame G and Androulaki E. *Bitcoin and blockchain security*. Artech House, 2016.
12. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system, http://www.bitcoin.org (2008, accessed 03 November 2017).
13. Rahmadika S, Rusmin PH, Hindersah H, et al. Providing data integrity for container dwelling time in the seaport. In: *Proceedings—2016 6th international annual engineering seminar, INAES 2016*, 2017, pp. 132–137.
14. Chord algorithm. *Online*, https://flylib.com/books/en/2.959.1.141/1/ (accessed 25 October 2017).
15. Stoica I, Robert TP, David L, et al. Chord: a scalable peer-to-peer lookup protocol for Internet applications. *IEEE/ACM Trans Netw* 2003; 11(1): 17–32.
16. Chord protocol. Online, https://en.wikipedia.org/wiki/Chord_(peer-to-peer) (accessed 11 November 2017).
17. Jaime GJ and Alfonso GC. Overview and challenges of overlay networks: a survey. *Int J Comput Sci Eng Surv* 2011; 2: 19–37.
18. Doval D and O'Mahony D. Overlay networks: a scalable alternative for P2P. *IEEE Internet Comput* 2003; 7(4): 79–82.
19. Karl W and Gervais A. *Ethereum eclipse attacks*. Zurich, Switzerland: ETH Zurich, 2016, pp. 1–7.
20. Heilman E, Kendler A, Zohar A, et al. Eclipse attacks on bitcoin's peer-to-peer network. In: *USENIX security Symposium*, 2015, pp. 129–144.
21. Karame GO, Androulaki E and Capkun S. Two bitcoins at the price of one? Double-spending attacks on fast payments in bitcoin. In: *IACR Cryptology ePrint Archive*, 2012, p. 248.
22. Otte P, et al. TrustChain: A Sybil-resistant scalable blockchain, Future Generation Computer Systems (2017), pp 1–11. http://dx.doi.org/10.1016/j.future.2017.08.048 (accessed 24 November 2018).
23. Rogaway TSP. Cryptographic hash-function basics: definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance. *Fast Softw Encryption* 2012; 3017: 371–388.
24. Hussein S and Badr S. Healthcare Cloud Integration Using Distributed Cloud Storage and Hybrid Image Compression. *Int J Comput Appl* 2013; 80(3): 9–15.
25. Chord. *Online*, http://www.dcs.ed.ac.uk/teaching/cs3/ipse/chord-desc.html (accessed 18 November 2017).
26. Narayan P. *Building blockchain projects: develop real-time DApps using Ethereum and JavaScript*. Birmingham-Mumbai: Packt Publishing.