Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?

Christian Esposito
University of Salerno

Alfredo De Santis
University of Salerno

Genny Tortora
University of Salerno

Henry Chang
University of Hong Kong

Kim-Kwang Raymond Choo
University of Texas at San Antonio

Editor:
Kim-Kwang Raymond Choo
raymond.choo@fulbrightmail.org

One particular trend observed in healthcare is the progressive shift of data and services to the cloud, partly due to convenience (e.g. availability of complete patient medical history in real-time) and savings (e.g. economics of healthcare data management). There are, however, limitations to using conventional cryptographic primitives and access control models to address security and privacy concerns in an increasingly cloud-based environment. In this paper, we study the potential to use the Blockchain technology to protect healthcare data hosted within the cloud. We also describe the practical challenges of such a proposition and further research that is required.

Healthcare is a data-intensive domain where a large amount of data is created, disseminated, stored, and accessed daily. For example, data is created when a patient undergoes some tests (e.g. computerized tomography or computerized axial tomography scans), and the data will need to be disseminated to the radiographer and then a physician. The results of the visit will then be stored at the hospital, which may need to be accessed at a later time by a physician in another hospital within the network.

It is clear that technology can play a significant role in enhancing the quality of care for patients (e.g. leveraging data analytics to make informed medical decisions) and potentially reduce costs by more efficiently allocating resources in terms of personnel, equipment, etc. For example, data

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 ©2018 IEEE

captured in paper form is hard to capture in systems (e.g. costly and data entry errors), costly to archive, and being available when needed. These challenges may lead to medical decisions not made with complete information, the need for repeated tests due to missing information or data being stored in a different hospital at a different state or country (at the expenses of increasing costs and inconvenience for the patients), etc. Due to the nature of the industry, ensuring the security, privacy, and integrity of healthcare data is important. This highlights the need for a sound and secure data management system.

HEALTH RECORDS IN ELECTRONIC FORMS AND HEALTH INFORMATION SYSTEMS

Generally, *Electronic Medical Records* (EMRs) contain medical and clinical data related to a given patient and stored by the responsible healthcare provider.[1] This facilitates the retrieval and analysis of healthcare data. To better support the management of EMRs, early generations of *Health Information Systems* (HIS) are designed with the capability to create new EMR instances, store them, and query and retrieve stored EMRs of interest.[2] HIS can be relatively simple solutions, which can be schematically described as a graphical user interface or a web service. These are generally the front-end with a database at the back-end, in a centralized or distributed implementation.

With patient mobility (both internally and externally to a given country) being increasingly the norm in today's society, it became evident that multiple stand-alone EMR solutions must be made interoperable to facilitate sharing of healthcare data among different providers, even across national borders, as needed. For example, in medical tourism hubs such as Singapore, the need for real-time healthcare data sharing between different providers and across nations becomes more pronounced.

To facilitate data sharing or even patient data portability, there is a need for EMRs to formalize their data structure and the design of HIS. *Electronic Health Records* (EHRs), for example, are designed to allow patient medical history to move with the patient or be made available to multiple healthcare providers (e.g. from a rural hospital to a hospital in the capital city of the country, before the patient seeks medical attention at another hospital in a different country).[3] EHRs have a richer data structure than EMRs. There have also been initiatives to develop HIS and infrastructures that are able to scale and support future needs, as evidenced by the various national and international initiatives such as the Fascicolo Sanitario Elettronico (FSE) project in Italy, the epSOS project in Europe, and an ongoing project to standardize sharing of EHRs.[4,5,6]

Recently, the pervasiveness of smart devices (e.g. Android and iOS devices and wearable devices) has also resulted in a paradigm shift within the healthcare industry.[7] Such devices can be user-owned or installed by the healthcare provider to measure the well-being of the users (e.g. patients) and inform/facilitate medical treatment and monitoring of patients. For example, there is a wide range of mobile applications (apps) in health, fitness, weight-loss, and other healthcare related categories. These apps mainly function as a tracking tool, such as registering user exercises/workouts, keeping the count of consumed calories, and other statistics (e.g. number of steps taken), and so on.

There are also devices with embedded sensors for more advanced medical tasks, such as bracelets to measure heartbeat during workouts, or devices for self-testing of glucose. For example, Leu and collaborators proposed a smartphone-based wireless body sensor network to collect user physiological data using body sensors embedded in a smart shirt.[8] The data (e.g. user's vital signs) can be continuously gathered and sent in real-time to a smart device, before being sent to a remote healthcare cloud for further analysis. Another example is Ambient Assisted Living solutions for healthcare designed to realize innovative telehealth and telemedicine services, in order to provide remote personal health monitoring.[9]

These developments have paved the way for *Personal Health Records* (PHR), where patients are more involved in their data collection, monitoring of their health conditions, etc, using their smart phones or wearable devices (e.g. smart shirts and smart socks).[10,11]

There are, however, a number of challenges associated with PHRs. For example, can we rely on the data collected by the patients themselves? Should the relevant healthcare providers certified data collected by the patients, and if so, how can this be done? Who should be legally liable for a misdiagnosis or delayed diagnosis, due to decisions being made on the data sent from the patient's device that is subsequently determined to be flawed or inaccurate (e.g. due to a malfunction sensor)?

Despite such challenges and potentially thorny legal issues, having a HIS based on an ecosystem of solutions that is able to seamlessly exchange data among themselves and provide the abstraction of a single health data storage for any given patient (e.g. physically distributed among multiple concrete software instances at multiple healthcare providers and mobile apps) will benefit all users, ranging from patients to healthcare providers to governments.

Cloud computing is a potential solution, due to the capability to support real-time data sharing regardless of geographical locations, to provide resource elasticity as needed, and to handle big data (e.g. hosting of big data analytical tools) to obtain useful insights from the analysis of big healthcare data for research and policy decision making.[12,13]

In Figure 1, we demonstrate how cloud help facilitate sharing of healthcare data among providers, supporting each provider in managing their data, providing a seamless way of exchanging and potentially certifying data between EHR and PHR, and providing a unified/comprehensive view of (the scattered) healthcare records for each patient. In other words, (federated) cloud computing can be used to interconnect the different healthcare providers and their PHR solutions, used by the providers to deal with any sudden or seasonal changes, and so on.

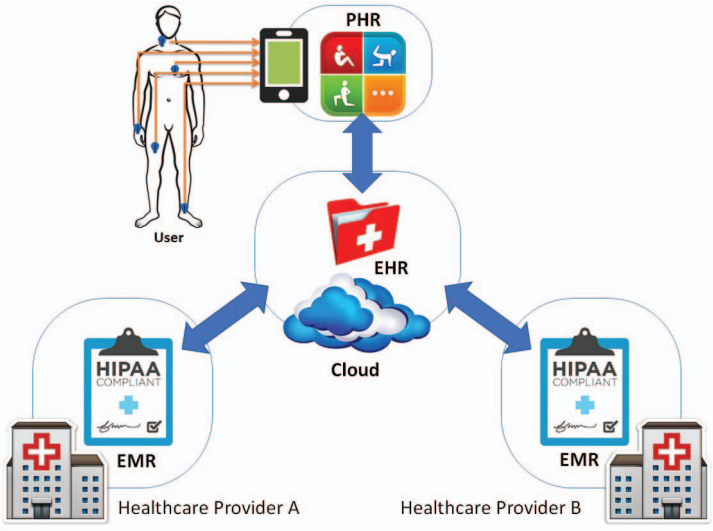

Figure 1. A conceptual cloud-based EMR/EHR/PHR ecosystem.

SECURITY AND PRIVACY

Healthcare data contain personal and sensitive information that may be attractive to cybercriminals. For example, cybercriminals seeking to benefit financially from the theft of such data may sell the data to a third-party provider, who may perform data analysis to identify individuals who may be uninsurable due to their medical history or genetic disorder. Such data would be of interest to certain organizations or industries.

Therefore, ensuring the security of the EMR/EHR/PHR ecosystem and the underlying systems and components that form the ecosystem is crucial, yet challenging due to the interplay and complexity between the systems and components. Moreover, the privacy and integrity of healthcare data must be protected not only from external attackers, but also from unauthorized access attempts from inside the network or ecosystem (e.g. employee of the healthcare provider, or cloud

service provider). The attacks (e.g. leakage or modification of data) can be intentional and unintentional, and organizations may be penalized or held criminally liable for such incidents, for example under the Health Insurance Portability and Accountability Act.

How to secure EMR/EHR/PHR ecosystem and ensure privacy and integrity of the data is an active research area. Approaches include using cryptographic primitives, such as those based on public key infrastructure and public clouds to ensure data confidentiality and privacy.[14] For example, data is encrypted prior to outsourcing to the cloud. However, this limits the searchability of the data, in the sense that healthcare providers have to decrypt the (potentially big) data prior to searching on the decrypted data, resulting in increases in time and costs for the data retrieval and diagnosis (e.g. download, decrypt, and search).[15]

Access control models have also been used to regulate and limit access to the data, based on predefined access policies.[16] Such models can be particularly effective for external attacks, but are generally ineffective against internal attackers as they are likely to be authorized to access the data. There have also been approaches to integrate access control with some cryptographic primitives, such as attribute-based encryption.[17]

BLOCKCHAIN TO THE RESCUE?

There has been recent interest in utilizing blockchain (made popular by the successful Bitcoin) in the provision of secure healthcare data management.[18,19,20] Broadly speaking, blockchain is a technology able to build an open and distributed online database, which consists a list of data structures (also known as blocks) that are linked with each other (i.e. a block points to the following one, hence the name blockchain). These blocks are distributed among multiple nodes of an infrastructure, and are not centrally stored. Each block contains a timestamp of its production, the hash of the previous block and the transaction data, and in our context, a patient's healthcare data and the healthcare provider information.

Figure 2 describes our conceptual blockchain-based EMR/EHR/PHR ecosystem. Specifically, when new healthcare data for a particular patient is created (e.g. from a consultation, and medical operation such as a surgery), a new block is instantiated and distributed to all peers in the patient network. After a majority of the peers have approved the new block, the system will insert it in the chain. This allows us to achieve a global view of the patient's medical history in an efficient, verifiable, and permanent way. If the agreement is not reached, then a fork in the chain is created and the block is defined as an orphan and does not belong to the main chain. Once the block has been inserted into the chain, the data in any given block cannot be modified without modifying all subsequent blocks. In other words, modification can be easily detected. As block content is publicly accessible, healthcare data needs to be protected prior to the data being in the block (e.g. obfuscated and perhaps, encrypted).

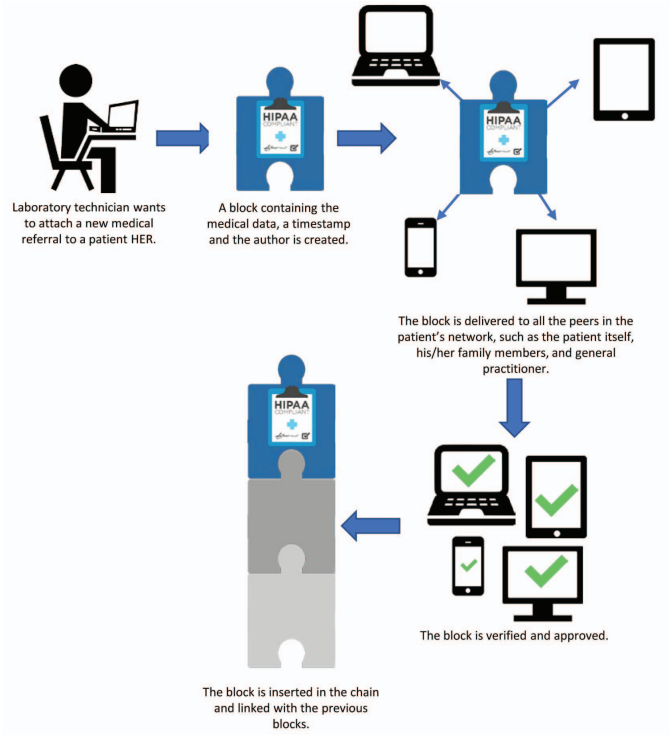

Figure 2. A conceptual blockchain-based EMR/EHR/PHR ecosystem.

Conceptually, blockchain is secure by design that provides the capability to achieve decentralized consensus and consistency, and resilience to intentional and/or unintentional attacks. Key benefits of deploying a blockchain in our approach are as follows:

1. Agreement can be reached without the involvement of a trusted mediator; thus, avoiding a performance bottleneck and a single point of failure;
2. Patients have control over their data;
3. Medical history as a blockchain data is complete, consistent, timely, accurate, and easily distributed; and
4. Changes to the blockchain are visible to all members of the patient network, and all data insertions are immutable. Also, any unauthorized modifications can be trivially detected.

As with any security solutions, there are limitations associated with a blockchain-based approach that need to be carefully studied. For example, blockchain technology can be somewhat disruptive and requires a radical rethink and significant investment in the entire ecosystem (e.g. replacement of existing systems and redesigning of business processes). In other words, before taking the plunge, healthcare providers particularly publicly funded providers will need to undertake a cost benefit analysis to understand the return on investment and any potential implications (e.g. legal and financial). For example, the same record can reside in multiple nodes of the network, located in different countries with different privacy and data protection requirements (e.g. EU and US).

CHALLENGES

While data integrity and distributed storage/access of blockchain offer opportunities for healthcare data management, these same features also pose challenges that need further study.[21]

The strong data integrity feature of blockchain results in immutability that any data, once stored in blockchain, cannot be altered or deleted. However, if the record is healthcare data, then such

personal data would come under the protection of privacy laws, many of them would not allow personal data to be kept perpetually—Article 17 of the soon-enforceable General Data Protection Regulation in the EU has strengthened the rights of individuals to request personal data to be erased. One of the principles of the Organization for Economic Cooperation and Development privacy guideline, on which many data protection laws are based, provides the right-to-erasure to individuals. Given the sensitivity of healthcare data, anyone planning to use blockchain to store them cannot ignore this legal obligation to erase personal data if warranted.

Another practical issue is on how fit it is for blockchain to store healthcare data. Blockchain was originally designed to record transaction data, which is relatively small in size and linear. In other words, one only concerns itself about whether the current transaction can be traced backwards to the original "deal". Healthcare data, such as imaging and treatment plans, however, can be large and relational that requires searching. How well blockchain storage can cope with both requirements is currently unclear.

In order to deal with these challenges, many have suggested the notion of off-chain storage of data, where data is kept outside of blockchain in a conventional or a distributed database, but the hashes of the data are stored in the blockchain. This is said to be the best of both worlds, as healthcare data is stored off-chain and may be secured, corrected, and erased as appropriate. At the same time, immutable hashes of the healthcare data are stored on-chain for checking the authenticity and accuracy of the off-chain medical records.

This idea, however, is not without potential challenges. With the tightening of data protection laws around the world and the attempts by privacy commissioners to regard metadata of personal data as personal data, it may not be very long that hashes of personal data are considered as personal data; then the whole debate of whether blockchain is fit to store personal data may start all over again.

REFERENCES

1. M. Steward, "Electronic Medical Records," *Journal of Legal Medicine*, vol. 26, no. 4, 2005, pp. 491–506.
2. R. Hauxe, "Health Information Systems—Past, Present, Future," *Int'l Journal of Medical Informatics*, vol. 75, no. 3–4, 2006, pp. 268–281.
3. K. Häyrinena et al., "Definition, Structure, Content, Use and Impacts of Electronic Health Records: A Review of the Research Literature," *Int'l Journal of Medical Informatics*, vol. 77, no. 5, 2008, pp. 291–304.
4. M. Ciampi et al., "A Federated Interoperability Architecture for Health Information Systems," *Int'l Journal of Internet Protocol Technology*, vol. 7, no. 4, 2013, pp. 189–202.
5. M. Moharra et al., "Implementation of a Cross-Border Health Service: Physician and Pharmacists' Opinions from the epSOS Project," *Family Practice*, vol. 32, no. 5, 2015, pp. 564–567.
6. S.H. Han et al., "Implementation of Medical Information Exchange System Based on EHR Standard," *Healthcare Informatics Research*, vol. 16, no. 4, 2010, pp. 281–289.
7. D. He et al., "A Provably-Secure Cross-Domain Handshake Scheme with Symptoms-Matching for Mobile Healthcare Social Network," *IEEE Transactions on Dependable and Secure Computing*, vol. PP, no. 99, 2016; doi.org/DOI: 10.1109/TDSC.2016.2596286.
8. F.Y. Leu et al., "A Smartphone-Based Wearable Sensors for Monitoring Real-Time Physiological Data," *Computers and Electrical Engineering*, 2017.
9. M. Memon et al., "Ambient Assisted Living Healthcare Frameworks, Platforms, Standards, and Quality Attributes," *Sensors*, vol. 14, no. 3, 2014, pp. 4312–4341.
10. P.C. Tang et al., "Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption," *Journal of the American Medical Informatics Assoc.*, vol. 13, no. 2, 2006, pp. 121–126.
11. S. Marceglia et al., "A Standards-Based Architecture Proposal for Integrating Patient mHealth Apps to Electronic Health Record Systems," *Applied Clinical Informatics*, vol. 6, no. 3, 2015, pp. 488–505.

12. A. Mu-Hsing Kuo, "Opportunities and Challenges of Cloud Computing to Improve Health Care Services," *Journal of Medical Internet Research*, vol. 13, no. 3, 2011.
13. V. Casola et al., "Healthcare-Related Data in the Cloud: Challenges and Opportunities," *IEEE Cloud Computing*, vol. 3, no. 6, 2016, pp. 10–14.
14. S. Nepal et al., "Trustworthy Processing of Healthcare Big Data in Hybrid Clouds," *IEEE Cloud Computing*, vol. 2, no. 2, 2015, pp. 78–84.
15. G.S. Poh et al., "Searchable Symmetric Encryption: Designs and Challenges," *ACM Computing Surveys*, vol. 50, no. 3, 2017.
16. Q. Alam et al., "A Cross Tenant Access Control (CTAC) Model for Cloud Computing: Formal Specification and Verification," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 6, 2017, pp. 1259–1268.
17. M. Li et al., "Scalable and Secure Sharing of Personal Health Records in Cloud Computing Using Attribute-Based Encryption," *IEEE Transactions on Parallel and Distributed Systems*, vol. 8, no. 3, 2016, pp. 2084–2123.
18. F. Tschorsch and B. Scheuermann, "Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, 2016, pp. 2084–2123.
19. A. Azaria et al., "MedRec: Using Blockchain for Medical Data Access and Permission Management," *Proceedings of the 2nd Int'l Conference on Open and Big Data* (OBD 16), 2016, pp. 25–30.
20. J. Zhang, N. Xue, and X. Huang, "A Secure System for Pervasive Social Network-Based Healthcare," *IEEE Access*, vol. 4, 2016, pp. 9239–9250.
21. J. McKinlay et al., "Blockchain: Background, Challenges and Legal Issues," *DLA Piper Publications*, 2016; doi.org/https://www.dlapiper.com/en/uk/insights/publications/2017/06/blockchain-background-challenges-legal-issues/.

ABOUT THE AUTHORS

Christian Esposito is an adjunct professor at the University of Naples "Federico II," where he received his PhD in computer engineering and automation. He is also a research fellow at the University of Salerno, Italy. Esposito's research interests include reliable and secure communications, middleware, distributed systems, positioning systems, multiobjective optimization, and game theory. Contact him at christian.esposito@dia.unisa.it.

Alfredo De Santis is a professor of computer science and the department director at the University of Salerno, where he received a degree in computer science. His research interests include data security, cryptography, digital forensics, communication networks, data compression, information theory, and algorithms. Contact him at ads@unisa.it.

Genny Tortora is a full professor of computer science and was dean of the faculty of Mathematical, Natural, and Physical Sciences from 2000 to 2008 at the University of Salerno, where she received a degree in computer science. Her research interests include software-development environments, visual languages, geographical information systems, biometry, and virtual reality. Contact her at tortora@unisa.it.

Henry Chang is an adjunct associate professor at the Department of Law in the University of Hong Kong. His research interests include technological impact on privacy. Chang is a fellow of the British Computer Society and a member of the Hong Kong/Guangdong ICT Expert Committee on Cloud. Contact him at hcychang@hku.hk.

Kim-Kwang Raymond Choo is the holder of the cloud technology endowed professorship in the Department of Information Systems and Cyber Security at the University of Texas at San Antonio. His research interests include cyber and information security and digital forensics. He is a senior member of IEEE, a fellow of the Australian Computer Society, and has a PhD in information security from Queensland University of Technology, Australia. Contact him at raymond.choo@fulbrightmail.org.