2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering

On the design of a Blockchain-based system to facilitate Healthcare Data Sharing

Anastasia Theodouli, Stelios Arakliotis, Konstantinos Moschou, Konstantinos Votis, Dimitrios Tzovaras
CERTH / ITI
Thessaloniki, Greece
{anastath , saraklio, konsmosc, kvotis, tzovaras}@iti.gr

***Abstract*—Blockchain technology though originally designed for keeping financial ledgers, recently has found applications in many different fields including healthcare. Sharing healthcare data for research purposes will boost research innovation in this area. That being said, healthcare data sharing raises many privacy and security issues for the Patients who share their data. In this work, we present the potential of Blockchain technology to facilitate (i) private and auditable healthcare data sharing and (ii) healthcare data access permission handling by proposing a blockchain-based system architecture design.**

Keywords- blockchain; healthcare data; smart contracts; privacy; security; pseudonymity; auditing; data integrity

I. INTRODUCTION

A Blockchain consists of a continuously growing list of records called blocks. Each block represents a set of transactions and is cryptographically linked to its previous block thus forming a chain. A Blockchain is managed by a peer-to-peer network of nodes that validate new blocks using a consensus algorithm. The consensus algorithm ensures that the next block in a blockchain is the one and only version of the truth, thus preventing powerful adversaries from successfully forking the chain.[1] As a result, all nodes of the network contain the same replica of data, eliminating the need of a central trusted authority to manage data. [1]

Blockchain, being a cutting-edge technology and an emerging research field, has numerous applications to several domains, e.g. in cryptocurrencies, Digital Auctions, Digital Supply Chains, IoT and smart cities, Digital Identities, etc. [2] Applications of Blockchain to healthcare domain have been extensively explored to enable interoperability between several Health Units in a secure and auditable way. [3, 4, 5]

Access of medical research centers to healthcare data stored on Web / Cloud Clinical Platforms can have a positive impact on medical research innovation. In such a case, Medical Researchers can have access to a distributed 'pool of data' of medical treatments and healthcare outcomes based on values stored via eHealth and mHeath in web/cloud clinical Platforms. Moreover, by enabling medical researchers to filter out specific features of the data they are looking for, one could achieve a facilitation in the formation of demographic cohorts, and also enhance precision medicine.

[1]https://www.coindesk.com/short-guide-blockchain-consensus-protocols/

That being said, healthcare data are highly sensitive and Data Owners, i.e. Patients, may hesitate to share their data for research purposes despite the positive impact that such a sharing can have as outlined above, since an inappropriate disclosure of their data and/or of their identities could have a direct impact on their health, and/or indirect financial or social implications as regards their employers, involved insurance companies, and so on.

To alleviate Patients' concerns as regards their data sharing, we present our contribution, a blockchain-centric system architecture design which is used to ensure (i) shared data integrity, (ii) patient pseudonymity, (iii) auditing and accountability, and (iv) workflow automation by leveraging inherent properties of the blockchain technology like immutability, auditability, and accountability combined with the usage of smart contracts, a transaction-aware state-machine mechanism which enables a (quasi) Turing-complete fully-programmable logic in the way that the Blockchain state changes; these scripts are automatically executed upon a pre-defined set of rules included within the smart contracts. Moreover, the usage of smart contracts is quite tailored to our approach that tackles with complex workflows.

The remainder of this paper is structured as follows, in section 2, we discuss related work. In Section 3 we present our setting (blockchain model proposed, involved entities, incentives of involved entities for system adoption and level of trust among them). In section 4, proposed system architecture is presented. Section 5 presents the smart contracts used, while section 6 gives an overview of the supported use case scenarios and their corresponding workflows. Section 7 outlines the added value of the system. Finally, section 8 concludes the paper.

II. RELATED WORK

In this section, we discuss works that focus on healthcare data sharing/management leveraging (i) Blockchain infrastructure, (ii) other technologies such as cloud computing and big data.

A. Healthcare data management with Blockchain

Blockchain has been proposed as an appropriate infrastructure for healthcare data sharing by the authors of this work [4]. With an aim to facilitate healthcare data interoperability between institutions, the authors also introduced a new consensus algorithm, called 'Proof of Interoperability' that was based on conformance to the Fast Healthcare Interoperability Resources (FHIR) protocol. In this work [5], authors illustrate how to apply blockchain

2324-9013/18/31.00 ©2018 IEEE
DOI 10.1109/TrustCom/BigDataSE.2018.00190

technology in pervasive social network (PSN)-based healthcare.

Healthcare data is highly sensitive and there is a need to protect it from unwarranted access. Towards this end, authors of this work [3] present MedRec, a novel, decentralized record management system to handle Electronic Health Records (EHRs), using blockchain technology. The block content represents data ownership and viewership permissions shared by members of a private, peer-to-peer network. Via smart contracts on an Ethereum blockchain, they log patient-provider relationships that associate a medical record with viewing permissions and data retrieval instructions (essentially data pointers) for execution on external databases. MedRec architecture further extends its value proposition in empowering medical researchers to mine in the Blockchain network getting anonymized medical data as mining rewards.

In medChain project[2], a federated Blockchain based on the Ethereum platform is used for Patient Data Storage and Retrieval. Patient Data Privacy is ensured using encryption with Patient's private key and hashing of electronic Protected Health Information (ePHI) before being stored on the medChain Blockchain.

Authors of this work [6] propose the usage of smart contracts deployed on a private, permissioned Ethereum blockchain to govern Clinical Trial Authorization (CTA) details and a private IPFS network to store the data structure that holds the clinical trial protocol whenever large file storage is required with an aim to improve data transparency in clinical trials.

Various design aspects as well as technology requirements and challenges of a blockchain platform architecture for clinical trials and precision medicine have been discussed in frames of this work [7]. Usage of the public Bitcoin Blockchain network with an aim to enhance transparency and traceability of the Consent given by Patients involved in Clinical Trials is discussed in frames of this work [8].

B. *Healthcare data management without Blockchain*

Healthcare data sharing among interested stakeholders (e.g. public health institutions, research institutions, patients, etc.) as regards multi-source, heterogeneous data using Cloud computing and Big Data analytics techniques has been explored in frames of this work [9]. In the data management layer of their proposed architecture, they propose techniques based on distributed parallel computing and distributed file storage based also on memory analysis, to cope with scenarios of real-time analysis of big data stored on their infrastructure.

Compared with the techniques proposed in this work, in our design, we neither store nor process the collected data on-chain. What we store on-chain, is metadata (hashed data, data reference URLs, and permissions) that enable the data sharing in a secure, private and auditable way. As regards the Patient data kept on-chain, their storage can be regarded as $O(n)$, where n is the number of data records. This is due to the fact that we store *hashed* Patient data which are of a fixed length and thus not affected by the actual size of the data record before hashing. Other than that, due to the distributed nature of the Blockchain, data stored on-chain are replicated to all the nodes of the network avoiding a single point of failure for the system, i.e. if a node fails, assuming m network nodes, there are still $m-1$ nodes holding the data.

[2] https://www.medchain.us/#

III. PROPOSED SETTING

The involved entities, their incentives to use the system, the level of trust among them and the Blockchain model assumed are analyzed below:

A. *Patients*

Patients who want to share their healthcare Data acknowledging (i) the benefits of such a sharing regarding medical research boosting in general, (ii) the positive impact on their own healthcare treatment and outcomes in the long run. On the other hand, Patients don't want to compromise privacy and security of their Data when sharing them. Moreover, according to the proposed system design, there is no significant overhead for Patients to share their data to the Blockchain network. Patients use dedicated Web / Cloud Platforms to export their data in the appropriate format. In our setting, it is assumed that Patients are trusted and the data they upload are correct. Patients are also enabled to filter their historical data and check past transactions informing them who accessed their data, when, and what data did they access.

B. *Web / Cloud Platforms*

Web / Cloud Platforms having their own local databases that keep Patient healthcare data. They can export the Data in an appropriate format for sharing with the upper layers of the system. No need to be Blockchain nodes. They can use the upper layers of the system (see Architectural layers of the system in Figure 1 below) as a Blockchain-as-a-Service (BaaS) infrastructure and this decreases capital expenditure for integrating with the system and thus increases chances of the system adoption from their part. In general, Web / Cloud Platforms database administrators could be regarded either as honest-but-curious which means that they will follow the agreed with the Data Owner protocol and return the results of the computations done on its side, however they may look at the data they processes, or as malicious, which means that they may not follow the agreed protocol and/or may not return the results of the computations. In our setting, the *malicious* Web / Cloud database administration security threat model is assumed, in which the administrator may see and alter data but not deny access to them and measures to tackle with this threat are described in Section 4 below.

C. *Medical research centers*

Medical research centers who want access to the Healthcare Data stored on Clinical Platforms for research purposes. E.g. they might want to use a common pool of data from which to define demographic cohorts or enhance precision medicine practices. They are not by default trusted

so there is a need for an off-chain verification of their Identity before being accepted as nodes of the Blockchain network.

D. Validators

Validators are a subset of the Blockchain network nodes which assemble new blocks of valid transactions. All verified Entities participating as nodes of the network will/can be Validators.

E. Blockchain model assumed

The Blockchain model assumed is a *consortium* blockchain in which identities of medical research centers that participate as nodes of the network are assumed to be verified *off-chain*. Once the medical research centers are verified and allowed to be network nodes, they are considered to be *trusted* by all the other peers of the network.

IV. System Architecture

In this section, architectural layers, components and interactions among the components of our proposed system are presented.

A. Layer1: Web / Cloud Platforms

In this layer, there are multiple Platforms either Web hosted or provided as Cloud Services which store Patient healthcare Data on their own local databases. Such an example Web Clinical Platform is myAirCoach[3]. They can export Patient data in a format appropriate for exchange over Restful Web Services. The data are being hashed before being transferred so as to avoid data leakage during their transfer between layers 1 and 2. The data transfer cost between layers 1 and 2 is also typically decreased due to the hashing.

B. Layer2: Cloud middleware

This is the cloud middleware, which connects multiple VMs that are set up in order to ensure that there is no single point of failure as opposed to a centralized setting in which one dedicated server hosts the middleware infrastructure. This component connects the Web / Cloud Clinical Platforms located at layer 1 with the consortium Blockchain network located at layer 3. It interacts with layer 1 by receiving the data via RESTful API over HTTP(s) and then stores them to the Blockchain by contacting with the dedicated smart contracts at layer 3 using an appropriate API to interact with these smart contracts.

C. Layer3: Blockchain network

This is the consortium Blockchain network. The smart contracts that administer the data sharing and permission management are deployed in this Blockchain network. Communication between Layers 2 and 3 is achieved via an appropriate API.

A key feature in this architectural design is that Layers 2 and 3 are exposed to the Web / Cloud clinical platforms located at Layer 1 as a Blockchain-as-a-Service (BaaS). This means that Platforms located at Layer 1 do not need to be Blockchain nodes in order to send the data; they only need to export the data in an appropriate format that can be consumed by the Web services exposed at Layer 2. This increases adoption of the system from multiple / heterogeneous Platforms as it is not bound to their local infrastructure (e.g. which DBMS they use).

The system architecture representing the above described layers together with the components in each layer and interactions among them is shown in Figure 1 below.

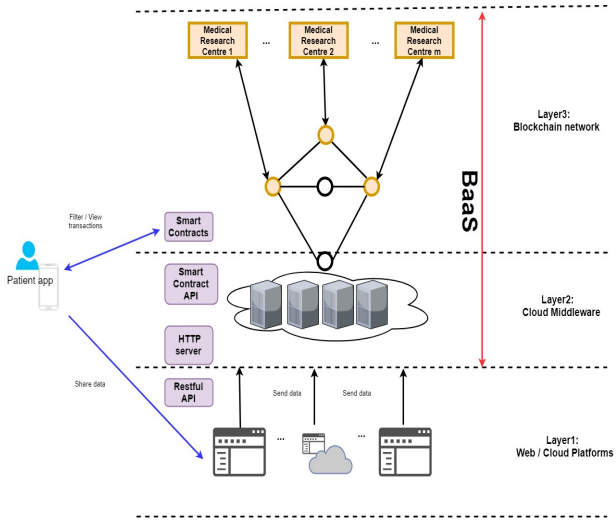

Figure 1. System Architecture

V. Smart Contracts

In this section, the Smart Contracts used by our system are analyzed.

A. Registry Contract (RC)

This smart contract acts as the Registry of all the Users of the System. Users can be separated in two different categories, i.e. (i) medical research centers and and (ii) Patients. RC contains a mapping between all system Users uniquely identifying field with a unique smart contract address that is called Patient Data Contract (PDC) and corresponds to each Patient Data. This uniquely identifying field should have the following properties (i) be unique per Patient, (ii) not able to reveal their identities in a direct or an indirect manner. Note that in the case that the User is not a Patient the address of the PDC can be a null or empty field that does not point to an existing PDC deployed within the Blockchain.

B. Patient Data Contract (PDC)

This smart contract is unique to each Patient and contains the hashed Patient healthcare Data along with a URL pointing to the Patient healthcare Data in the Web / Cloud Clinical Platform local database. The hashed copy of the data is used in order for the Entities that want to access the Data (medical research centers) to be able to verify the data integrity to tackle with the *malicious* database administrator

[3] http://myaircoach.eu/myaircoach/

threat model that has been assumed as already explained in section 3 above.

C. Permissions Contract (PC)

This smart contract administers the Permission management of Patient Data. In particular, it contains a mapping between the Patient Data Contract address, the data requesting Entity (medical research center) uniquely identifying field with a field called 'Permissions Status' which contains the Patient approval to access their data.

The Smart Contracts of the system along with the data they contain and high-level relationships among them are depicted in Figure 2 below.

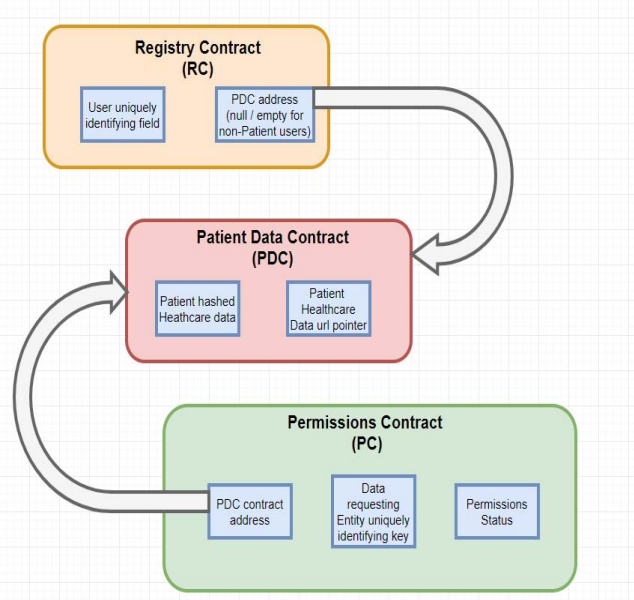

Figure 2. Smart Contracts along with their data and relationships between them

VI. Use Case Scenarios

In this section, the following use case scenarios of our proposed system are presented, (i) User Registration, (ii) Patient Data Sharing, and (iii) Medical Research centers Request Permissions to access Patient Data.

A. User Registration

Users[4] using a dedicated application UI register to the RC by entering a uniquely identifying field generated for them. In case that the User to be registered is a Patient (option ticked within their UI), then a new PDC is created within the RC and the PDC address is written as a reference back to the RC.

B. Patient Data Sharing

For pseudonymity issues, Patients can connect with the Web / Cloud Clinical Platform via which they will share their data in two ways (i) with an account on the Platform, (ii) without any account.

[4] In our setting, User can be either a Patient or a medical research centre.

In the case of sharing data with an account on the Clinical Platform, the uniquely identifying field of the Patient should be given by the user via the Patient application UI separately and not to be stored on the database, so that there is no matching in the Clinical Platform database between the field and personal data like name, username, emails, etc. Since the Patient has an account on the Web Platform in this case, some (encrypted) data on the URL stored on the Blockchain can enable data requesting entities to retrieve data back on the Web / Cloud Platform.

In the case of sharing data *without* an account on the Clinical Platform, the user will upload and store data on the Clinical Platform anonymously (e.g. with two factor authentication, email verification, JWT temporary token login, etc.). The data will be stored there along with the uniquely identifying field of the Patient. In this case, there is no problem to store the field, since the Patient does not store personal data on the Clinical Platform and thus the uniquely identifying field and personal data cannot be matched; patient pseudonymity i.e. *who* shares data via the Platform is thus preserved. The Clinical Platform is used for exporting data in correct format for communication with the BaaS (e.g. RESTful web service with JSON message format exchange that sends POST HTTP(s) requests to the HTTP server located at the layer 2) and also for sharing the Blockchain stored URL for the medical researchers to access the data. Since the Patient has not an account on the Web Platform in this case, the uniquely identifying field is used to match the user with their data.

The workflow of this use case scenario is the following:

1. Web / Cloud Platform hashes locally Patient Data and sends them along with the uniquely identifying field of the Patient to the cloud HTTP server which calls the smart contract API.
2. The smart contract API searches the RC for the uniquely identifying field of the Patient. If not found, adds a new record to the RC for this Patient, then creates a new PDC contract and stores the data of the Patient into this newly created PDC. If already found, it locates the PDC and stores the data in the related fields of the PDC (this is actually a data update).

C. Request Permisions

A key design concept of our proposed system is that the medical Research centers should not know whose Patient data they access. This ensures Patient pseudonymity, i.e. data requesting entities know what the healthcare data are, but they do not know whose data they are. The only prerequisite is that the Patient has been registered with the system which serves as an implicit consent that they allow their healthcare data to be accessed by requesting Entities in a pseudonymous way.

The workflow of this use case scenario is the following:

1. The data requesting Entity, which in our setting can be a medical research center, calls a dedicated

function within the RC and selects *randomly* a Patient. The reason for random selection is that there should be no difference in the times each Patient is selected to be notified to provide access to their data, consider e.g. the scenario in which Patients are *sequentially* selected from the RC to approve access to their data. In such a case, the higher a Patient is in the Registry, the more times they will be notified to provide access to data requesting entities. Note that to implement this functionality, a dedicated function within the RC should exist which in order to distinguish between Patients and other Users e.g. by checking the PDC address value; in the case of a non-Patient User this filed should have an empty or null value.

2. A temporary permission contract (PC) between the Entity and the Patient with Permission Status 'Pending' is created.
3. A notification is sent to the Patient via the dedicated Patient application UI to either approve/reject giving permissions to the requesting Entity.
4. If not accepted, the temporary PC is deleted (destructed). A related message is sent to the Entity.
5. If accepted, the Status of the Permission Contract is changed to 'Approved'.
6. When accepted, a related message is sent to the Entity and the Entity can use (i) the URL to access online the healthcare data, and (ii) the hashed data to check the data integrity by hashing the data provided by the URL resource and matching them with the hashed data obtained from the Blockchain. The access to the data may be time restricted and fine permissions can have also been granted (i.e. give access to only portions of the data).
7. A transaction is logged to the Blockchain about this access for auditability/accountability reasons.

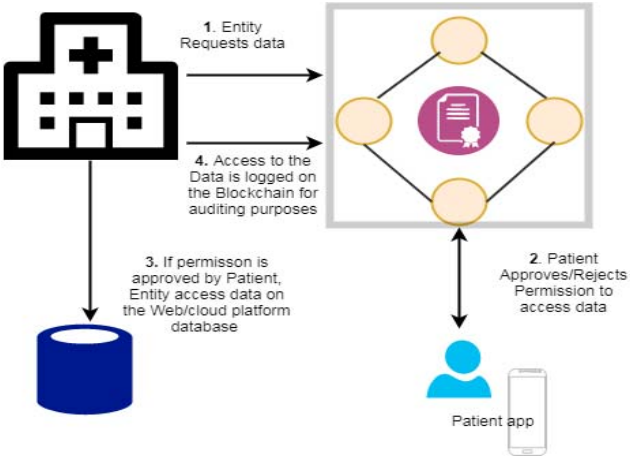

Figure 3. High-level description of 'Request Permissions' use case scenario

VII. System Value

A. Data Integrity

The data requesting entity, after obtaining permission by the Patient to access their data can see if the data downloaded from the Web / Cloud clinical Platform (after being hashed), match the hashed data stored on the PDC Smart Contract. This is significant if we assume a *malicious* administrator security model in the web/cloud clinical platform database side.

B. Patient Pseudonymity

Patient Pseudonymity is ensured since only the uniquely identifying field of the Patient is stored on the RC Smart Contract and *none of their personal data are stored* there. Patients are selected in a *random* order so there is no chance to track patients according to their registration order (no information leakage).

C. Workflow automation

Each time an Entity needs to access the Patient healthcare data, the procedure is done automatically via smart contracts. The Patient needs just to approve the Permission (once) via a notification and then a PC contract is created. Each subsequent time, the smart contract automatically checks for permissions and gives access. This way the workflow is automated; no paperwork or personal contact is needed and the service is available on a 24/7 basis.

D. Auditing and accountability

Via the dedicated Patients application UI, Patients are allowed to filter the transactions with their uniquely identifying field and view *who* accessed their Data, *what* Data they accessed, and *when*. Thus, auditability as regards the actions performed on the system is ensured. Given that only (off-chain verified) trusted Entities participate into the Blockchain network also acting as Validators, *immutability* property of Blockchain is ensured from which fact *accountability* is also derived.

VIII. Conclusions

In this paper, we presented the design of a system that enables healthcare data sharing and permission management in a secure, private and auditable way by leveraging unique properties of the Blockchain technology. Security of the system is also enhanced by enabling the requesting data entities to check with the assistance of the Blockchain infrastructure the *integrity* of the data that they access. The proposed system can be leveraged in frames of the healthcare data exchange and interoperability between National Contact Points (NCPs) in frames of the KONFIDO project.[5]

We showed how the proposed system has added value in frames of (i) Patient Data Integrity, (ii) Patient pseudonymity, (iii) Workflow automation, and (iv) Auditing and accountability.

Taking the above into account, it is evident that the proposed system, if adopted by the requesting data involved

[5] http://www.konfido-project.eu/konfido/

entities (i.e. the medical research centers), will clearly have an impact on medical research innovation boosting.

ACKNOWLEDGMENT

Authors acknowledge support from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643607 (myAircoach) and No 727528 (KONFIDO).

REFERENCES

[1] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." (2008): 28.

[2] Islam, SM Riazul, et al. "The internet of things for health care: a comprehensive survey." IEEE Access 3 (2015): 678-708.

[3] Ekblaw, Ariel, et al. "A Case Study for Blockchain in Healthcare:"MedRec" prototype for electronic health records and medical research data." Proceedings of IEEE Open & Big Data Conference. 2016.

[4] Peterson, Kevin, et al. "A Blockchain-Based Approach to Health Information Exchange Networks." (2016).

[5] Zhang, Jie, Nian Xue, and Xin Huang. "A Secure System For Pervasive Social Network-Based Healthcare." IEEE Access 4 (2016): 9239-9250. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Nugent T, Upton D and Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts [version 1; referees: 3 approved]. F1000Research 2016, 5:2541 (doi: 10.12688/f1000research.9756.1)

[7] Shae, Zonyin, and Jeffrey JP Tsai. "On the Design of a Blockchain Platform for Clinical Trial and Precision Medicine." Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on. IEEE, 2017.

[8] Benchoufi M, Porcher R and Ravaud P. Blockchain protocols in clinical trials: Transparency and traceability of consent [version 1; referees: 1 approved, 1 not approved]. F1000Research 2017, 6:66 (doi: 10.12688/f1000research.10531.1)

[9] Zhang, Yin, et al. "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data." IEEE Systems Journal 11.1 (2017): 88-95.