ENABLING THE REALIZATION OF NATIONWIDE INTEROPERABILITY AND PATIENT CENTERED RESEARCH THROUGH A BLOCKCHAIN SOLUTION

Mathew E. Rose, MB BCH BAO, MS
(mathewrose@rcsi.ie)
August 8, 2016

Abstract

The Nationwide Interoperability Roadmap, Precision Medicine Initiative, Patient Centered Outcomes Research, and other national healthcare initiatives identify the need for interoperable electronic healthcare records to provide better patient care, decrease healthcare costs, and enable large scale data analysis to improve healthcare standard of practice, advance medical technology, and guide policy making. We propose the use of a permissioned distributed blockchain solution to address these pitfalls in current systems nationwide. The solution allows sharable data to be captured from healthcare providers, other stakeholders, and IoT devices, while enabling large scale de-identified data analysis. Through the use of blockchain, a distributed network, and hash cryptography we reduce the risk of ransomware attacks, offer immutable data storage, increased data security, prevention of data theft, and enable healthcare stakeholders to meet national initiative objectives.

1. Introduction

The adoption of electronics and the internet in today's world has made a huge impact on society and the individual. It has done so for two key reasons, it aims to improve quality of life and provide better access to information. Healthcare is no different in this respect. In 2009 the Health Information Technology for Economic and Clinical Health Act (HITECH), a national initiative to adopt electronic health records (EHR) was put into place, because the value of having broader access to medical records allowed for improved patient care, research, and administrative tasks. By 2015, 96% of non-federal acute care hospitals across the US had possessed certified health IT.[1] Among these institutions, the number sending electronic clinical summaries had increased from 41% in 2008 to 82% in 2015.[2] Yet the sending of information is only one of four components that make up interoperability among medical institutions. The other three consist of finding, receiving, and integrating medical records into patient care. Further complicating the nation's mission to establish interoperability of health information is new technology. With the adoption of activity tracking devices and increased monitoring of individuals activities through the 'internet of things' (IoT), there is even more information. All this information needs to be captured and meet the 4 domains of interoperability, which can be used by clinical decision makers to help improve patient care and health outcomes.

Unfortunately, as with anything good there is always something that challenges its positive nature. The growing use of electronic medical records and IoT is now endangering individual privacy, security, and even patient safety. Patient identity theft for fraudulent acquisition of

[1] http://dashboard.healthit.gov/quickstats/pages/certified-electronic-health-record-technology-in-hospitals.php
[2] http://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-interoperability-2015.php

controlled medications is a growing problem and the electronic availability of health records is giving notorious individuals better access.[3] Hackers have evolved from just stealing personal information to sell and are now holding entire hospital networks hostage.[4] As a result, it is critical that the healthcare industry among others develop methods to enable the electronic capturing and sharing of data while ensuring personal and institutional security. We believe blockchain technology can offer such a solution.

The purpose of this paper is to demonstrate a method of using blockchain technology and cryptography in healthcare to enable the realization of the nationwide interoperability, improved patient care, reduced healthcare costs, and advanced patient centered research.

2. What is blockchain?

Although blockchain can be related to Bitcoin,[5] it does not have to be and therefore a general convention has been established within blockchain technology community: Blockchain with a capitalized "B" refers to the database/ledger that forms the backbone of Bitcoin. Whereas blockchain with a lower case "b" refers to all other blockchains not associated with Bitcoin.[6]

blockchain is simply a relatively new type of database that can be used to store information in a distributed or decentralized manner. Multiple structures can be used to establish a blockchain system design, but certain basic components apply to all blockchain technology.

Figure 1- Centralized vs Decentralized vs Distributed Networks

blockchain at its core consists of data stored in a building block like manner that is replicated and placed in multiple decentralized/distributed locations called nodes. Each block is unique and will only fit onto the block that came before it and into the block that comes after it. Any change to the data in a particular block would change its structure and therefore in turn cause it not to fit in place. Additionally, if data from one node or several nodes is deleted, exact copies from still existing nodes will replace the missing data. This means that the data in every block is always the same, every block is placed in the same order, and any data deleted will always be recovered

[3] http://www.healthcareinfosecurity.com/3-stolen-health-databases-reportedly-for-sale-on-dark-web-a-9227
[4] http://www.healthcarebusinesstech.com/hackers-hospital-data/
[5] Satoshi Nakamoto, Bitcoin a Peer-to-Peer Electronic Cash System (https://bitcoin.org/bitcoin.pdf)
[6] Luis Carranza, An introduction of the current state of play, blockchain Conference Dublin Ireland 2016

giving the technology an immutability characteristic - data cannot be changed once placed on the blockchain.

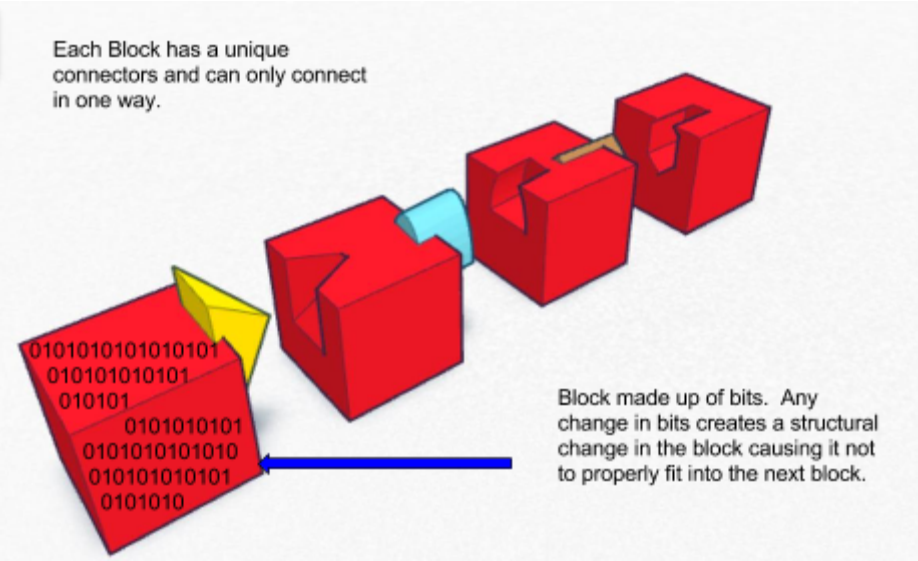


Figure 2- blockchain core

Given that the technology is stored identically in a decentralized/distributed way, no one individual, group, or organization can absolutely control the block without an agreed consensus. In other words a change would have to be made simultaneously in more than 51% of the nodes to create a change in the established block. This means that the data is resistant to being held hostage from a malicious attack and data immutability is further sustained.

2.1 Types of blockchain

blockchain technology can be subcategorized into attributes that further define the design on top of its core. A blockchain can either be Permissioned or Permissionless, which can be further subdivided into Public or Private. Permissioned means that a central party or consortium gives an individual permission through some type of "Know Your Customer" (KYC) procedure allowing them to access and place information on the blockchain.[7] Therefore, it can only be a designed as private type of blockchain. A permissionless blockchain is as the word describes open to anyone who wants to place data on the blockchain and therefore can be either public or private.

2.2 Hash cryptography- A 2nd component of blockchain security

blockchain refers to the method of storage, while cryptography addresses the format of the data placed in storage. Cryptography uses mathematical algorithms to encrypt and decrypt data with

[7] Tim Swanson, Consensus-as-a-service: a brief report on the emergence of permissioned, distributed ledger systems

the intent of allowing only those for whom the data is intended to read it. In general these algorithms take data input, break it down into its binary components (i.e. 0's and 1's) and combine it with a unique key to convert it to output that is completely illegible, random (according to laws of statistics), but reproducible. Each output created will have a unique length depending on the key used and the size of the data input. The specific bits used for the input creates a unique fingerprint in the output. The slightest change of a single bit in the input will completely change the output and as a result change the fingerprint.[8] When combining the fingerprint with a public key that is sent to the receiver of the data, a digital signature is created. The digital signature allows the receiver to verify the data was sent by the party they believe created the data and not someone else. This encrypted data can then be trusted by the receiver and decrypted using a specific key to convert the output back into the original data. Such functionality makes it nearly impossible for a malicious attacker to decrypt and/or change data being sent between parties.

There are numerous cryptographic algorithms in existence and each has their own level of strength in preventing a non-intended party from being able to view the unencrypted data. Current focus is mostly on hash algorithms with the main one being in the SHA2 family.[9] The majority of blockchain technology companies today use a subtype of this family called SHA256 to encrypt data. As computer technology advances, the security of these specific algorithms weakens. It is for this reason that the SHA3 family of algorithms was created and other cryptographic families will be created in the future. Therefore, even though the blockchain solution offers increased security now, it will need to have the capability to undergo updates in the future to further increase cryptographic algorithm security.

3. A blockchain solution to enable nationwide interoperability of health information technology

An ideal blockchain solution for the use of maintaining healthcare data needs to meet the following criteria:

- Be immediately accessible by all hospitals, treating physicians, emergency personnel and other stakeholders.
- Allow limits to be placed on the data so that administrators, government agencies, insurance companies and other agencies only have access to the information needed; not the entire records.
- Data is collected and stored so that it can be stripped of any personally identifiable information (PII), but also retrieved in association with identifiable information.
- Data storage systems need the ability to be HIPAA compliant.
- Data must be secure from malicious attacks and held hostage.
- Data needs to be unmodifiable once recorded in order to prevent illegal tampering.
- Implementation of the blockchain solution must occur smoothly with little impact on the individual inputting data.

---

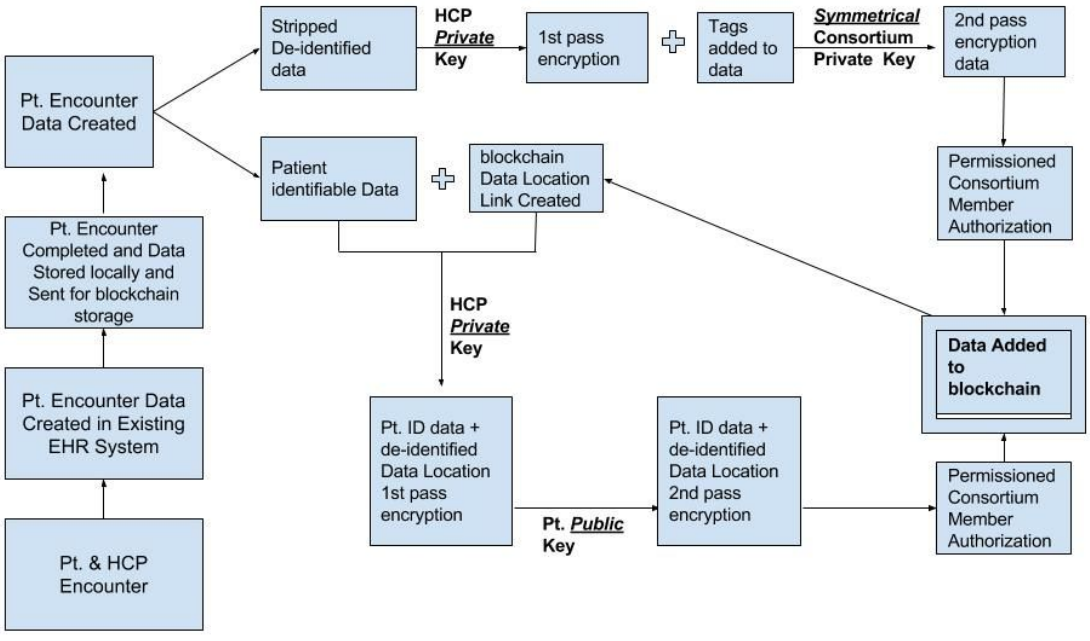
[8] Andreas M. Antonopoulos: "Consensus Algorithms, Blockchain Technology and Bitcoin" University College London Lecture (https://www.youtube.com/watch?v=sE7998qfjgk)

[9] NIST Special Publication 800-197 Revision 1: Recommendation for applications using approved hash algorithms

3.1 Proposed blockchain Solution

We propose the use of a permissioned distributed blockchain solution that uses two layers of key pairs (private key and public key) and a symmetrical consortium key for data encryption. A consortium distributed storage network will be established consisting of Hospitals and other approved stakeholders throughout the United States. Each member organization will undergo an extensive background check to confirm their validity then permissioned access to the blockchain by a consortium committee and assigned a public and private key. The public key is sent to all consortium members for data decryption and the private key is kept confidential within the organization. Every patient will also be given a private and public key, which are tied to biometric security measures (i.e. fingerprint scan, corneal scan, voice/facial recognition) to ensure identity of patient and provide for information access in the event of an emergency. The patient's biometric private key will give hospitals/other organizations permission to access health records for a specified period of time required to ensure their health encounter can be properly documented and used by the treating health provider. The patient's public key will allow healthcare providers to add data to the blockchain in the patient's absence. As a secondary measure, patient's will be asked to appoint an individual (Friend, family member or power of attorney) to act as a provider of access, whose biometric scan will be associated with the patient's private key in the event they are not able to provide access themselves Figure 3 and 4 demonstrate the flow of how data is stored and retrieved from the blockchain respectively.

Figure 3- Schematic for Data Addition to a blockchain



Schematic for Data Addition to a blockchain

Key
Pt. = Patient
HCP= Healthcare Provider/Healthcare organization
EHR= Electronic Healthcare Records

3.2 What can be stored on the blockchain?

Any data generated involving health information can be stored on the blockchain. Such data could be, but is not limited to: medical records, surgical records, radiology, lab investigations, medical device data, administrative records, billing records, administrative records, population studies, and recordings generated from IoT devices. Data will first undergo a process of stripping personally identifiable information leading to the generation of 2 types of data:

1. Data that is completely stripped of personally identifiable information (This is the data that can be accessed by approved organizations for large scale data analysis and organizational research).
2. Data that could not be stripped of personally identifiable information (This information only available during patient encounters).

Note: All data stored should include tag and metadata information (Billing, Insurance, GP, Surgical Records, etc) to enable better data filtering/analysis, reduced information redundancy, and increased data retrieval speeds. While a timestamp is automatically made based on data position in the blockchain, the data should also include an attached timestamp prior to encryption as it adds layer of verification of authenticity for future audits comparing timestamp.

3.3 How should the data in a blockchain be stored in the nodes?

A key feature of the blockchain is its redundancy on multiple nodes. The redundancy can be performed in 2 ways:

1.Each node has a full intact copy of the blockchain, which will require all nodes to have large data storage capabilities and potential exclude small healthcare providers.

2. Application of the storage method developed by Storj.io applied to the permissioned blockchain proposed.[10] In this method data is cut into shards, a specific sized piece of a file (1MB or 10MB, etc), which are stored separately on multiple nodes in the network. If one node goes down, then the information on that node will be rebuilt in another node. This improves speed of transmission for data retrieval. Additionally, since the entire blockchain is not stored on one node, security is increased and its susceptibility to being hacked is further reduced. Lastly, this second method of storage means that any computer in a health organization with available hard drive space can act as a node. Thus, healthcare providers of any size can participate in the consortium. Costs associated with upgrading storage hardware is significantly reduced for anyone individual node.

3.4 Does it matter if different software is used to obtain the data?

The software/method used to collect patient data does not matter, unless the software does not meet current standards for Health IT. The process of encrypting and storing data on the blockchain focuses on breaking down primary data to its basic bit structure. These bits are then stored on a similar backbone, the blockchain, eliminating the need for EHR systems to access

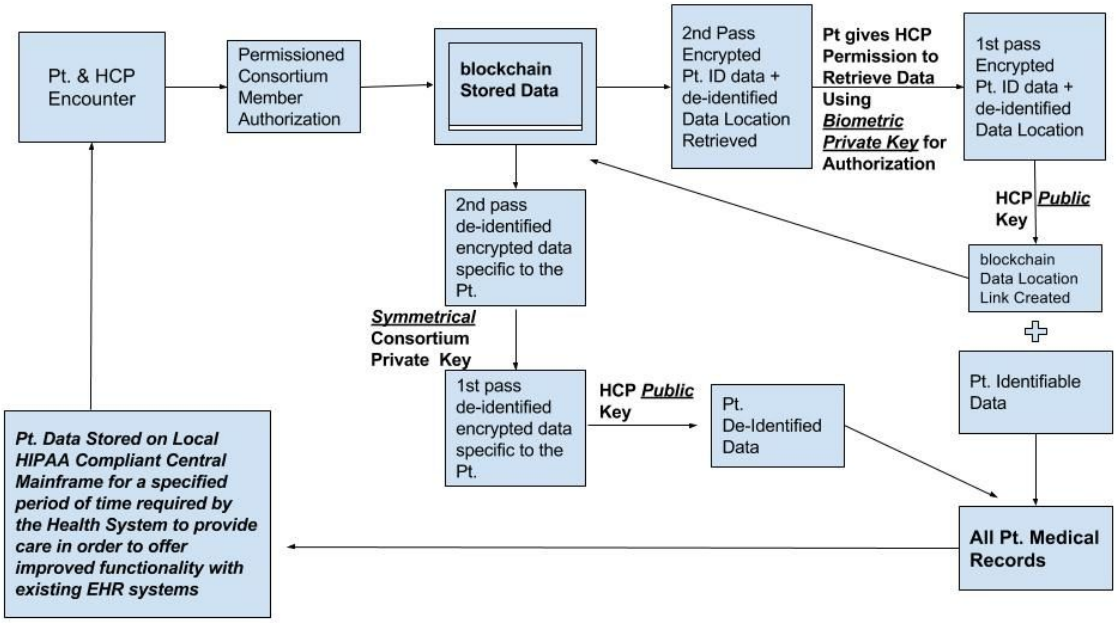
[10] Shawn Wilkinson, Storj A Peer-to-Peer Cloud Storage Network (https://storj.io/storj.pdf)

one another's databases. The implementation of integrating the Healthcare Enterprise (IHE) core strategy with Digital Imaging and Communications in Medicine (DICOM) and Health Level 7 (HL7) standards enables the interoperability among most of these existing EHR Systems.[11] In the event an EHR system used by a healthcare provider does not comply with these standards, the differences may create a challenge for data integration into each other's EHR software. This challenge can be addressed by the development of a translator that converts the code structure of the non-compliant EHR system into an agreed upon standards used in the blockchain. Such a step would smooth the transference of data between all EHR systems.

3.5 How can the de-identified data be obtained for patient encounters?

All de-identified data is stored on the blockchain without a way to trace it back to identifiable patient data. In order to enable de-identified data to be re-connected to its identifiable properties, we propose that links for storage locations be created as de-identified data is added to the blockchain. De-identified data will be encrypted and placed on the blockchain using a organizational specific private key. Once placed, the data has a specific location on the blockchain for which a link identifying its specific location will be sent back to the system creating the data and attach the link to patient Identifiable information prior to its encryption. Figure 3 and 4 show how these links will be incorporated into the identifiable data.

Figure 4- Schematic for Data Retrieval from a blockchain



Schematic for Data Retrieval from a blockchain

Key
Pt. = Patient
HCP= Healthcare Provider/Healthcare organization
EHR= Electronic Healthcare Records

[11] http://www.himss.org/library/interoperability-standards/basics

3.6 Maintaining HIPAA compliance

To maintain HIPAA compliance each organizational member of the consortium should already have an existing HIPAA compliant Electronic Healthcare Record System in place. Implementing the blockchain solution will change the data storage structure on the back end of those existing systems. Therefore the consortium will generate a HIPAA-Compliant Storage plan in coordination with each independent member/Institution. The nature of the blockchain solution addresses each of the components required for this plan. Each node in the consortium holds an exact copy of the blockchain, which enables data to be retrieved from another node should any issue develop (for example, fire, vandalism, system failure, and natural disaster) within one node or multiple nodes. Such a storage plan will address A Data Backup Plan, A Disaster Recovery Plan, Emergency mode operation plan, Testing and revision procedures, and Applications and data criticality analysis.[12]

3.7 Security of blockchain solution

From 2014 to 2015 there was a 62-fold increase in hacking incidents targeting health information.[13] blockchain technology as described above prevents data from being successfully attacked, corrupted or held hostage. With periodic maintenance and upgrades the data can be kept secure indefinitely. The use of a public key to biometric key association will further secure the data and prevent data loss in the event someone loses their private key. Costs associated with these devices today is relatively nominal and therefore should not be a barrier to use.

4.Potential gaps in standards

blockchain technology is a relatively new technology not included in existing IHE, HL7, or DICOM standards, amongst other standards influencing Health IT. The proposed blockchain solution would not change the way data is collected or format type of image files used. It would change the way data is stored. Therefore, if this proposal is adopted by the Health IT community existing standards will need to be updated to include the technology as the storage method for health related data.

Another potential gap is the current design of IoT devices, such as activity monitors, phones, etc. Today's IoT devices create continuous monitoring of the individual and a potential invasion of privacy. As a result IoT device standards need to be established to help improve personal privacy and protection of personally identifiable information.

5. Value of blockchain to the healthcare system

blockchain technology offers significant value to the healthcare system in the United States. It acts as a nationwide interoperable database from which healthcare providers, insurance companies, government agencies, manufacturers, distributors, dispensers, and other stakeholders

---

[12] TechTarget E-Guide: Managing Data and Developing a HIPAA-Compliant Storage Plan, (http://searchhealthit.techtarget.com/tip/Tip-Developing-a-HIPAA-compliant-storage-plan)

[13] http://dashboard.healthit.gov/quickstats/pages/breaches-protected-health-information.php

can obtain both personally identifiable information and de-identified data. These stakeholders can retrieve de-identified data to complete organizational functions, patient research, create policy, improve patient outcomes, reduced healthcare costs, and numerous other functions for which such data would be needed. In the situations where personally identifiable information is needed, individuals can be contacted and have the opportunity to provide or deny consent to use the information. This separation of personally identifiable information and de-identified data provides a framework on which IoT devices can be developed to enable objective functionality while maintaining personal privacy. Lastly, through the integration of data obtained into patient care, redundant tests/procedures can be eliminated reducing costs in healthcare.

6. Conclusion

The implementation of the distributed blockchain permissioned solution proposed would enable interoperability among stakeholders, improved patient outcomes, enhance patient research, reduce healthcare costs through reductions in redundancy, increase security, and improve personal privacy among today's IoT. As such, the technology will further enable the objectives of the Nationwide Interoperability Roadmap[14], Patient Centered Outcome Research, Precision Medicine Initiative[15], delivery system reform, and other national healthcare delivery priorities to be met.

[14] https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf

[15] https://www.nih.gov/precision-medicine-initiative-cohort-program