Blockchain: solving the privacy and research availability tradeoff for EHR data

A new disruptive technology in health data management

Gábor Magyar
Blockchain Expert Zrt.
Budapest, Hungary
gabor@blex.io

***Abstract* — A blockchain powered Health information ecosystem can solve a frequently discussed problem of the lifelong recorded patient health data, which seriously could hurdle the privacy of the patients and the growing data hunger of the research and policy maker institutions. On one side the general availability of the data is vital in emergency situations and supports heavily the different research, population health management and development activities, on the other side using the same data can lead to serious social and ethical problems caused by malicious actors. Currently, the regulation of the privacy data varies all over the world, however underlying principles are always defensive and protective towards patient privacy against general availability. The protective principles cause a defensive, data hiding attitude of the health system developers to avoid breaching the overall law regulations. It makes the policy makers and different – primarily drug – developers to find ways to treat data such a way that lead to ethical and political debates. In our paper we introduce how the blockchain technology can help solving the problem of secure data storing and ensuring data availability at the same time. We use the basic principles of the American HIPAA regulation, which defines the public availability criteria of health data, however the different local regulations may differ significantly. Blockchain's decentralized, intermediary-free, cryptographically secured attributes offer a new way of storing patient data securely and at the same time publicly available in a regulated way, where a well-designed distributed peer-to-peer network incentivize the smooth operation of a full-featured EHR system**

Keywords— Health data storage, patient privacy, blockchain, EHR, cryptographically secured data storage, multisignature data access

I. WHAT IS BLOCKCHAIN?

There is no widely accepted formal definition what blockchain actually is. However, it is accepted, that blockchain is a cryptographically secured, immutable, write once, read many type data structure, consisting of blocks of records, where the blocks are linked together using an unmodifiable key referencing mechanism. Blockchain is also a distributed data structure across a business network where there is no central entity responsible for the governing and management the data recorded. Blockchain is working over a network of single, independent entities, which form together a peer-to-peer network. Also, values recorded on the blockchain are synchronized between the peers and a consensus mechanism provides the common sense and commonly accepted validity of data. So, the blockchain data structure consist of the following components: (1) secured list of blocks containing data records, (2) a peer-to-peer network containing identical exemplars of the blockchain data structure, (3) a consensus mechanism securing the synchronized growth of the blockchain, (4) a security mechanism providing the immutability of the data structure. All of the four properties are equally important and so it has important consequences.

The availability of the data in a blockchain can be permissioned or permissionless. In case of a permissioned blockchain the security, validity and integrity of the blockchain is usually ensured by a well-known legal entity. In case of permissionless blockchain the security and integrity of the blockchain is secured by the majority of the independent peers and no central authority can influence the execution of the consensus rules. Also any changes in the consensus rules requires the agreement of the majority of the peers.

A blockchain structure is commonly used to store any kind of value as a distributed ledger which tracks the history of the value. In order to perform secure transactions between two different parties on the blockchain there is no need a trusted third party as the consensus rules of the network ensures the honest behavior of the parties. Instead of intermediaries and central authorities the trust is enforced by the network itself. Therefore the blockchain technology can be called as the technology of the trust.

Any data or transaction recorded on the blockchain cannot be changed any more, this ability of the system is not a pure technical limitation but comes from the inherent properties of the technology. As in legacy technologies the read-only attributes could be easily modified by an authorized system administrator, in case of blockchain this option is simple not an option. The security of the blockchain network is the appropriate combination of the cryptographically secured encryption function and the common investment of the network peers via the difficulty and cost of the consensus mechanism employed by the network. The longer a transaction is available on the blockchain the higher resources are necessary to perform a malicious attack against the information stored. Essentially, every newly created block of transactions attached to the chain creates a new amount of security for the already existing information stored.

978-1-5386-4636-6/17/$31.00 ©2017 IEEE

Blockchain is often called as distributed ledger, where the ledger can be seen as a tamper proof method to share information. The simplified mechanism of the distributed ledger technology consist of six consecutive steps, (1) broadcast a transaction and collect them in a transaction pool (2) create a block from the transaction pool and validate it (3) broadcast the created block to the peers (4) the members of the network validate the block and (5) the block is added to the chain (6) the starting transaction is settled and there is no way to change it any more.

The main advantages of the technology is its transparency, security and instant transaction speed. An additional and unique advantage of the technology that it makes possible the disintermediation which essentially establish the distributed electronic health records.

II. Problems of the centralised EHR Systems

A. Current Healthcare Infrastructure

Currently, there are a huge amount of separate health information systems, which hold the data of the individual patients in huge silos of health information. These information system are organized by different ways. The ways of organizing data depends on the goals of the health care provider's business. It is totally different in case of a diagnostic center or in case of a general practitioner. Anyway, in both cases ultimately (name, value) pairs describe the results of an encounter and different structuring procedures integrate the data into EHR records. The (name, value) pairs are implicitly always extended by several essentials attributes, where the time of the event represents a crucial role.

In order to integrate these isolated data silos a series of interfaces are built and maintained continuously. To resolve the problem of interfacing the different health data recording systems a wide range of protocols have emerged. Among the existing protocols there is no widely accepted one, so the problem of the data exchange is still a serious everyday problem.

However laws and incentive programs have made health care data more accessible, the vast majority of hospital systems cannot share patient data easily and safely. There are only rather general jurisdiction concerning patient privacy and there is no uniformly available technical solution ensuring patient privacy. Health medical information structures are so complex and context dependent that the many classification systems, protocols and medical thesaurus are rarely, if ever, used in their whole capacity in practice.

At the same time, the realignment from procedure and profession based health care delivery to the holistic approach based care delivery requires care providers accessing patient data under care delivery in an improved way, which also requires the higher level availability of an integrated EHR data structure.

To achieve a higher level and improved care outcome the research activities and second level health providers like drug companies, service industries also demanding for access to individual patient data to fulfill their job on a higher level. At the same time the personal health monitoring devices also deliver a big bunch of data, where the security and safety issues stick up again.

Hence the current solutions resulting a difficult choice between improved care deliveries on the price of decreased patient privacy. This tradeoff demands a solution as the problem emerges with varying intensity time to time.

Unfortunately, legal solutions and governmental promises are not ultimate solutions as experiences shows that changes in the technical and power structures can essentially hurt the existing centralized structures and organizational authorities. At the same time malicious attacks cannot be excluded which can ruin the existing security and consequently the privacy structures.

B. Simplified working model of patient-provider relations

The holistic care delivery approach assumes the access to the holistic view of the patient data. Different working models are in operation in different national health care systems.

What is common that the Principal Care providers are actively coordinate and collaborate with other care providers and ancillary health organizations like laboratories and pharmacies. The most common way of working to have a central responsibility – e.g. a general practitioner, or a responsible physician – who orders services and the contracted service providers delivering the requested data. These approaches can work rather well in case of a particular care event either focusing on a particular illness or focusing on a particular care level. In emergency situations, however, the centralized data storage systems have chance to provide the necessary data which could be life savers. However, centralized systems can provide the requested data in a rather smooth way, the main concern against it the centralization itself. There are too much authorization in the hand of the central authorities, which results an overcomplicated permission system, which contains the possibility of data breaches and information leaking.

C. Tradeoff between care delivery and reimbursement

Health care data manipulation is an existing problem in such systems where the care delivery influences the hospitals or other care provider's direct income. Registering the data of the care is an axiom in medical health delivery as the documentation is the base of later investigations, still happens that institution incentives are influences the history of a patient's data. Several times patients are reporting that there are unknown data in the central registries contradicting their own memories about symptoms, findings and therapies. Having an immutable blockchain based patient history such contradictions may not occur.

D. Availability of patient data

Data is one of the most valuable assets a health care system is producing. Access to this asset creates a lot of questions. The actors who are interested in these data are primarily the patient, the service person who delivers the care and third parties, who can use these data for different interests. Jurisdiction should provide a clear guideline that which third

parties, in what situations and for what purpose cab access the data. Currently, the regulation is unclear and inconsistent, therefore the health care systems contain generally inconsistent solutions if they even contain any solution.

III. A REGULATION TO START WITH - HIPAA

As the national regulations are different, we shall use the principles of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) to start with. In this paper we do not want to give an overall picture of the corresponding regulation, we only use some basic principles, which can be found more or less in every National jurisdiction.

According to the HIPAA rules, "The Privacy Rule… (applies) to health plans, health care clearinghouses, and to any healthcare provider who transmits health information in electronic form". In addition to these entities, those actors that act on behalf of the primary health providers are also responsible and should suit to the regulation. These actors are Business Associates who acts via Business Associate Contracts, which regulate the terms and conditions of patient data management.

The essential part of the HIPAA rules is summarized as "all individually identifiable health information held or transmitted by a covered entity or its business associates in any form or media, whether electronic, paper, or oral" are private patient health information. De-identified health information is defined as "Health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information". De-Identified data use restrictions are summarized by the following, "There are no restrictions on the use or disclosure of de-identified health information. De-identified health information neither identifies nor provides a reasonable basis to identify an individual". The boundary of identifiable data to de-identifiable data is defined as any information that may restrict the possible number of individuals a collection of information is associated with to a certain percent of the whole population. This percent could be argued and can depend on local conditions.

Fast local regulation and rules depend on the local jurisdiction, the basic principles are very similar to each other.

Concerning the principles listed above one can state that Blockchain technologies are especially suitable for fulfilling the demands of secure technology, privacy requirements, lack of centralized government and the reducing of the costs of providing a viable security infrastructure.

IV. CRYPTOGRAPHIC TOOLS AND SMART CONTRACTS

In order to understand the strength embedded in the blockchain based distributed ledger technologies and the special requirements of the health care delivery industry, we should mention several basic cryptographic notion especially suitable as being building blocks of the described infrastructure.

A. Public-key cryptography (PKI)

The Public-key cryptography have been well-known for more than 40 years, the technology implicated a series of methods and algorithms which are widely used in our everyday life. From the point of view of a distributed health information system the PKI algorithms are fundamental security ingredients of the blockchain based systems. Every entity in a blockchain based system are identified using a private-public key pair.

B. Digital signatures

Digital signatures are consequences of the PKI schemas. Digital signatures are used to secure the origin of information or a complete bunch of information. When the origin of the information is secured by authorized digital signatures than the chance to modify the information is practically impossible. This feature is used to ensure the non-repudiation of events.

C. Blind signature

Blind signature is a form of digital signature in which the content of a message is disguised before it is signed. This tool is suitable to hide sensitive patient information before transfer but still providing a proof about the content.

D. Multisignature

Multisignature refers to requiring more than one key to authorize a transaction. Multisignature is often called (m,n)-type signature, where an information is signed by n parties and the signature of m parties are necessary for authorization. Multisignature algorithms has a wider variation which are useful for solving different use cases in health ecosystems.

In a cryptographically secured healthcare information system the above mathematical tools can give a reliable toolset to build different cooperative subsystems where the privacy of the patient are secured and still the interested actors can access the appropriate information.

E. Smart contracts

Smart contracts are a blockchain based invention, however it is not an integral part of the blockchain concept. Smart contracts at the same time fundamentally important parts of many blockchain based ecosystems like Bitcoin and Ethereum. Smart contracts are similar to the database triggers, where based on the transaction data real time operations are performed at the very time of the transaction, which allows sophisticated behavior. Smart contracts were originally designed to perform legal contracts in electronic form. The smart contracts, which are computer scripts are performed on the underlying peer operator machines of the blockchain and the results of the scripts become integrated, immutable chunks of the blockchain data. Smart contracts are used to influence the default operation of the virtual machines executing the transaction.

We shall list in the next chapters what are those application fields where blockchain technology will disrupt the existing solutions.

V. EHR Application model

A. The Healthcare EMR storage network

In a global healthcare ecosystem the legacy EMR are centralized structures subject to hacking, strict security regulations and significant overhead costs. The actors are fundamentally classified as information producers and information users.

The information producers in case of a blockchain based infrastructure are providing information towards the blockchain where an incentivized peer-to-peer network is emitting it. During the emission of the information every peer do the same procedure according to a consensus rule system, resulting the information validated and secured. The validation happens by applying the validity rules to the information published by the actors while security happens by a special securing procedure called mining.

Mining itself is a resource demanding activity which buries the incoming information into the blockchain. As the mining procedure is costly the information security is growing together with the growths of the blockchain itself. The storage of the information is separated from the systems of the providers and happens through API calls. The providers have the patient's public key information and either applying some kind of blind signature or applying some kind of multi-signature, the security of the outgoing information is ensured. Miners, who perform the validation, securing and eventually storage of the information have no access to the information content itself.

The encrypted information is available only by controlled procedures where the access to information is logged and eventually the information owner is notified about the fact of accessing his/her data. The range of information providers are much wider in our cases as in legacy health information systems, having access to the blockchain using API-kits the unavailable part of providers can source information, like wearables, implants, home medical devices via the fast growing IoT technologies, as well.

The other part of actors are the users of the information, including the primary health service providers, research communities, financing institutes, emergency care providers and the patients themselves. Using the innovations of the cryptography industry, blind signatures, multi-signatures, hierarchical signatures and other secure procedures can ensure, that in case of an information request only the authorized requestors have access to the information while on the route no one can read any open text data.

The requested information travelling from the provider to the requestor is encrypted along the whole route independently from the number of transmitters involved. As the access keys are distributed evenly, no central organization has exclusive access to the data. This also means that the privacy issues remain on the level of the patient-information requestor level and there are no centrally managed databases involved.

The decentralized storage and decentralized management of the health data prevents organizing expensive and complex maintenance and auditing institutions while the incentives of the parties dynamically stabilize the underlying network. In case of complex queries specialized secondary network of aggregation servers have the task to provide the response where the de-identification filters can rule out the data not compliant. Also, as every transaction – both for storing and for querying – have its calibrated cost, where the transaction income can be shared for example between the patient and the independent data service provider. However, it is a crucial point of the blockchain operations, the cost structure can serve different purposes and should be designed carefully.

Blockchain-enabled health IT systems can provide technological solutions to many challenges, including health data interoperability, integrity and security, portable user-owned data. Blockchain could enable data exchange systems that are cryptographically secured and irrevocable. This would enable seamless access to historic and real-time patient data, while eliminating the burden and cost of data reconciliation.

B. Permissioned and permissionless availability

Creating EHR systems based on distributed blockchain technology is a complex task. The infrastructure to be achieved should contain both permissioned and permissionless subsystems.

Permission servers are applied on both institutional and health care organizational levels. At these levels permissions are organized into permission classes where registration is necessary for the individual actors who are using the permissions issued. Direct access to patient data presumes always permissioned access where the owner of the permission is identified. Fraudulent access to patient data is reported and cause defined consequences.

Permissionless access to the system is sufficient in such cases where patient data in open text do not occur, these situations are basically the area of information storing and forwarding. No open text information travel in the system, as the key infrastructure enables access to the end users only.

C. Incentives and operational costs

Blockchain ecosystems are based on some kind of incentives and some kind of consensus mechanisms. Permission based system, however, have no need for direct incentives as system owners define the incentives in an implicit manner. Also, permissioned systems are frequently proprietary systems where the consensus rules are embedded into the operation rules. On the other hand, the permissionless, open systems should have an incentive to take part in the game, which are usually coming up in the form of an artificial token, whose value creates a connection between the financing and operating parties. Tokens are already used in health economy systems as an independent measure of the value of different services. However, the tokens used for incentivizing the blockchain operators server primarily as payment tools.

The health tokens introduced are redeemed by the insurance companies and used by the operators of the storage and security service machines, in this way the solid network of servicing computers ensures the smooth operation of the system where the maintenance, backup and operation costs are undertaken by them.

VI. SUMMARY

In this paper I briefly introduce a new approach of an integrated health information model, where a complex and highly connected EHR can operate. The system is based on a newly discovered operational environment, the blockchain, where the heavy use of cryptographic tools makes possible to build a decentralized and openly extendable network. Elements of the solution are introduced. The new technology solves an essential problem of accessing data without endangering personal privacy. Also, the technology opens new opportunities for automatic personal monitoring devices and takes a step towards involving new service providers as well.

REFERENCES

[1] Adam Back. Hashcash - A Denial of Service Counter-Measure, Online, 2002.

[2] Andrew Poelstra. "Distributed Consensus from Proof of Stake is Impossible" (PDF).

[3] Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, Steven Goldfeder: Bitcoin and Cryptocurrency Technologies with a preface by Jeremy Clark - Draft — Feb 9, 2016

[4] Centers for Disease Control Prevention. “HIPAA privacy rule and public health. Guidance from CDC and the US Department of Health and Human Services.” In: (2003.)

[5] Rosenberg (ed.) Handbook of Financial Cryptography and Security. CRC Press, 2011.

[6] Chaum, A. Fiat, M. Naor. Untraceable electronic cash. CRYPTO 1998.

[7] Digital Assets on Public Blockchains - White Paper - BitFury Group - Mar 15, 2016

[8] HHS.gov. “H. H. S. O. of the Secretary Summary of the HIPAA Privacy Rule”. In:(2013).url:www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[9] Incentive Mechanisms for Securing the Bitcoin Blockchain - White Paper - BitFury Group -Dec 07, 2015

[10] Joseph Poon, Thaddeus Dryja: The Bitcoin Lightning Network: Scalable O-Chain Instant Payments - January 14, 2016 - DRAFT Version 0.5.9.2

[11] Katz, Jonathan, and Yehuda Lindell. Introduction to Modern Cryptography, Second Edition. CRC Press, 2014.

[12] Leslie Lamport. Time, Clocks, and the Ordering of Events in a Distributed System. Communications of the ACM, 21(7):558{565, Jul 1978.

[13] Nakamoto, Satoshi. Bitcoin: A peer□to□peer electronic cash system . (2008)

[14] Nick Szabo. Formalizing and Securing Relationships on Public Networks. http://szabo.best.vwh.net/formalize.html, Sep 1997.

[15] Chrissa McFarlane, Michael Beer, Jesse Brown, Nelson Prendergast: Patientory: A Healthcare Peer-to-Peer EMR https: //patientory.com/patientory_whitepaper.pdf, May, 2017

[16] Peter Todd. Near-zero fee transactions with hub-and-spoke micro-payments. http://sourceforge.net/p/bitcoin/mailman/message/33144746/, Dec 2014.

[17] Pieter Wuille. BIP 0032: Hierarchical Deterministic Wallets. https://github.com/bitcoin/bips/blob/master/bip-0032.mediawiki, Feb 2012.

[18] Proof of Stake versus Proof of Work - White Paper - BitFury Group - Sep 13, 2015

[19] R. Anderson. Security Engineering (2nd ed). Wiley, 2008.

[20] S. Haber, W. S. Stornetta. Secure names for bitstrings. CCS, 1997.

[21] Vitalik Buterin. "On Stake". "Hard Problems of Cryptocurrencies". Retrieved 23 January 2016. one thing has become clear: proof of stake is non-trivial

[22] Wood, Gavin. "Ethereum: A Secure Decentralised Generalised Transaction Ledger" (PDF). Retrieved 23 January 2016. Ethash is the planned PoW algorithm for Ethereum 1.0