Blockchain EHR System: Integrity, Privacy, Scalability

Anwar Walid and Wenpu Chen
Nokia Bell Labs, Columbia University
July 31, 2016

ABSTRACT

In this paper we present our vision for the application of blockchain to e-Health. We propose an architecture built on the blockchain (or permissioned ledgers?) to enable health systems to efficiently manage and secure digital assets. The system enables patients and health provider to communicate healthcare data with integrity and privacy, and enables health provider to store healthcare records such as medical bills and client-physician communications to claims and disputes. The cryptographic security enhances the security of such records, while the immutable, irrevocable nature of transactions is intended to make claims processing more efficient and simplify dispute resolution. Health records secured via a blockchain could make it easier for patients to share records with multiple providers, while retaining control over those records. The new solution enables the vision of Interoperable Health IT Infrastructure where individual and community health is improved through seamless health information flow and availability to the right people, at the right place and at the right time.

TABLE OF CONTENTS

1. INTRODUCTION AND MOTIVATIONS

Providing efficient and secure access to Electronic Health Records (EHR) is a vital next step in activating patients in their care and improving the health system. Current systems remain largely closed systems preventing interoperability, portability and collaboration. Indeed, the vision of Interoperable Health IT Infrastructure is to enables

improved individual and community health through seamless health information flow and availability to the right people, at the right place and at the right time.

Current systems, even with their siloed design, still have security and privacy vulnerabilities. Migrating to more open and interoperable designs pose even greater challenges. There is a real concern about both people's and entities' access levels to patients' EHRs. A patient's EHR might be fragmented and accessible from several sites (by visiting different doctors' offices, hospitals, providers, etc.). Security defects in some of these systems could cause the disclosure of information to unauthorized persons or companies, and health data therefore need protection against manipulations, unauthorized accesses and abuses, which includes taking into account privacy, trustworthiness, authentication, responsibility and availability issues. EHRs also have difficulties in maintaining data privacy, to the extent that administrative staff could access information without the patient's explicit consent. Additionally, as the new vision is of a learning health system where individuals are at the center of their care generating longitudinal health status through their wearables, large volumes of data need to be efficiently stored and retrieved.

In this paper we outline our proposal for EHR system based on the blockchain. The blockchain is well-suited to address the above challenges. It facilitates an architecture that is globally accessible and yet localized to the patient's needs. Each patient's medical history can be securely recorded as they journey through life, following them as needed. The system is more secure offering protection from massive breaches. With blockchain, the attack surface is a single patient, and a distributed ledger doesn't have the same kind of vulnerabilities to ransomware or social engineering attacks.

Although the data is distributed, access is global, so data analysis on a large sample of patients would be possible, thus improving population health. Of course, all access to patient data would be opt-in by design, indeed by platform, because the patient has the cryptographic last word on the matter. Furthermore, as healthcare slowly shifts focus from acute and centralized to chronic and distributed, collecting health data "telemetry" from distributed patients will be an invaluable capability well-suited to the blockchain

The blockchain: There are two key parts of Blockchain technology: (i) A distributed ledger system, which is an append-only database whose uploads verified cryptographically in a distributed and transparent manner, and (ii) a consensus protocol that allows participants to agree on appending data. It is the process in which a majority (or in some cases all) of network validators come to agreement on the state of a ledger. It is a set of rules and procedures that allows maintaining coherent set of facts between

multiple participating nodes. In the case of Bitcoin, the “longest chain” – the chain with the most proof-of-work – is considered to be the valid ledger.

The cryptographic security enhances the security of such records, while the immutable, irrevocable nature of transactions is intended to make claims processing more efficient and simplify dispute resolution. The value of the blockchain may be viewed in terms of enabling an overall highly reliable EHR system with minimal cost, where the cost can be viewed as:
Cost = Risk of failure * Cost of failure + (1- Risk of failure) * Normal cost
The blockchain technology can significantly decrease the Risk of Failure and thus the overall cost.

2. BLOCKCHAIN SOLUTION ALTERNATIVES

The design characteristics of blockchain may be tailored to meet the demands of particular system objectives. The blockchain ledger can be designed to be permissioned or permissionless. A permissioned system is one in which identity for users is whitelisted (or blacklisted); it is the common method of managing identity in traditional finance. In contrast, a permissionless system is one in which identity of participants is either pseudonymous or even anonymous. Bitcoin was originally designed with permissionless parameters although as of this writing many of the on-ramps and off-ramps for Bitcoin are increasingly permission-based [1].

Permissionless (public) ledgers like Bitcoin and Ethereum [2][3] use proof of work as their consensus protocol, the network is robust as long as no one gain control of over 51% voting power (in both cases, the computing power). On the other hand permissioned ledgers like Ripple [4] use different consensus protocol. less participants means larger potential of unbalanced computing power distribution, which means the POW is no longer a valid protocol. On the other hand, having pre-selected participants means we have stronger assumption - there exists trust in the network. The Ripple consensus protocol family is a protocol set that fits the demand of partially trusted community. It is a well-balanced solution compared with the traditional centralized system and the totally distributed Bitcoin blockchain. We view it as a good representation of our ‘partially trusted’ in real world.

There are a variety of trade-offs between permissioned and permissionless systems involving speed, cost reduction, censorship, reversibility and finality. And due to their gated approach, permissioned systems as a whole are capable of clearing and settling

assets faster and are cheaper to maintain than capital-intensive Permissionless systems.

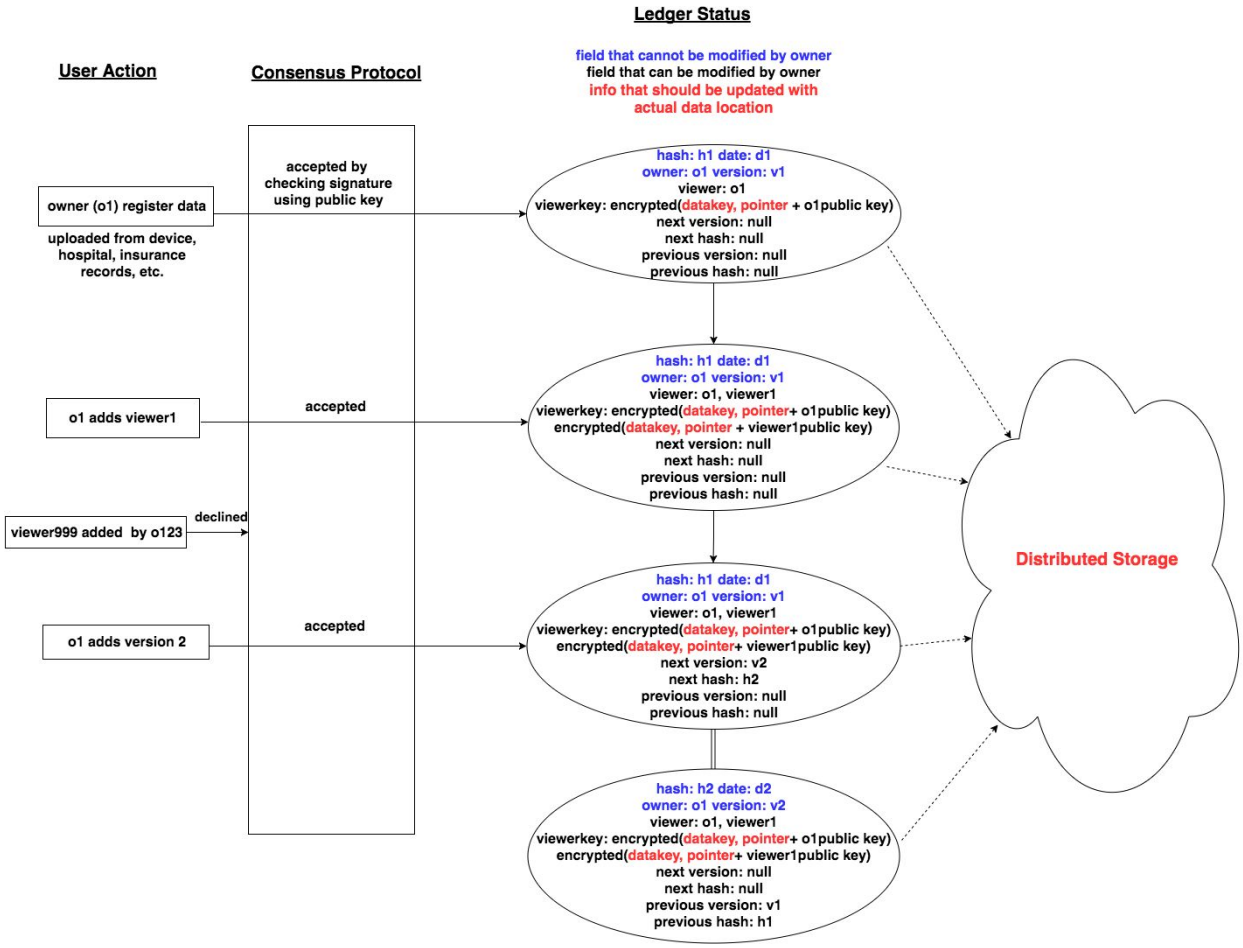
In our proposal, described next, we consider a permissioned ledger design.

3. PROPOSAL

3.1 Use case

To illustrate our proposal, we consider four types of users in the system. Doctors, Patients, Insurance companies and Legal parties. We build use cases for all of them to create a complete ehealth ecosystem. For doctors we have Hospital record registration, Patient health record integrity check, Read access under control. For patients: Device record registration, Personal data read access control, Hospital record integrity check. For insurance company: Patient record integrity check, Hospital record integrity check, Read access under control. For legal parties: Ledger data can be used as an evidence for clients.

3.2 Implementation structure



The blockchain system content flow acts like the traditional centralized database. Users can have multiple database actions (see left), and the public ledger will act corresponding to those actions (see right). There are two main difference: 1 The actually content is saved in distributed storage, somewhere separated from the ledger. The ledger only save name, verification and pointer (the red part). 2 Instead of using a firewall to protect both writing and reading access the centralized database, we use the consensus protocol (see middle) to protect the writing access of ledger. And to protect the integrity and avoid damage from internal database, the ledger reading access is open to all nodes.

To conclude, our design contain the Ledger function (see section 3.3) and Consensus protocol (see section 3.4)

3.3 Ledger function

The purpose of ledger is to create a public file that everyone have the ability to verify item without help from third-party. It is one of the two blockchain key parts. The following are our ledger functions designed for the ehealth public ledger system.

3.3.1 User registration

- Each registered user will be given a predetermined key pair
- Each registered user can create his own key pair if he can provide valid signature.
- Any device user, health service provider, insurance company registered at the company will be provided with an unused key-pair

3.3.2 Create and register data

- Data1 is created by owner1
- Data1 is uploaded to centralized/distributed third-party database, encrypted.
- Owner1 upload the hash, date, version info, valid viewer info, encrypted pointer, encrypted key for each viewer, and sign this registration using his private key
- All other nodes check the signature and reach a consensus to accept the update

3.3.3 Data version update

- Owner1 has a new version for the same data called Data2
- Owner1 update the previous registration by editing the ‘next version’ field: null -> Data2
- Owner1 update the ‘next hash’ field with Data2’s hash
- Owner1 create a new registration for Data2
- Owner1 sign using private key and other nodes will check the signature before acceptance
- Just the same as the list structure same for further update Data3, Data4, Data5 …

3.3.4 Add/delete viewer

- Owner1 add viewer1 under the field ‘viewer’
- Owner1 create a new encrypted pointer and key for viewer1 and add it under field ‘viewer key’
- Sign the update using private key and let other node check
- When delete a viewer just delete the corresponding item under the above two fields

3.3.5 Read the data + integrity check

- Owner1 agree viewer1 to read Data1

- Owner1 update the registration and encrypted pointer/key in the blockchain
- Viewer1 went to the blockchain and gain the pointer/key
- Viewer1 use his own private key to read the pointer/key
- Viewer1 download the data following the pointer
- Viewer1 read the data using the key provided by Owner1
- Viewer check the hash of d1 with the one registered in blockchain

3.3.6 Protection from third party database damage

- Data1 is damaged when stored in third party
- When viewer1 retrieve the data he calculate the hash
- Compare with blockchain
- Not equal -> discard and notify owner1

3.3.7 Protection from invalid upload

- OwnerI want to register data2 under Owner1's name
- OwnerI cannot provide the current signature
- Other nodes disagree on acceptance
- Nothing registered

3.3.8 Existence proof

- Owner1(i.e. the health provider) destroyed Data1 and want no one find it
- Viewer1(i.e. the patient) previously downloaded the data and have a local copy
- There is a conflict between owner1 and viewer1
- Viewer1 show the lawyer the local copy of data1
- Lawyer calculated the hash of data1
- Compare with blockchain -> same
- Since delete/modify hash and date is prohibited, owner1 cannot deny

3.3.9 Protection from unwanted viewer

- Viewer999 want to read data1
- Viewer999 is not registered by owner1
- All the link / pointer under data1 is encrypted
- Viewer999 cannot have the access
- Data1 itself is encrypted in database -> admin of the database cannot read it

3.4 Consensus protocol

The consensus protocol is equally important as ledger function in any blockchain design. Depending on the user distribution and property of community, we should choose consensus protocol with care. In our approach, we choose the adjusted Ripple

consensus protocol family. Unlike the ripple who treat all node in community uniformly, we initially set a voting power distribution among different interest groups.

Whenever a new participant join the network, it will be given a voting label. voting nodes sharing the same label will share the preset voting power. The labels are 1 Data generators (patients etc) 2 Data receivers (healthcare providers etc) 3 Data monitors (trusted parties etc).
This will prevent any potential interest group become overwhelming.

Compare with the permissionless protocol like proof-of-work, our approach will have a lower cost on computing power, faster update time and prevent from the 51% attack. On the other hand, our approach also have lower failure risk compare with traditional centralized system

3.5 Privacy and Security
Privacy is of the most important feature we care about. Since the ledger only contains Hash, we no privacy concerns.
Security issues can come from both ledger part and consensus protocol part. For the ledger function, see section 3.3.6 to 3.3.9.

For the consensus protocol, as long as honest parties are the majority for all the three voting parties, we no security problem. If there is one voting party fail, the other two parties will still keep the consensus working. If two of the three voting parties fail, the network will fail and integrity history will not be verified afterwards.

There are two solution to deal with a network fail:
1 Keep the key hash root updated and published in public (twitter, newspaper, official pages etc) since the device will can keep the related merkle tree to the root, the root is sufficient to prove the integrity of certain piece of data. one just to prove the hash is generated BEFORE the network failure. 2 After the network failure, one can retrieve the valid hash roots and form a new network starting from the last valid root. It will be valid until the next attack.

3.6 Performance and Cost
Major cost will be storage cost. For each node join the network, we will have to store the hash copy. For small devices one can choose to keep the hash root and hash they personally related and discard the rest to save storage space. For large facilities they have to keep the complete hash history copy.

3.7 Additional comments

Note that the system expected cost should be compared with that of a comparable centralized system (including connection, admin integrity, location, policy, maintenance cost, etc). However, given the failure cost can be large in interoperable health systems, we have enough reason to implement the blockchain solution to minimize integrity, privacy and access risks.

4. DISCUSSIONS AND FUTURE DIRECTIONS

One possible future direction is to build additional Ledger functions that suits the need of EHR systems. Additional research and investigation also needed on the consensus protocol and its parameter adjustments to enable more robust system. Another area of research is on cryptography for more secure ledger.

On the other hand, the estimation of actual cost of both the blockchain solution and the centralized solution is still an issue to be investigated and the risks evaluated.

5. REFERENCES

[1] “Consensus-as-a-service: a brief report on the emergence of permissioned, distributed ledger systems,” Tim Swanson, April 6, 2015.
[2] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system[J]. 2008.
[3] Ethereum White Paper - A Next-Generation Smart Contract and Decentralized Application Platform https://github.com/ethereum/wiki/wiki/White-Paper
[4] Schwartz D, Youngs N, Britto A. The Ripple protocol consensus algorithm[J]. Ripple Labs Inc White Paper, 2014: 5.Bell