Scalable Architecture for sharing EHR using the Hyperledger Blockchain

Andressa Fernandes
Federal University of ABC – UFABC
andressa.fernandes@ufabc.edu.br

Vladimir Rocha
Federal University of ABC – UFABC
vladimir.rocha@ufabc.edu.br

Arlindo F. da Conceição
Universidade Federal de São Paulo
arlindo.conceicao@unifesp.br

Flavio Horita
Federal University of ABC – UFABC
feahorita@gmail.com

***Abstract*—Blockchain technology has been applied in several areas, ranging from financial to health. Although this technology offers benefits, for instance the immutability, anonymity, and decentralization of information, its use presents some problems, been the scalability in terms of storage size one of them. This paper presents a scalable architecture for sharing electronic health records using a multi-channel hyperledger blockchain. The architecture uses one blockchain to record patient visits and one blockchain for each health institution to record links that point to Electronic Health Records (EHRs) stored in external systems. Among the main conclusions, the final results highlight the scalability of the proposed architecture, when compared to the related work models in the heterogeneous network.**

***Index Terms*—Blockchain, EHR, Hyperledger, Healthcare**

I. Introduction

Health Information Systems (HIS) are essential for improving efficiency in patient care. HIS provides health professionals the means to gather, compile, and analyze patients' data to make the best decision of its treatments. However, the information unavailability (by theft, loss or server outage) or how reliable is it (by modifying its integrity) in a distributed and unreliable system are still open problems to overcome [1].

Blockchain is a technology that enables reliable and secure information storage using a combination of distributed consensus procedures with cryptography techniques [2], [3]. The use of this technology started in 2008 with the creation of Bitcoin cryptocurrency by Satoshi Nakamoto [4]. The main objective of Bitcoin was to allow anonymous, auditable, reliable and secure financial movements, preventing trusted third parties, like banks, to intermediate these movements.

The concept used in Blockchain is the *distributed ledger*, representing an ordered and consistent chain of financial transactions, distributed across multiple nodes on an untrusted peer-to-peer network, using the proof of work to perform the consensus among them.

Following Bitcoin's success, other Blockchain technologies were developed to expand its usage beyond the financial context. In this sense, technologies as Ethereum [5] and Hyperledger [6], include the use of smart contracts [3], small programs stored in Blockchain, that allow performing operations on it. The use of smart contracts expands the power of Bitcoin's Blockchain, allowing to store states depending on the domain where the Blockchain is applied. In this case, the state not only needs to be an account balance in the financial context but any information, a patient medical record in the context of HIS [5].

Blockchain is expected to have a significant impact over the next few years [7], and to be used in several areas, for instance, supply chain management, Internet of Things (IoT) [8], Health [9], among others.

In the health domain, Blockchain can be applied in several contexts, for example, to control the access of sensitive data, to audit the medical financial payments, to allow transparency on the inventory drugs movements, to register the patients' appointments and visits, among others. Recent research, however, points out that despite its potential, Blockchain cannot be applied in all cases [10].

A problem that arises in Blockchain is the lack of scalability given by the amount of information that must be transmitted and stored on computers that are part of the system. To put in perspective, in the financial domain, the total financial transactions made by Bitcoin have reached 400 million transactions in ten years [11], Although this size is not a problem for today's hard disks, while the total transactions exceeds billions of transactions, the required space will become a problem, as analyzed in section VI. In this context, in Brazil's health domain, there were 1.4 billion patient visits just in 2018 carried out by its Unified Health System (SUS) to the government report [12]. In China, to [13], there were approximately 7 billion patient visits in 2017. In this context, if each patient visit is considered one transaction in the chain, the total amount of space and the architecture to support it must be considered to successfully apply the Blockchain in this domain.

The discussion above is the main motivation for this paper, in which we propose a scalable architecture for sharing Electronic Health Records (EHR) among patients, health professionals, and health institutions. For that, we define the components necessary to record the patient visits in the chain, and how these components must be related to support the access of EHRs stored in external storage systems, for example, clouds or HIS.

The rest of the paper is organized as follows. Section II

978-1-7281-4659-1/20/$31.00 ©2020 IEEE
DOI 10.1109/ICSA-C50368.2020.00032

presents the general concepts with regards to the Blockchain and the Hyperledger. Section III discusses related works. Section IV presents the proposed scalable architecture and Section V focuses on the components interaction through three common usage scenarios. Section VI compares the architecture with some existing alternatives in terms of size scalability. Finally, we conclude our work and point out some future directions in Section VII.

II. Blockchain

Blockchain is a technology that implements this similar to a *distributed ledger*, whose main characteristics are decentralization, availability, integrity, auditability, and privacy [14].

Decentralization allows information stored in the Blockchain to be replicated across multiple computers, avoiding a single point of failure of centralized servers. Availability permits to access the information whenever it is needed, regardless of whether some computers fail. Integrity is related to information preservation, protecting it from improper changes. Auditability offers the ability to track all information that has been stored in Blockchain. Finally, privacy allows members to remain anonymous.

In the following subsections, we present the concepts that comprise the Blockchain that allow implementing the aforementioned characteristics.

A. Blocks and Transactions

From a technological point of view, the Blockchain consists of an orderly and consistent chain of blocks, each one having a header and the data to be stored.

The header contains several attributes, some of them were the identifier, the previous block, and the signature. The identifier represents a globally unique value, created with a mathematical function (hash function) that encloses all block information. The previous block is responsible for chaining the blocks. For this, each new block added in the chain will have the value of the identifier of the previous block, thus creating a logical chain of links. The signature allows identifying the creator of the block.

The data registers the transactions belonging to the block. These transactions can represent any type of information. For example, in a financial system, the transaction would be a monetary transfer between two people. In health information systems, in turn, the transaction could be the document (or a URL address) that contains the patient's electronic health record.

B. Replication in peer-to-peer (P2P) networks

Offering resources (documents or media files) has been deployed since the very beginning of the internet with the client/server architecture [15]. In this architecture, a computer called server provides the resources to client computers, which are responsible for requesting them.

A P2P network, used by Blockchains, is a connected network of computers where each of them, called *peer*, can be both client and server at the same time. The P2P network emerges as an alternative to the client/server architecture, whose main problem is the server failure, which lets resources inaccessible for requesting them. Thus, the main objective of the P2P network is to increase the availability of the resources [16].

In this goal, each time a peer downloads a file, e.g., the chain of blocks, this file will be replicated on both, that can be downloaded by a third peer. Note that the resource availability will increase each time it is downloaded.

P2P networks are currently used not only for chain replication in the Blockchain technology but also in many other systems. The best-known example of its application is in file sharing via BitTorrent [17] and Skype video conference [18].

C. Permissioned and Permissionless

To Buterin [19], there are two types of Blockchains: permissioned and permissionless. In a permissioned Blockchain, any entity (whether to be a person, institution, etc.) could perform operations in the chain. In this sense, the Blockchain must allow those mutually unreliable entities to perform modifications on the chain without relying on a centralized trusted third-party. On the other hand, permissionless Blockchains were proposed with the objective that only a set of entities perform operations in the chain. In this sense, the permissionless Blockchain mainly operates with a known and frequently verified set of participants.

In the permissionless Blockchain, the entity becomes a participant of the network after joining it and can execute all the activities allowed, for instance audit or propose transactions, among others. Examples of technologies that implement a permissionless Blockchain are Bitcoin [4] and Ethereum [5].

In the permissioned Blockchain, the entity must be authenticated and authorized before to join the system, after which it will be transformed into a participant. As a participant, he can execute just specific activities, as view the blocks or audit the chain. Validation of an entity is usually performed by a trusted entity for managing permissions, called a certificate authority. An example of the technology that implements a permissioned Blockchain is the Hyperledger Fabric [6].

D. Smart contracts

The evolution of Blockchain has been classified into two generations. The first generation (used in Bitcoin) was oriented towards financial transactions, where value is transferred from one entity to another. For this, the features (hard-coded statically in the Blockchain) need to ensure that the entity has the necessary balance to perform the transaction. Thus, when a computer receives the transaction, it uses these features to maintain system consistency.

The second-generation (used by Ethereum and Hyperledger) appears around 2015. The main innovation is that new features can be dynamically added, meeting business requirements of different contexts, not only the financial one. For example, in health information systems, it is used to record the patient's health records. These features are called smart contracts, a concept proposed by Szabo in 1996 [20].

E. Hyperledger

Hyperledger [1] is an organization that provides several open-source Blockchain, been the Hyperledger Fabric one of them. Its goal is to provide a decentralized platform that allows creating specific functionalities for each kind of business through the use of smart contracts.

The Hyperledger Fabric is composed of certificate authorities, committing peers, endorser peers, orderers, and clients. Also, the components communicate among them using channels, that are structures specifically created to allow transactions privately and confidentially, isolating different domain applications. Thus, a channel is how components can communicate securely and reliably within the Blockchain.

Fabric certificate authorities are responsible for two tasks: first, to make sure that any component (user or smart contract) that wants to use the system is whom it claims to be (in other words, by recognizing the authenticity of the component); second, to authenticate the component and authorize it to use certain functionalities (e.g., perform transactions) or access other components after certification.

Committing peers are responsible for persisting the chains transmitted through the channels created in the system. Thus, they store the various Blockchains, one for each channel created. This *'one chain per channel'* approach brings two benefits: privacy and scalability.

Concerning privacy, a component will not be able to access (i.e., viewing or modifying) a chain from a committing peer associated with the channel if the component does not have access to that channel. In this context, privacy and confidentiality of the information are safe from components that did not.

With regards to scalability, note that having several Blockchains, one per channel will allow distributing the number of transactions and information stored among several committing nodes, increasing both the number of requests that a node could fulfill and the amount of data stored, thus increasing the system scalability.

Endorsing peers are responsible for two tasks: first, for collecting transactions from customers; second, for analyzing, using smart contracts, whether the transaction has any associated policies or rules that must be followed.

Ordering peers are responsible for two tasks: first, for receiving client transactions; second for ordering these transactions to maintain the Blockchain consistent (i.e., the same on all committing peers). In this sense, all ordering peers acting on a specific chain must agree on the order in which transactions will be added to it by committing peers. For that, transactions are received using Apache Kafka [21] technology, which allows them to be stored in a distributed and fail-tolerant way. Ordering, in turn, is achieved using Zookeeper [22].

Client applications are responsible for performing transactions in the system, sending them to the endorsing and ordering peers. In Hyperledger Fabric, a client could be a real person (who uses the system through an application), or other systems (who use the system through communication interfaces).

[1] https://www.hyperledger.org

III. Related work

This section describes relevant works that use the Hyperledger Blockchain for sharing EHRs. Although this citation-sequence is not extensive, it explain chronologically how the state of the art evolved in this research field.

In [23], the authors propose an architecture that uses a blockchain for sharing data in a mobile environment. The architecture focuses on the user, as a centerpiece of data sharing, and is comprised by six entities: users, wearable devices, healthcare providers, health insurance companies, the hyperledger blockchain, and a database cloud. The user is the owner of the data, responsible for securing, denying and revoking access to other entities. The wearable devices entity format the data collected by the devices and send it to the user account and the blockchain entity. The healthcare providers (a physician or a nurse) take care of users and can visualize its EHRs given the consent provided by them. The insurers' entity is used by the user to request some quotes to the insurers' companies, choosing the most appropriate health plan. On the other hand, the companies may request access to user data also. The blockchain network is used for collecting data from wearables and healthcare providers, for recording the permission access generated by the user, and for accessing the EHRs stored in the database cloud. The database cloud stores user information, data requests, data access information, and privacy policies.

In [24], the authors propose a Hyperledger-based architecture focusing on managing the user identity and privacy over the network. For that, the architecture distinguishes patients from doctors and third-party users. Besides, the architecture uses an external storage system to store the EHRs, while the Blockchain only registers pointers (as links) to the EHR location. To access these records, the architecture deploys Hyperledger's smart contracts for verifying patient's read or write policies.

In [25], the authors propose a conceptual architecture focusing on identity and access management. The architecture is comprised of four modules: the clients, an application server, a database, and an authentication server. The client (a mobile or web app) communicates with the application server, which, in turn, sends the user credentials to the authentication server to verify the client. If the verification is succeeded, the client could access the database (for retrieving the EHRs) through the application server. The purpose of separating authentication and application servers is to decouple the logic and allow modularity. Thus, different application servers could access a different database using the same authentication server. In this architecture, the Hyperledger Blockchain is used to register the permissions (grant and revoke) given by the client to other clients to access its EHRs. Although the work also analyses the architecture scalability, it was deployed only in a single organization.

In [26], the authors propose the EMRShare, a three-layer architecture for sharing electronic medical records. The layers (Blockchain, Service, and Dispatcher) are used to facilitate the integration of the architecture with existing external systems. In EMRShare, the dispatcher layer is responsible for receiving client requests and to start the corresponding service in the service layer. The service layer encapsulate all blockchain requests coming from the dispatcher and is responsible for sending the requests to the blockchain. The Blockchain layer is responsible for registering access permissions, allowing reads and writes to the chain, and for storing the record metadata. It is important to note that EMRShare uses an external storage system to store the EHR. This work also analyses the architecture scalability, deploying it in four organizations.

In [27], the authors propose an architecture-specific to monitor patients that are using IoT and wearable devices. The architecture is comprised of two blockchain groups: medical devices blockchain and consultation blockchain. The former is responsible for recording all health data collected by the devices, and the latter is responsible for recording all patient's EHRs. It is important to note that in the blockchain of the first group, each patient creates and has access to its blockchain, but is persisted only during the period of treatment. The second blockchain, in turn, will be always available, but only for health professionals.

In [28], the authors describe a layered software architecture for supporting emerging healthcare and life science use cases, being sharing EHR's one of them. The architecture is comprised of a front-end layer, API services layer, smart contract layer, the hyperledger blockchain layer, and the HIPAA cloud layer. The front-end layer is responsible for capturing the client requests and call specific services defined in the API layer. The API layer provides REST APIs to access the business-specific rules implemented in the Smart Contract layer. The Smart Contract layer is divided into contracts specific to the core business and in reusable modular components, transversal to the contracts aforementioned, as that data encryption, permission access, notifications, among others. The hyperledger and the cloud layer, both are deployed in a cloud computing environment. The former is used as a smart contract container and the latter is responsible for register all health data, based on the Health Insurance Portability and Accountability Act (HIPAA). This architecture only uses one blockchain, which can lead to some scalability issues but could be enhanced to use several blockchains.

For a more recent review, please see [29] and [30] that also address the use of Blockchain in Healthcare for sharing EHR from a research perspective.

Although the alternatives use hyperledger for sharing EHRs, our architecture focus on system scalability in a multi-institution environment, each one having a different and restricted infrastructure to support the blockchain deployment. To the best of our knowledge, none of the previous works analyses the scalability in terms of storage and how the Hyperledger components could be related to allowing sharing EHRs in this environment.

IV. Proposed Architecture

Figure 1 presents the overall architecture used by our work. The architecture is comprised of four components, connected, which are the clients, the Blockchain, the storage system, and the certificate authorities. In the following, we present the details of each component.

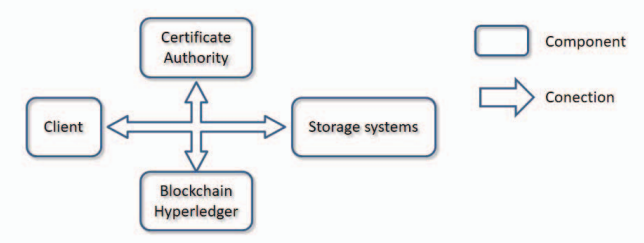

Fig. 1. Architecture for sharing EHR based on Hyperledger Blockchain.

A. Client component

The client component is an electronic device (a smartphone or a laptop) that represents a real-world entity in the healthcare context. For example, it may be a person (a patient or a physician) or a health institution (a hospital or a clinic). The component is responsible for: (i) sending and receiving patient electronic health records; (b) allowing or denying the access of its records to other clients. As the client must interact with all other components, it communicates with them through reliable and secure APIs (Application Programming Interfaces). Reliable in the sense that the information will not be modified and secure in the sense that the information will not be accessed by clients without permission. Besides, the component stores the client's public and private keys. The former represents a unique entity identifier that will be used for other entities to validate the entity's information. The latter is used by the entity for signing the events performed by it. In this paper, the credential loss issue is not being analyzed.

B. CA component

The Certificate Authority (CA) component is responsible for creating the entity credential (i.e., both public and private keys) and for delivering it to the entity, which will store them into the client component. For this, first, the CA must validate if the entity that wants to use the architecture present valid and trusted information. In our architecture, we divide the CA component into three sub-components: Patient CA, Health Professional CA, and Health Institution CA.

Patient CA (PatientCA) component is responsible for validating patient' information before delivering the keys. If the information presented is valid, the CA generates the keys and send them to the patient. From this point, the patient becomes a client of the system. Note that the patient anonymity holds in that it will use only an identifier (the public key) to validate its actions and not real-world information, as its name or its social security number. In our architecture, this component can be deployed by each health institution but is more common

to be deployed by a more country-wide organization, as a federal government ministry of health. If both alternatives are deployed, the latter will be responsible for validating the entities information and creating its access credentials (i.e., the keys) and the former will be responsible for allowing access to the institution's information (i.e., giving access to the communication channel and therefore its associated blockchain).

Health Professional CA (ProfessionalCA) and Health Institution CA (InstitutionCA) both also need to validate health professional and institution information, respectively. Both could be deployed by each state (or region, depending on the country), or also by the ministry of health.

C. Storage Component

The storage component is an external system responsible for providing the storage of electronic documents, specifically the patients' electronic health records (EHRs). This component is divided into two sub-components: a personal storage module and a HIS storage module.

In the personal storage module, the client could storage EHRs in its cloud storage account, for instance DropBox, Google Drive, Microsoft OneDrive. among others. With this approach, the architecture provides the means to allow patients for owning their information, instead of trusting in third-parties to do it. Note that at any moment the patient can deny access to its EHRs without asking permission from anyone.

In the HIS storage module, we assume that health institutions already have repositories to storage their patients' information. Also, it is assumed that the module allows both searching medical records (e.g., using SQL queries or REST methods) and transferring them in a scalable way.

Besides, independently of the sub-component used, we assume that the storage system has a module that allows the client authentication and authorization for accessing its stored resources (e.g., using OAuth2, OpenID, among others) and an access point (e.g., an URL through a REST service) to obtain the EHRs.

D. Hyperledger component

Our architecture is comprised of a global blockchain and several smaller local blockchains, one for each health institution.

The global blockchain is responsible for recording all patients visits to a health institution. For this blockchain to be deployed, each health institution must have a committing peer that stores the chain. It is important to note that even if a patient visits the same institution several times, this blockchain will store just one of these visits.

The blockchain for each health institution (called local Blockchain) is responsible for recording all EHRs associated to its patients. For this blockchain to be deployed, we assume that the health institution has the minimal infrastructure required just to run a hyperledger network. Our work defines the health institution infrastructure in two groups, based on its devices' processing power, network connectivity, and IT employees.

- *Large health institution*: it has a cluster of 5 or more servers; 8 GB or more of RAM for each server; distributed hard disk of 5 TB or more; 100 Mbps of bandwidth. Besides, it has a fixed IT team.
- *Small health institution*: it has 5 or fewer servers (not configured as a cluster); 2 GB or more of RAM; a hard disk of 1 TB; 10 Mbps of bandwidth. Besides, it does not has a fixed IT team.

V. Components Interaction

To understand how the components interact, we analyze the architecture lifecycle based on three common usage scenarios for sharing EHRs as defined in [1].

A. Assumptions for demonstration scenarios

For all scenarios described below, we assume the following initial conditions: (i) both health professionals and institutions were registered by the Professional CA and Institution CA, respectively; (ii) A Patient CA CA_{wc} was deployed, but each institution could have deployed its Patient CA also; (iii) Each institution has created its communication channel with the registered health professionals that work on it. As a consequence, there is one Blockchain per institution.

Besides, we assume that a global Blockchain b_{wc} was deployed. As mentioned, it is responsible for registering only the institutions visited by the patient. Thus, even if a patient visits the same institution multiple times, this Blockchain will register only one of them. It is worth noting that all entities have access to it.

B. Scenario 1: Patient visiting a hospital

In the first scenario, a person (not registered in the system) visits a physician m_1 who works in hospital h_1. Note that, as the person is not registered, she does not have any previous EHR. Figure 2 shows the steps required for the physician treats the person.

In step 1, the person (using a mobile app $app1$) is registered in CA_{wc} as a patient, obtaining a unique identifier p_1 and its public and private keys. At the end of this process, the patient will become a client, whose access credential will be stored on its electronic device (mobile phone). In step 2, patient p_1 request a *read-only* access to hospital' Patient CA CA_{h1} (specifically for accessing the EHRs that will be created at the end of the visit). In step 3, the physician m_1 treats the patient and creates an EHR using the institution's HIS. Next, the HIS register the visit metadata in the hospital's local Blockchain b_1, for instance the patient identifier and how to access the created electronic document ehr_1. To access the EHR, it is possible to use a REST service with a URL (url_1) that receives the patient's credential and returns the encrypted documents associated with it (encrypted with the patient's private key). In step 4, the HIS will register in b_{wc} that patient p_1 visited h_1, only if it was the first time.

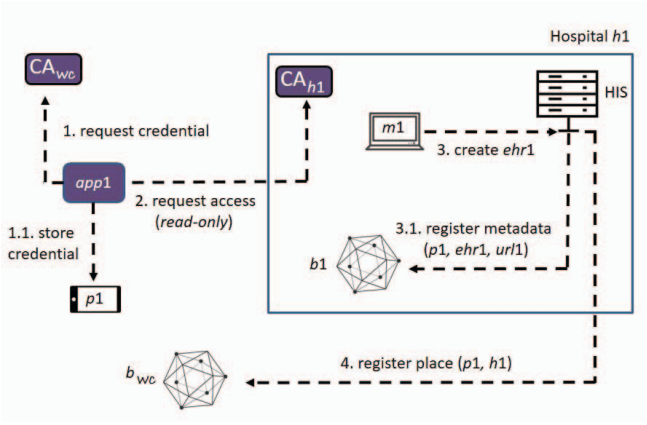

Fig. 2. A person, becoming a patient, visiting a hospital h_1.

C. Scenario 2: Patient visiting a Blockchain-connected heart clinic

In the second scenario, patient p_1 will perform an electrocardiogram at heart clinic h_2 with the same physician m_1. Note that m_1 works in both hospital h_1 and clinic h_2, therefore with access to both local Blockchains b_1 and b_2, respectively. Figure 3 shows the steps required for the physician to visualize the EHR ehr_1 created in the previous scenario and perform the exam.

In step 1, the physician m_1 requests to Blockchain b_{wc} the institutions that the patient p_1 visited, obtaining h_1. In step 2, the physician requests to Blockchain b_1 for the metadata of all documents associated with p_1. Specifically, she will obtain the visit metadata with a link to access the ehr_1. Next, the physician performs the electrocardiogram and gives the result to the patient p_1 as an electronic document ehr_2 (e.g., a PDF or Word document).

In step 3, p_1 sends ehr_2 to its personal storage module, in this example will be the Dropbox [2]. At the end of this process, the storage system will return a URL link url_2 that will allow accessing the document via the Internet. It is important to highlight that it is essential to encrypt the document with the patient's private key. In step 4, the physician register in b_2, through the HIS at h_2, the exam metadata, for instance the patient identifier and the link url_2. In addition, in step 5, the HIS also register in b_{wc} that patient p_1 visited h_2 (only for the first time). Finally, in step 6, patient p_1 request a *read-only* access to hospital' Patient CA CA_{h2}.

D. Scenario 3: Patient visiting a non-connected clinic

In the third scenario, patient p_1 will be treated at a walk-in clinic h_3 by a nurse m_2 who only works at hospital h_3 (therefore, he has no access to local Blockchains b_1 and b_2). Figure 4 shows the steps for the nurse to visualize the patient's EHRs ehr_1 and ehr_2 created in the previous scenarios.

In step 1, the nurse m_2 requests to Blockchain b_{wc} the institutions that the patient p_1 visited, obtaining h_1 and h_2.

[2]https://www.dropbox.com

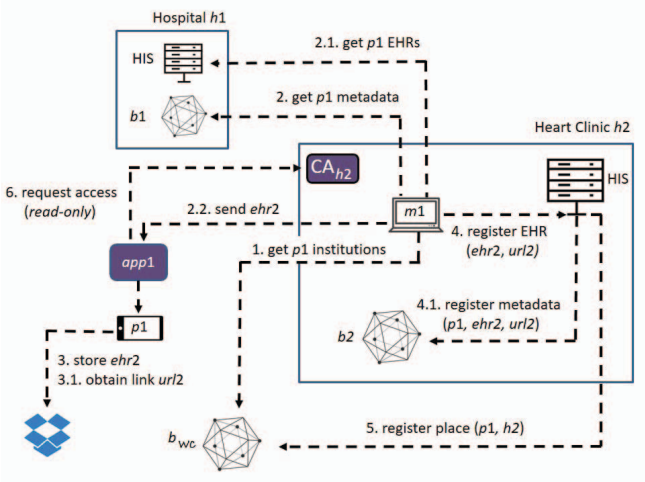

Fig. 3. Patient visiting a heart clinic h_2 with access to h_1.

In step 2, the nurse request the credential to p_1 and an *access code*, in order to access the EHRs stored in the hospital and heart clinic's HIS. This code, encrypted with p_1's private key, informs an HIS to allow the access to patient's EHR. In step 3, the nurse, using the code, obtains the encrypted EHRs (ehr_1 and ehr_2) from REST services deployed at the hospital h_1 and the heart clinic h_2, and decrypt them with p_1' public key. In step 4, the nurse treats the patient and creates an EHR ehr_3. Next, the HIS register the visit metadata in the clinic's local Blockchain b_3. In step 5, the HIS will register in b_{wc} that patient p_1 visited h_3. Finally, in step 6, patient p_1 request a *read-only* access to clinic' Patient CA CA_{h3}.

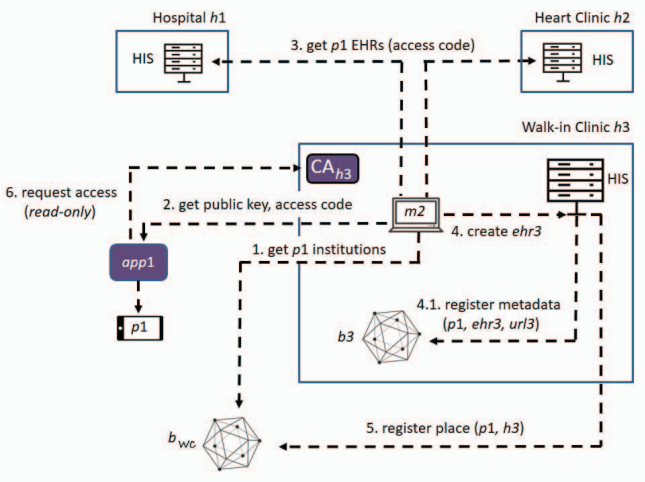

Fig. 4. Patient visiting a walk-in clinic h_3 without access to h_1 nor h_2.

VI. Architecture Analysis

In this section, we analyze the proposed architecture in terms of storage size scalability, comparing our multi-channel solution with the most used approach, which is to have just one Blockchain to records all events.

As mentioned in Section IV-B, the architecture needs the Certificate Authorities (i.e., PatientCA, ProfessionalCA) for being deployed until it is proper used. This prerequisite is not an easy task to do. To the best of our knowledge there are only very few countries using blockchain technology for health on a national scale, Estonia (with an estimated population of 1,3 million people [31]) was the first country adopting it [32]. For larger and decentralized countries, as Brazil, in which the country is divided into states and each of them has its own autonomy, it could be difficult to deploy a national CA. This difficulty arises because it is necessary to align the federal and state public policies, as well as the infrastructure designated to health. Of course, this behavior is inherent to all systems that use a permissioned Blockchain approach.

For estimating the storage scalability size of the architecture, composed by the global Blockchain and the local Blockchains, it is necessary to define the information that these Blockchains will store in its blocks, based on the primitive data types and sizes. Table I shows the main Hyperledger primitive data types, with its size, obtained from [33].

TABLE I
PRIMITIVE DATA SIZE ON HYPERLEDGER

Data	Size	Frequent uses
Int	256 bits	numbers in general
Bool	8 bits	flags
String	8 bits for each char	names and descriptions
DateTime	64 bits	dates and hours

The data model stored in the global Blockchain is shown in Listing 1 that is responsible for aggregating the patient's identification (PatientID) with the institution's identification (InstitutionID).

Listing 1. Data model for global Blockchain

```
struct UniqueVisit {
    String[20] PatientID;
    String[20] InstitutionID;
}
```

In this data model, each identification has 20 bytes (based in the data size of Table I), representing the public key of the entity. The total cost of this model, which will be persisted as a transaction in the block, is 40 bytes.

Now, we assume that a person visits one-to-ten different health institutions in his lifetime. As previously mentioned, the patient could visit several times the health institution, but the global Blockchain just persists in one of them. Table II shows the best- and worst-case scenarios to store all Blockchain visits transactions for Brazil, with a population estimated in 211 million people [34].

TABLE II
GLOBAL BLOCKCHAIN MODEL SIZE

Visits / Qty. Institutions	Visits x UniqueVisit (40 bytes)	In GB
211,836,829 / one	8,473,473,160	7.89
2,118,368,290 / ten	84,734,731,600	78,91

TABLE III
PROJECTION OF BLOCKCHAIN SIZE IN LARGE HEALTH INSTITUTIONS

Years	Patient Visits	Average year size (MB)
1	26,000	5.25
5	130,000	26.28
10	260,000	52.56
20	520,000	105.13

In the worst case, Table II shows that the entire chain fits in a normal hard disk because the higher value is approximately 79 GB (assuming, as mentioned, that a normal hard disk has 1 TB of size). This result indicate that the global blockchain could be deployed in both small and large health institution infrastructures.

Listing 2. Data model for Institution Blockchain

```
struct NewRecord {
    DateTime timestamp;
    String[20] publicKeyPatient;
    String[20] publicKeyDoctor;
    String[50] healthRecord;
    String[50] link;
    String[64] hashData;
}
```

In the institutional local Blockchain, we store the data model suggested by [1] whose purpose is to aggregate the patient visiting metadata, i.e., the patient and doctor entities, as well as the link pointing to the EHR. The data model is shown in Listing 2.

In this model, the attribute *timestamp* represents the date that the patient visits the doctor, each of them has its own identifier (*publicKeyPatient* and *publicKeyDoctor*, respectively). While, the attribute *healthrecord* represents some keywords that could be added to the record, e.g., electrocardiogram or blood exam. The third attribute *link* represents a URL that allows retrieving the EHR, associated with the visit, which is stored in some external storage system. Finally, the attribute *hashData* allows ensuring the EHR integrity by storing the hash value made from the document. As shown in Table I, the total cost of the *NewRecord* model is 212 bytes.

When estimating the total amount of data stored in each local blockchain, we first need to determine how many patient visits are in a large and small health institution. To [35], a large health institution in Brazil has 26,000 visits per year. On the basis of these data, Table III shows the projection of patient visits from one to twenty years, with the respective total blockchain size to store them, using the model of Listing 2.

When relying on a small health institution that may have 10,000 or fewer visits per year [36], Table IV details the projection for one to twenty years versus the Blockchain size.

Note that in both cases (i.e., large and small health institutions), the entire blockchain size has at maximum 100 MB, fitting in normal hard disks.

On the other side, it is important to comprehend that the size required by the related works that use just one Blockchain for storing all patient visits. In Brazil, to the management report

TABLE IV
PROJECTION OF BLOCKCHAIN SIZE IN SMALL HEALTH INSTITUTIONS

Years	Patient Visits	Average year size (MB)
1	10,000	2.02
5	50,000	10.10
10	100,000	20.21
20	200,000	40.43

of 2018 [12], there were 1,4 billion patient visits through its public healthcare system called SUS [37]. Based on this information, Table V shows the projection for one to twenty years using Listing 2.

TABLE V
PROJECTION OF SIZE USING JUST ONE BLOCKCHAIN

Years	Average size (TB)
1	0.26
5	1.34
10	2.69
20	5.39

Regarding the results of Table V, in the first year, large and small health institutions will be able to store the entire Blockchain in its hard disks. However, from the fifth year and forward, only the larger health institutions could store this model without made infrastructure changes, because of the 1 TB hard disk size limit of the smaller ones. From the twentieth year, even the large ones will have to increase its distributed storage servers. This finding thus provide evidence that it is necessary to divide the entire Blockchain into several smaller ones as proposed in our architecture, using the global and the local Blockchains.

VII. CONCLUSIONS AND FUTURE WORK

HIS is essential for providing relevant information about patients to support decision-making of health professionals. Blockchain is a technology that allows reliable and secure data storage and has been used in several areas, including health.

One of the problems using Blockchain in HIS is the size scalability for storing medical records that may arise when just one chain is used. To increase the scalability, we propose a multi-chain architecture and formulate the interactions of the components necessary to share EHRs. Through analysis, our architecture behaved in a more scalable way than the works using just one chain.

As future work, we will analyze the performance scalability of the architecture, that is, what should be the Hyperledger network needed to manage patient visits in terms of transactions per second.

ACKNOWLEDGEMENTS

FH would like to thank the financial support provided by CNPq (Grant Nro. 437937/2018-6) and FAPESP (Grant Nro. 2018/25805-5).

REFERENCES

[1] A. F. da Conceição, F. S. C. da Silva, V. Rocha, A. Locoro, and J. M. M. Barguil, "Eletronic Health Records using Blockchain Technology," in *Anais do I Workshop em Blockchain: Teoria, Tecnologias e Aplicações*. Porto Alegre, RS, Brasil: SBC, 2018.

[2] M. Swan, *Blockchain: Blueprint for a new economy*. O'Reilly Inc., 2015.

[3] F. Greve *et al.*, "Blockchain e a revolução do consenso sob demanda," *Minicursos do SBRC)*, vol. 36, 2018.

[4] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[5] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, 2014.

[6] E. Androulaki *et al.*, "Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains," in *Proceedings of the Thirteenth EuroSys Conference*, ser. EuroSys '18. New York, NY, USA: ACM, 2018, pp. 30:1–30:15.

[7] D. Tapscott and A. Tapscott, *Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World*. Brilliance Audio, 2016.

[8] S. Huh *et al.*, "Managing IoT devices using blockchain platform," in *Advanced Communication Technology*. IEEE, 2017, pp. 464–467.

[9] P. Zhang *et al.*, "Blockchain technology use cases in healthcare," ser. Advances in Computers. Elsevier, 2018.

[10] C. Monteil, "Blockchain and health," in *Digital Medicine*. Springer, 2019, pp. 41–47.

[11] B. Luxembourg, "Blockchain total number of transactions," https://www.blockchain.com/charts/n-transactions-total, 2019, last access: 02/07/2019.

[12] M. da Saúde, "Relatório de Gestão (in portuguese)," 2018, http://bvsms.saude.gov.br/bvs/publicacoes/relatorio_gestao_2018.pdf.

[13] Y. Chen *et al.*, "Blockchain-Based Medical Secure Storage and Medical Service Framework," *Journal of Medical System*, 2019.

[14] D. Drescher, *Blockchain Basics: A Non-Technical Introduction in 25 Steps*. Apress, 2017.

[15] M. v. Steen and A. S. Tanenbaum, *Distributed Systems: Principles and Paradigms (3rd Edition)*. CreateSpace Independent Publishing Platform, 2017.

[16] A. Oram, *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*. O'Reilly Media, 2001.

[17] B. Cohen, "The BitTorrent protocol specification," 2008.

[18] S. Baset and H. Schulzrinne, "An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol," *Proceedings - IEEE INFOCOM*, 2005.

[19] V. Buterin, "On public and private blockchains," *Ethereum blog*, 2015.

[20] N. Szabo, "Smart contracts: Building blocks for digital markets," http://www.alamut.com/subj/economics/nick_szabo/smartContracts.html, 1996, last access: 28 Mar. 2019.

[21] J. Kreps, N. Narkhede, and J. Rao, "Kafka: A distributed messaging system for log processing," in *Proceedings of 6th International Workshop on Networking Meets Databases (NetDB), Athens, Greece*, 2011.

[22] P. Hunt *et al.*, "Zookeeper: Wait-free coordination for internet-scale systems," in *Proceedings of the 2010 USENIX Conference*, 2010, pp. 1–14.

[23] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, "Integrating blockchain for data sharing and collaboration in mobile healthcare applications," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017, pp. 1–5.

[24] A. Juneja and M. Marefat, "Leveraging blockchain for retraining deep learning architecture in patient-specific arrhythmia classification," in *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, March 2018, pp. 393–397.

[25] T. Mikula and R. H. Jacobsen, "Identity and access management with blockchain in electronic healthcare records," in *2018 21st Euromicro Conference on Digital System Design (DSD)*, Aug 2018, pp. 699–706.

[26] Z. Xiao *et al.*, "Emrshare: A cross-organizational medical data sharing and management framework using permissioned blockchain," in *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, Dec 2018, pp. 998–1003.

[27] O. Attia, I. Khoufi, A. Laouiti, and C. Adjih, "An IoT-Blockchain Architecture Based on Hyperledger Framework for Healthcare Monitoring Application," in *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, June 2019, pp. 1–5.

[28] F. Curbera *et al.*, "Blockchain: An enabler for healthcare and life sciences transformation," *IBM Journal of Research and Development*, vol. 63, no. 2/3, pp. 8:1–8:9, March 2019.
[29] M. H. Kassab, J. DeFranco, T. Malas, P. Laplante, g. destefanis, and V. V. Graciano Neto, "Exploring research in blockchain for healthcare and a roadmap for the future," *IEEE Transactions on Emerging Topics in Computing*, pp. 1–1, 2019.
[30] S. Cichosz, M. Stausholm, T. Kronborg, P. Vestergaard, and O. Hejlesen, "How to use blockchain for diabetes health care data and access management: An operational concept," *Journal of Diabetes Science and Technology*, vol. 13, p. 193229681879028, 07 2018.
[31] Worldometers, "Estonia Population ," https://www.worldometers.info /world-population/estonia-population/, 2020, last access: 08/01/2020.
[32] Nortal, "Blockchain and healthcare: the Estonian experience," https://nortal.com/blog/blockchain-healthcare-estonia/, 2018, last access: 02/01/2020.
[33] H. Composer, "Hyperledger composer modeling language," https://hyperledger.github.io/composer/v0.19/reference/cto_language, 2019, last access: 17/01/2019.
[34] Worldometers, "Brazil Population ," https://www.worldometers.info /world-population/brazil-population/, 2020, last access: 08/01/2020.
[35] InCor, "Numbers of the Heart Institute - InCor (in portuguese)," http://www.incor.usp.br/sites/incor2013/index.php/conheca-estrutura /numeros, 2019, last access: 02/07/2019.
[36] J. Mendes, M. Cecílio, and V. Osiano, "Small sized hospitals from SUS in the state of São Paulo (in portuguese)," *BEPA [internet]*, vol. 11, no. 128, pp. 25–40, 2014, avaliable from: http://portal.saude.sp.gov.br/resources/ses/perfil/profissional-da-saude/ destaques//saude_em_dados_gais_17_hospitais_de_pequeno_porte.pdf.
[37] S. Silva and M. Santos, "Sistemas emergentes no ecossistema digital brasileiro de saúde pública: Uma abordagem sociotécnica," in *Anais Estendidos do XV Simpósio Brasileiro de Sistemas de Informação*. SBC, 2019, pp. 63–68.