


Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control

Xiao Yue¹ · Huiju Wang²  · Dawei Jin² · Mingqiang Li³ · Wei Jiang²

Received: 31 May 2016 / Accepted: 11 August 2016
© Springer Science+Business Media New York 2016

Abstract Healthcare data are a valuable source of healthcare intelligence. Sharing of healthcare data is one essential step to make healthcare system smarter and improve the quality of healthcare service. Healthcare data, one personal asset of patient, should be owned and controlled by patient, instead of being scattered in different healthcare systems, which prevents data sharing and puts patient privacy at risks. Blockchain is demonstrated in the financial field that trusted, auditable computing is possible using a decentralized network of peers accompanied by a public ledger. **In this paper, we proposed an App (called Healthcare Data Gateway (HGD)) architecture based on blockchain to enable patient to own, control and share their own data easily and securely without violating privacy, which provides a new potential way to improve the intelligence of healthcare systems while keeping patient data private.** Our

proposed purpose-centric access model ensures patient own and control their healthcare data; simple unified Indicator-Centric Schema (ICS) makes it possible to organize all kinds of personal healthcare data practically and easily. We also point out that MPC (Secure Multi-Party Computing) is one promising solution to enable untrusted third-party to conduct computation over patient data without violating privacy.

Keywords Healthcare data system · Indicator-centric schema · BlockChain · Healthcare data sharing · Privacy risk

Introduction

Implementation of EMR (Electronic Medical Record) has been considered to be a critical role to improve healthcare intelligence, quality, user experience and related costs. Kemkar et al. reported that the EMR system could eventually save more than billions annually [1]. Sharing of healthcare data will help us to become smarter, for instance, better understanding of patterns and trends in public health and disease to ensure better quality care [2]; better recommendations for exercising or doctors [3]; plan services that make the best of limited national health service budgets for the health and wellbeing of everyone, and so on. For discussion convenience, we use healthcare data to represent patient data, and healthcare data system to denote all kinds of systems that generate, access or store patient data, throughout the paper.

A big challenge for healthcare data systems to become smarter is how to gather, store and analyze personal healthcare data without raising privacy violations. For such systems, privacy concerns have proved barriers to adopt

This article is part of the Topical Collection on *Mobile & Wireless Health*

✉ Huiju Wang
wanghuiju.cn@hotmail.com

¹ School of Political Science and Public Administration, Huaqiao University, No.269 Chenghua North Rd. Quanzhou, Fujian, 362021, China

² School of Information and Safety Engineering, Zhongnan University of Economics and Law, 182 Nanhu Ave, East Lake High-Tech Development Zone, Wuhan, 430073, China

³ School of Public Administration, Zhongnan University of Economics and Law, 182 Nanhu Ave, East Lake High-Tech Development Zone, Wuhan, 430073, China

healthcare data system. Anecdotal evidences from recent years suggest that a lack of adequate security measures has resulted in numerous data breaches, leaving patients exposed to economic threats, mental anguish and possible social stigma [4].

Patient data is one valuable asset of patient. Despite privacy concerns, 90 % of Americans valued online access to their health records [5]. However, physicians and patients each have a different set of information gathered over the years. A central location where all of patient information can be compiled and accessible to both physicians and patients alike is highly demanded.

Healthcare data should not be trusted in the hands of third-parties, where they are susceptible to attack and misuse. In this work, based on blockchain storage platform, we designed the architecture of **HDG (Healthcare Data Gateway)**, a smart App on smartphone that enables patient to manage and control the sharing of their healthcare data easily. HDG is a combination of traditional database and gateway: it manages personal electronic medical data on blockchain storage system, evaluates all data requests by leveraging purpose-centric access control and utilizes secure multi-party computation to enable third-party to conduct processing on patient data without risking patient privacy. The proposed architecture does not depend on any third-party and no single party has absolute power to affect the processing.

Smartphone is one ideal platform to design such App for patients to manage their own data because of:

- its mature computing power which guarantees the implementation.
- the popularity of smartphone app stores, which allows an application to be installed in seconds, further lowering the barrier to user adoption.
- its popularity, which make it easy-to-get device to use HDG.
- speeding mobile wireless network, such as cloud-based 5G network [6], which guarantees the performance.

Blockchain is first used in Bitcoin as an accounting system with a public, secure, and verified ledger, and may be viewed as storage supply chain in which every operation may be verified, accountable and immutable. Such inherent characteristics make it a potential solution for healthcare data systems that concerns both sharing and patient privacy. To our best knowledge, we are the first one to import blockchain into the design of healthcare data system.

The structure of the paper is as follows. Section “Related Work” surveys previous work, and Section “HDG-centric Healthcare Ecosystem” presents the whole healthcare ecosystem around HDG. Section “Design of Healthcare Data Gateway” introduces the core design of HDG followed by the concluding remarks.

Related work

The existing works that most closely related to ours lay in the area of cloud-based healthcare data systems. We summarize them as follows.

Studies [7–9] designed a country level framework for electronic medical systems based on cloud models. For example, Patra et al. [9] proposed a cloud-based model to build one information system on a national level which provided a cost effective way in dealing with patient information for rural areas. People are encouraged to provide their personal healthcare information which will be stored in the health cloud and accessed by the medical professionals and policy makers to provide more medical services, such as remote disease diagnosis and control, etc. Rolim et al. [10] proposed a framework covering the process from data collection to data delivery. Using attached sensors on medical equipment, data may be collected and be stored in a cloud directly, which can be accessed by authorized medical professionals. Yin et al. [11] introduced a patient-centric system built on cloud with data collection layer, data management layer and data service layer.

Attribute based Encryption (ABE) is one of the most popular encryption schemes used in cloud computing. Barua et al. [12] and Narayan et al. [13] proposed a patient-centric and privacy-preserving ABE based access control model; Chen et al. [14] described a novel framework with Cloud-based Privacy-aware Role Based Access Control model which may be used for controllability, traceability of data and authorized access to healthcare data resources.

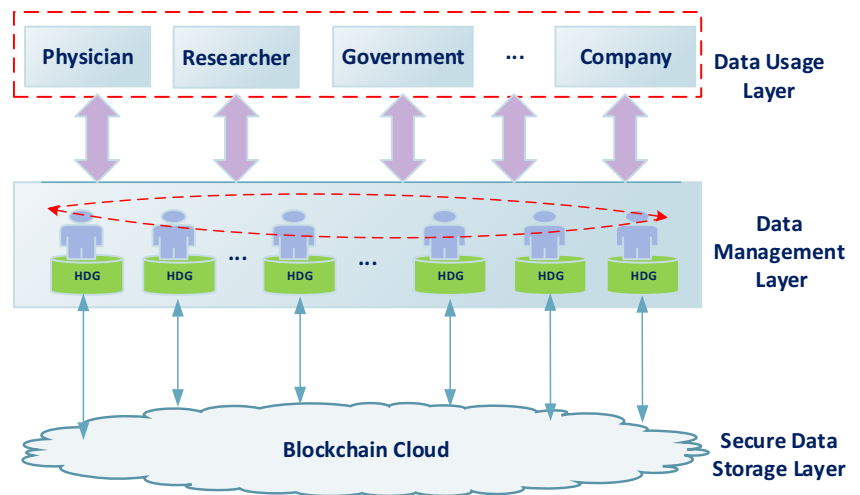
HDG-centric healthcare ecosystem

In healthcare data systems, patient data are scattered in different systems. Patients have no control of their own data. Our proposed architecture aims at enabling patient to manage all his own data through their data gateway securely (as illustrated in Fig. 1). It consists of three layers:

Storage Layer This layer provides scalable, secure, highly available and independent storage service for healthcare data, against confidentiality and integrity attacks. The data are stored in the private blockchain cloud. Blockchain may guarantee medical data cannot be changed by anybody including physicians and patients himself/herself internally and natively. Cryptographic techniques (i.e., encryption, hashing, signatures) are used to protect data.

Data Management Layer Consists of a set of individual’s HDGs that are independent and connected with each other. Data gateway is the hybrid of firewall concept and database

Fig. 1 HDG-centric Healthcare Ecosystem



concept. On one hand, it works as a gateway which evaluates all data accesses, incoming ones and outgoing ones, without bypassing it; on the other hand, it acts as a database manager which manages all kinds of heterogeneous personal data, including not only patient's own data but also authorized other's data. HDG treats metadata and data differently and mainly focuses on management of metadata (healthcare indicator, schema, etc.), index information etc. Metadata should be sufficient to allow performing queries before accessing the cloud to retrieve the data of interest.

Data Usage Layer Entities that use patient healthcare data are included in this layer. Typical examples are electronic medical record systems, data analytics algorithms (such as [15]), other's HDG, physicians, etc.

Design of healthcare data gateway

One HDG tour scenario

In this subsection, we take a guided tour of HDG through one scenario. To make this scenario possible, patients and practitioners are all assumed to be equipped with HDGs.

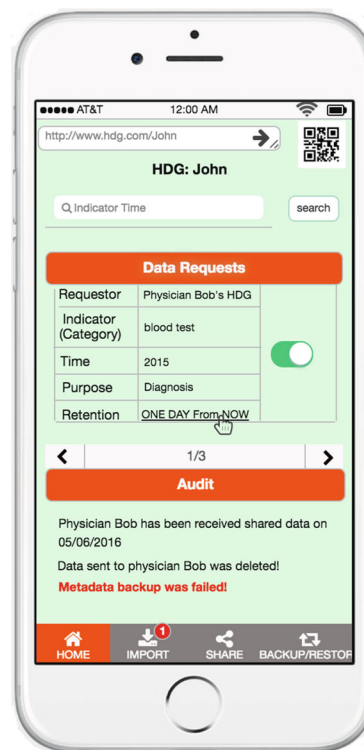
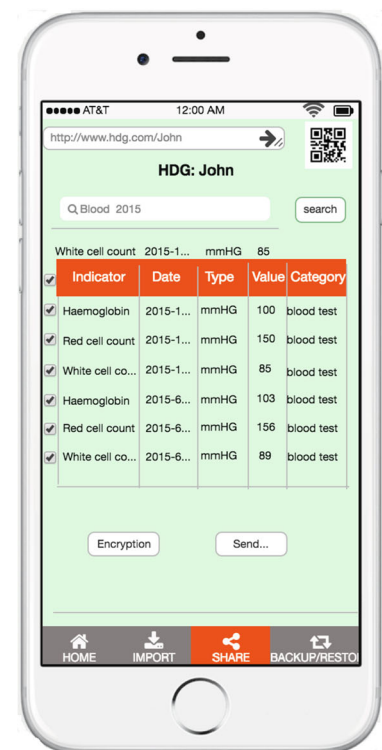
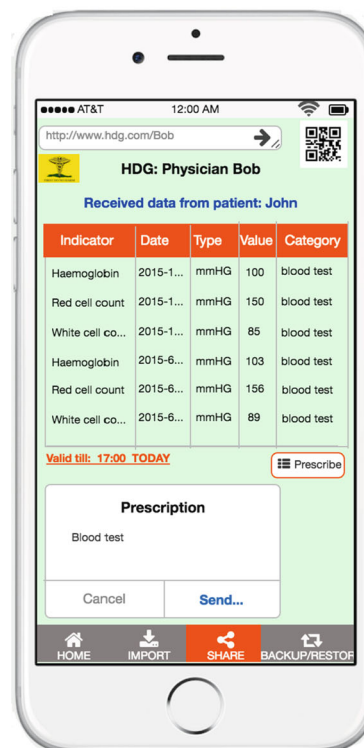
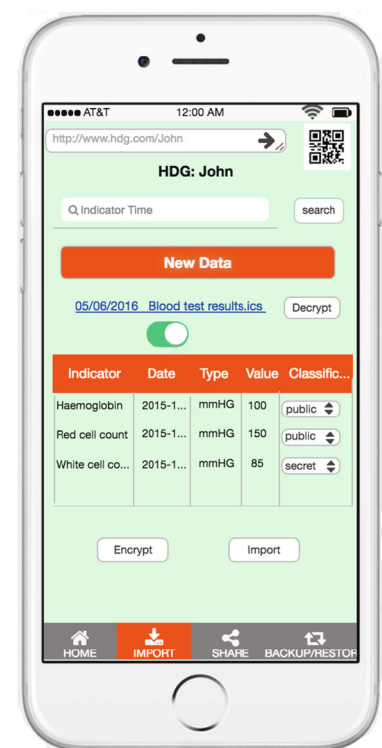
John is a patient. When he visits his physician Bob, he is free to decide, depending his willingness to grant Bob to access his blood test data. In the positive case, John decrypts his blood test data and encrypts it with a new key and sent the data and the new key to Bob (as shown in Fig. 2a and b). After receiving the data, Bob may use the key to query the data (as shown in Fig. 2c). While Bob's HDGs holds John's data replica, Bob cannot operate on the replica exceeding their authorities and their HDGs will enforce to destroy the replica after one day period because the authorized period is expired. Bob actions are recorded by their own HDGs and sent back to John for audit purpose.

Bob prescribes a blood test to John. The electronic blood test results are sent to John's account in the blockchain cloud in an encrypted form and fed to John's HDG. If the results contain information that John would like to keep secret, he simply masks this file so that it remains hidden from any user querying the healthcare data except him, i.e., indicator "White cell count" in Fig. 2d. The hospital keeps just the track of this medical act for administrative purposes without test results.

Finally, if the Ministry of Health decides to count how many people has the same problem with John. The Ministry of Health writes one counting function and submits it to run in the blockchain cloud. For the counting function, one possible secure code transformation method may be implemented by importing one random value: each HDG generates one random value and keeps it securely; then each HDG adds the random value and 1 to the counting value if the patient has the same disease. After the last HDG finishes the counting, it subtracts the sum of all random values from the result and returns the final value to the Ministry of Health. Such algorithm transformation is shown in Fig. 3. In this way, no one learns whether the HDG's owner has the same problem with Bob or not, but final result is computed.

ICS: A unified data schema

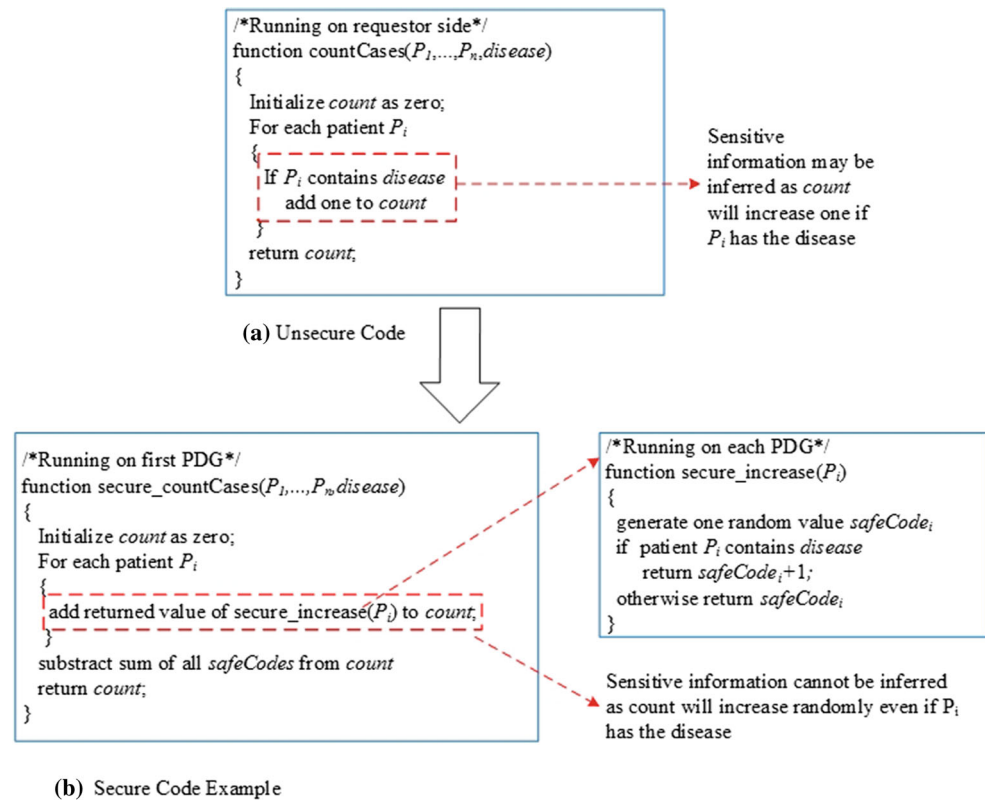
A challenging problem for HDG is how to manage all kinds of data, i.e., records, text, images, etc. Schema is the most popular data organization model adopted by database management system. Its key idea is to design one schema with multiple decomposed tables in which each table describes one type of entities; then upload data into tables. For HDG, we doubt its practicability as it requires common users to conduct some professional operations, such as defining schema and query.

Fig. 2 Illustration of core interfaces of HDG**(a)** Home user interfaces of John's HDG**(b)** Data sharing illustration of John's HDG**(c)** Data sharing illustration of physician's HDG**(d)** Data import user interface of John's HDG

HDG follows the “upload data once, query it many times” model. Data is immutable once uploaded. Accordingly there is no need to decompose tables to reduce

redundancy and achieve integrity. So, instead of organizing patient data into tables, we propose to use one simple and unified schema to model all kinds of data and liberate users

Fig. 3 Illustration of Secure Computing Code Transformation



from the schema issue. We use one simple “table” to organize one patient’s all data: $\langle Time, Indicator, Type, Value, Description \rangle$, where

- Time: denotes when the value is generated.
- Indicator: describes the meaning of the value.
- Type: data type of the value, such as unit, text, image, video, etc.
- Value: actual data value.
- Category: category of the indicator (as shown in Fig. 2a).

Besides, both indicator and indicator category information may be used to query data (as shown in Fig. 2a). It is noted that, for some complex value, indicator may composite multiple indicators. Take test results in a 2-dimension table as one example, the indicator may consist of two sub-indicators: column indicator and row indicator.

Each patient has one such schema table that takes patient’s identity as table name. As HDG may also store other’s data, we may create one such schema for each patient. We call such data organization method **Indicator-Centric Schema (ICS for short)** model.

Figure 4 shows how we organize one blood test results and one Visante OCT report of a patient into such ICS model. Personal blood indicators are stored in the table as shown in red dash rectangle and Visante OCT results are

stored as one image. Information about the medical act, such as who conduct the test in which lab, will be kept in hospital own system as it is not private information about patient.

Further, data is organized physically using index built on *Indicator (Category)* and *Time*. Popular multi-dimensional index such as KD-Tree [16, 17], R-Tree [18, 19], and composite index, etc., can all be adopted. Here, we propose multi-level index and multi-dimensional (*LD-Index* for short, as illustrated in Fig. 5) for our application. In one LD-Index, data is first indexed based on *Indicator (Category)* using Hash-index, then for each indicator (category), is further indexed on time using B⁺ tree index. In this way, LD-Index can be used not only as index but also as directory to organize data (especially for index on *Indicator (Category)* field). Data retrieval may be processed by traversing successively the hash index, B⁺ tree index and then data items in leaf node.

Comparing with traditional databases which use tables and records to organize data, ICS stores data at the table cell granularity, which makes it more flexible and easy to integrate healthcare data.

Privacy-aware data access control

In healthcare data system, healthcare data should be created and altered by physician professionals and cannot be

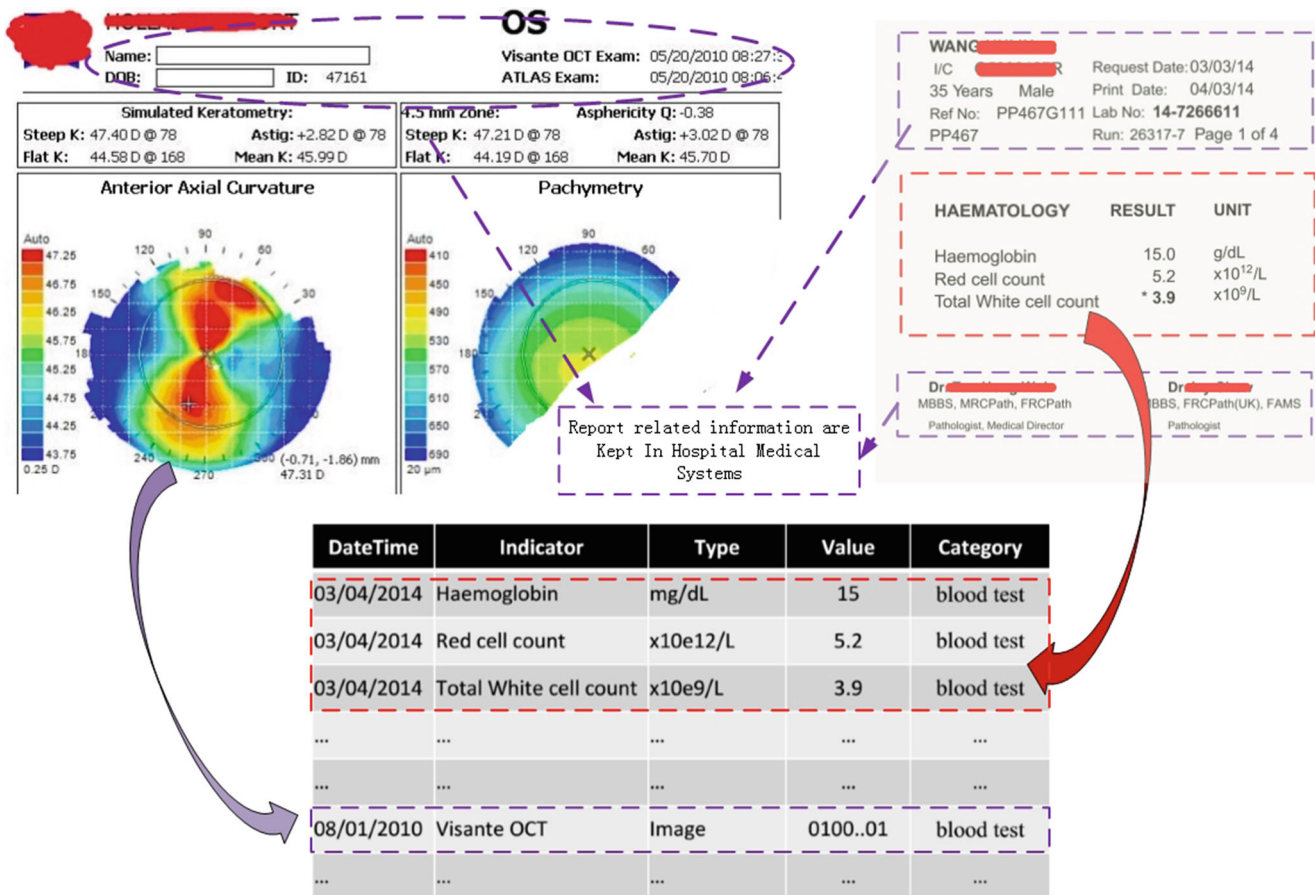


Fig. 4 Illustration of Indicator-Centric Schema

altered by anyone including patient himself/herself. Personal healthcare data is owned by patient and used by requestors under the authorization of data owner.

Privacy-aware data access policy cannot be easily achieved by traditional access control models. The first reason is that traditional access control models focus on who is performing which action on what data object, privacy policies are concerned with what data object is used for which purpose(s).

We propose one purpose-centric access control model. Based on ICS model, data request is expressed using one unified interface $\langle \text{Requestor}, \text{Time}, \text{Indicator (Category)}, \text{Purpose}, \text{Retention} \rangle$, where *Requestor* indicates source HDG that is requesting data, *Indicator (Category)* is the search keyword for the data item requested (i.e., indicator name, indicator category), *Purpose* describes access purpose, and *Retention* shows desired access period. In our access model, we divide data access purpose into two types: 1). raw purpose which is to access raw data. Typical example is that physician access healthcare data for diagnosis, and nurse for care. 2). process purpose which is to derive some results based on raw data, and does not concern raw

data. Each data request contains the unique URL, QR-code of target HDG and parameters of the data request.

Data query

Based on access purpose, we further divide data users into two types: *r-users* refer to users who need to access raw data, and *p-users* are users to use raw data to derive results.

For data query issued by *r-users*, HDG processes it by sending a replica of data to requestor's HDG, and requestor's HDG enforces sticky policy (such as data deletion) automatically without requestor's permission.

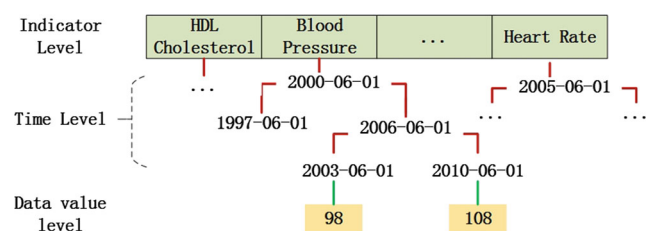


Fig. 5 Illustration of LD-index

Practically, sharing data means sharing the associated meta-data (so that the recipient user can understand the data), the cryptographic keys (so that her data gateway can decrypt them) and the sticky policy (so that her data gateway can enforce the expected access control rules).

Once a service accesses a piece of raw data, it could store it for future analysis, even disclose it to others. Thus, for *p-users*, one challenging problem is how to enable requestor to accomplish their computations while keeping data private. So if the purpose is just for processing, we adopt one secure data processing model, which is to never allow a service to observe the raw data, but instead, to let it run computations directly on the blockchain network and obtain the final results, which may be implemented by utilizing MPC (Secure Multi-Party Computation) technique as in [20]. It was already shown that any function can be securely computed, with security for malicious adversaries [21, 22], which means any data processing function may be transformed into secure code. Fig. 3 shows one running example.

Other functions

HDG is also equipped with some necessary data management functions and emphasize on privacy protection.

Anonymization. HDG can anonymize data which remove identified personal information if necessary. It implements the transformation on its own if possible.

Communication. HDG can communicate with other HDGs for data requests or collaboration.

Data backup and recovery. Backup information on HDG to cloud and recover information if necessary.

Conclusion

Sharing of healthcare data is vital important for intelligence healthcare services. Unfortunately, healthcare data are scattered in different healthcare data systems. And it is awkward to share data, as it will violate patient's privacy possibly. Healthcare data is one valuable asset of patient. It is natural to enable patients to own and control their data without compromising security or limiting the sharing of healthcare service. Our architecture enables this by utilizing blockchain platform as storage system, purpose-centric access as one access-control model, and unified and simple Indicator-centric schema as storage model. Based on our architecture, patients are not required to trust any third-party and are always aware that who is accessing his data and how it will be used. With a decentralized platform and HDG-based central control, making legal and regulatory decisions about collecting, storing and sharing patient data become simpler. HDG provides a potential way to house and

share healthcare data while keeping privacy and lays the root for smart healthcare services from the architecture point of view.

Acknowledgment This work is partly supported by the Humanity and Social Science Youth Foundation of Ministry of Education of China (Grant No.14YJC630181), the Fundamental Research Funds for the Central Universities of China(Grant No. JB-SK1206), the Scientific Research Fund for Talent Introduction of Zhongnan University of Economics and Law (Grant No. 21141611313), the Scientific Research Fund for Talent Introduction of Huaqiao University (Grant No. 12Y0324) and National Social Science Foundation for Young Scholars (Grant No.13CTJ003).

References

1. Kemkarl, O. S., and Dahikar, D. P. B., Can electronic medical record systems transform health care? potential health benefits, savings, and cost using latest advancements in ict for better interactive healthcare learning. *International Journal of Computer Science & Communication Networks* 2(3/6):453–455, 2012.
2. Bordea, G., Jothi, N., Rashid, N. A., and Husain, W., Data mining in healthcare: a review. *Procedia Computer Science* 72:306–313, 2015. doi:10.1016/j.procs.2015.12.145.
3. Zhang, Y., Chen, M., Huang, D., Wu, D., and Li, Y., idocor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. *Futur. Gener. Comput. Syst.*, 2016. doi:10.1016/j.future.2015.12.001.
4. ONA General Counsel.: Members and patient privacy: be aware and beware!. <https://www.ona.org/> (2016). [Online; accessed 28-Jul-2016].
5. The Office of the National Coordinator for Health Information Technology (ONC), USA.: The value of consumer access & use of online health records. <https://www.healthit.gov/> (2015). [Online; accessed 28-Jul-2016].
6. Chen, M., Zhang, Y., Hu, L., Taleb, T., and Sheng, Z., Cloud-based wireless network: Virtualized, reconfigurable, smart wireless network to enable 5g technologies. *Mobile Networks and Applications* 20(6):704–712, 2015. doi:10.1007/s11036-015-0590-7.
7. Hendrick, E., Schooley, B., and Gao, C., Cloudhealth: Developing a reliable cloud platform for healthcare applications. In: 2013 IEEE 10th Consumer Communications and Networking Conference (CCNC), pp. 887–891, 2013. doi:10.1109/CCNC.2013.6488579.
8. Gul, O., Al-Qutayri, M., and Vu, Q. H., Framework of a national level electronic health record system. In: 2012 International Conference on Cloud Computing Technologies, Applications and Management (ICCTAM), pp. 60–65 2012, p. Yeun, C.Y. doi:10.1109/ICCTAM.2012.6488072.
9. Patra, M. R., Das, R. K., and Padhy, R. P., Crhis: Cloud based rural healthcare information system. In: Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance, ICEGOV '12, pp. 402–405. ACM, New York, NY, USA, 2012. doi:10.1145/2463728.2463805.
10. Rolim, C. O., Koch, F. L., Westphall, C. B., Werner, J., Fracalossi, A., and Salvador, G. S., A cloud computing solution for patient's data collection in health care institutions. In: 2nd International Conference on eHealth, Telemedicine, and Social Medicine, 2010. ETELEMED '10. pp. 95–99, 2010. doi:10.1109/ETELEMED.2010.19.

11. Zhang, Y., Qiu, M., Tsai, C. W., Hassan, M. M., and Alamri, A., Health-cps: Healthcare cyber-physical system assisted by cloud and big data. *IEEE Syst. J.* PP(99):1–8, 2015.
12. Barua, M., Liang, X., Lu, R., and Shen, X., Espac: Enabling security and patient-centric access control for ehealth in cloud computing. *Int. J. Secur. Netw.* 6(2/3):67–76, 2011. doi:[10.1504/IJSN.2011.043666](https://doi.org/10.1504/IJSN.2011.043666).
13. Narayan, S., Gagné, M., and Safavi-Naini, R., Privacy preserving ehr system using attribute-based infrastructure. In: Proceedings of the 2010 ACM Workshop on Cloud Computing Security Workshop, CCSW '10, pp. 47–52. ACM, New York, NY, USA, 2010. doi:[10.1145/1866835.1866845](https://doi.org/10.1145/1866835.1866845).
14. Chen, L., and Hoang, D. B., Novel data protection model in healthcare cloud. In: 2011 IEEE 13th International Conference on High Performance Computing and Communications (HPCC), pp. 550–555, 2011. doi:[10.1109/HPCC.2011.148](https://doi.org/10.1109/HPCC.2011.148).
15. Zhang, Y., Grorec: A group-centric intelligent recommender system integrating social, mobile and big data technologies. *IEEE Trans. Serv. Comput.* 99:1–1, 2016.
16. Bentley, J. L., Multidimensional binary search trees used for associative searching. *Commun. ACM* 18(9):509–517, 1975. doi:[10.1145/361002.361007](https://doi.org/10.1145/361002.361007).
17. Bereczky, N., Duch, A., Németh, K., and Roura, S., Quadkd trees. *Theor. Comput. Sci.* 616(C):126–140, 2016. doi:[10.1016/j.tcs.2015.12.030](https://doi.org/10.1016/j.tcs.2015.12.030).
18. Guttman, A., R-trees: A dynamic index structure for spatial searching. In: Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data, SIGMOD '84, pp. 47–57. ACM, New York, NY, USA, 1984. doi:[10.1145/602259.602266](https://doi.org/10.1145/602259.602266).
19. Jin, P., Xie, X., Wang, N., and Yue, L., Optimizing r-tree for flash memory. *Expert Syst. Appl.* 42(10):4676–4686, 2015. doi:[10.1016/j.eswa.2015.01.011](https://doi.org/10.1016/j.eswa.2015.01.011).
20. Zyskind, G., Nathan, O., and Pentland, A., Enigma: Decentralized computation platform with guaranteed privacy. *CoRR abs/1506.03471*, 2015.
21. Chaum, D., Crépeau, C., and Damgard, I., Multiparty unconditionally secure protocols. In: Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing, STOC '88, pp. 11–19. ACM, New York, NY, USA, 1988. doi:[10.1145/62212.62214](https://doi.org/10.1145/62212.62214).
22. Goldreich, O., Micali, S., and Wigderson, A., How to play any mental game. In: Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing, STOC '87, pp. 218–229. ACM, New York, NY, USA, 1987. doi:[10.1145/28395.28420](https://doi.org/10.1145/28395.28420).