Medical Record System Using Blockchain, Big Data and Tokenization

Paul Tak Shing Liu[1,2(✉)]

[1] Social Mind Analytics (Research and Technology) Limited, Hong Kong, China
ogmaster2011@gmail.com
[2] Maximus Consulting (Hong Kong) Limited, Hong Kong, China

Abstract. This paper will discuss the major aspects of medical records, blockchain and big data. In turn, it will discuss the advantage and disadvantage of using blockchain on medical records storage and retrieval. It will also discuss the alternatives of using blockchain and big data techniques. Different aspects of medical records will be investigated briefly: (1) integrity, (2) viewing control, (3) viewing approval, (4) western medicine and chinese medicine practice, (5) storage size and duration, (6) deletion and purge, (7) file format conversion, (8) data migration, (9) report interpretation, etc. Characteristics of blockchain and big data analytics will be explored briefly with description. A conclusion will summarize the approaches. References will be provided for further research and investigation.

Keywords: Medical record · Blockchain · Big data analytic · Tokenization

1 Introduction

Medical Record. To improve our medical service for our patients around the world we must have a readily available standardized medical record history any time anywhere. We must also overcome the following challenges: (1) record type, (2) storage, (3) medical records and reports, (4) ownership and reading right, (5) data security, (6) western medical practice and Chinese medical practice, (7) patient's right of knowledge.

Medical record types include text reports, word documents, images, videos and data. They are all important in our medical history. They are all in different formats with possible free style.

Central storage is important too. It is because many doctors may not have their own or may have different medical record systems. Record formats may pose a challenge sometimes. Nevertheless, many formats can now be converted from one to another or vice versa. Yet another problem, forever increasing storage due to forever increasing number of reports/files is caused by the continuous contribution of reports or records from medical doctors or patients. All old reports or records must be archived regularly and must be transferred to new storage media from time to time.

K.-Y. Lam et al. (Eds.): ICICS 2016, LNCS 9977, pp. 254–261, 2016.
DOI: 10.1007/978-3-319-50011-9_20

Medical test result interpretation, diagnosis, treatments and recovering progress are medical reports prepared by medical professionals or medical doctors. Test result interpretation will be given by professional medical personnel in free style writing. However, those interpretations may be in different presentation styles and in free style writing. Similarly, diagnostics will be given in writing by medical doctor. They involve free style writing and can be difficult to be interpreted by both machine and patient. Medical treatments will be given in writing by medical doctor. They involve free style writing too and can be very different from doctor to doctor. Different doctor may also use different dosages for similar treatment. Recovering progress of patient will be observed and recorded in writing by medical doctor. Those progress reports too involve free style writing and are difficult to be interpreted by machine and patient. Also, extra patient data and recovering progress can be contributed and inputted by patient through web interface from time to time into medical record system. These extra data and reports are vital to every patient's recovering process and are important to medical doctor's diagnosis. Nonetheless, patients may use very different wordings and very free style writing too.

Medical record ownership and record sharing among doctors is extremely controversy within medical community. Many doctors worry about possible lawsuit initiation by others or even by patient if things go wrong. Medical related legal matter or lawsuit may incur from time to time. Therefore, all documents generated must be properly stored and must be produced to court in whole if requested. All records must not be tampered. All records must not be deleted for whatever reason.

Data security is important. It is necessary to protect the privacy of both patient and medical doctor. All kinds of IT security measures must be implemented for data protection. Many doctor's medical reports actually need some kind of data security. For example, IT security, backup, data recovery, etc. All these could jack up their clinic operation cost.

Patient's right of reading reports can be very tricky. Since all medical reports are owned by medical doctors, medical doctor has the right not to share the medical reports to anyone, including patients. All accesses must require the consent from the report owner which is the doctor. Report reading right may also induce lawsuit which is most worried by medical doctor. Report reading right for other medical personnel can be conditionally limited by report owner. It all depends on who will read the medical report. That medical personnel must require pre-arranged consent from the report owner. Can patient and/or other medical personnel view the patient's medical history without directly accessing the medical reports owned by other doctors? This is very important in a way that patient's conditions can be properly accessed without the jeopardy of delay due to the tedious requesting process of consent from other doctors.

Western medical doctor and Chinese medical doctor are very much different in terms of practice and report writing. Their reports must require proper translation back and forth. Simply translating Chinese into English is definitely not enough. It also need certain qualitative translation on medical terms.

If a medical doctor retires, what can we do with all the medical reports owned by that specific doctor? It is a tricky question. Besides, that specific medical doctor may want to release all medical reports with conditions. However, that doctor may want to

retain the secrecy of those reports. On the other hand, other doctors may not want to take up the responsibility of all those medical reports in case of possible lawsuits.

Finally, patients should have right to know who has viewed the related reports about specific patient. This is to avoid unauthorized access to patient's privacy. At the same time, patient should be able to report unauthorized access to his/her own medical history.

HK eHealth. There are few other criteria imposed by ehealth system (eHR) by Hong Kong government [1]. Those include (1) government-led, (2) data privacy and security, (3) not compulsory, (4) open and pre-defined common standards, (5) building block approach. Every record access of patient's record will be recorded and notification will be sent to corresponding patient. The record data are very restricted, including lab tests, x-ray, etc. Medical doctor's reports may not be there. All authorized medical personnel may read patient's records if corresponding patient agrees.

Blockchain. In electronic currency world, blockchain has been very popular for bitcoin transaction settlement as a public ledger [2]. However, blockchain itself has its own limitation and criteria. They are summarized as followed.

Blockchain can prevent duplicated transactions by accepting the very first arrived transaction for specific unspent electronic currency and rejecting other transactions for that unspent currency immediately thereafter. The inverse hash calculation used by the proof of work will reinforce such status for the long run. This is the heart of blockchain. It can then be extended to prevent duplication of work. Duplication of work can be avoided and can be detected immediately through the digital signature of the duplicated records. However, it is not possible, through digital signature, to detect similar, but different, records or files.

The proof of work uses the inverse calculation of hash value [3, 4]. The inverse hash calculation is difficult and the direct hash calculation is simple. Blockchain then forces transaction settlement or any electronic work to include the inverse hash calculation. Then direct hash calculation will be used to verify that the transaction settlement work has been completed.

Blockchain can prevent the deletion of work. Deletion or tampering of work will be notified in the chain if the direct calculation of the hash value of specific block is inconsistent.

Blockchain cannot reverse the work or transaction settlement. The reverse of work is not possible because it is not possible to recalculate the hash value of specific individual block without the recalculation of all hash values of all other blocks thereafter. However, unfortunately, this part of the whole protocol is mostly hacked since bitcoin's birth.

The guarantee of the existence of data records or files is not easy. If only digital signature is used, it cannot guarantee the existence of records or files. If the original piece of document is not within a specific block, there is no way to guarantee the existence of such document within specific system or internet or world wide web.

Big Data Analytics. Big data, nowadays, can achieve many things using artificial intelligence techniques. Popularity analytic tells which item or group of items is most popular from the data sources. With popularity analytic, we can detect similarity or

closeness among topics within large volume of data. Groups of items (or topics) can be system generated using closeness relationship formulation. Or they can be assigned by user. Sentiment analytic tells the positive or negative feeling towards specific item or topic. It can also tell the percentage towards positive feeling and the percentage towards negative feeling. Aspect mining tells the characteristics of specific item. Both popularity and sentiment analytics can be performed on specific aspect. Prediction of future public behavior towards specific item, or person can be performed in the near future. Strategy and action suggestions for specific item or person can be formulated thereafter.

Tokenization. Tokenization is very important in this respect of medical reports sharing. Token will not depend on operating systems and does not include content within. The storage server can examine the validity of the token before sending out the required report(s). The report owner can sign the request token sent by specific viewer with confidence.

2 Development Approaches

The following proposed features summarize what can be implemented to overcome the challenges described in introduction section for the development of a reasonably sound and safe medical record system that will benefit most patients.

Medical record types, formats and storage are the first few challenges in the whole development cycle. They can be resolved by today's technology without much challenge. Whether we use off the shelf ready-made software or customer-designed system, it is never a challenge right now. It is always good to have a centralized system with backup in different locations for data storage, new data insertion or instant data record reading. Different channels can be established for accepting documents or multi-media files. Standardization can be achieved by converting all files into a couple reading formats.

All medical tests, test interpretation, diagnosis, treatment and recovering progress reports are important data records or files. A general medical report system can utilize different channels for accepting the data or files. A web based good looking user interface can be developed for such purpose. It can upload individual files or a batch of files. As for the data or progress reports contributed by patients can be inputed through web based interfacing portal.

All uploaded data files and records can now be saved to a centralized storage area or database for future retrieval. All records or files must be digitally signed by owners/contributors and must designate the possible viewing counterparties. Each record will be used later for either patient's view, or original medical doctor's review, or be reviewed by other medical doctors for further diagnosis and treatment, or be retrieved by court for judicial process.

Next, we should consider measure to reduce the risk of purge and modification unintentionally or intentionally. We can implement blockchain mechanism directly on the medical reports.

Database, backup and duplicated hardware have been used historically to protect data from the risk of tampering. Hash value and digital signature or blind signature of each record can further reinforce protection tampering [5].

Blockchain mechanism can be implemented for all the encrypted medical data records and files [6–8]. Encryption of medical record or file is needed for further protection from unauthorized access. In turn, blind signature can be used for integrity check of the original medical record or file. In order that each record or file is not being deleted or tampered, each encrypted record or file should be included serially inside the blockchain implemented. At the same time, the blind signatures should also be included for the proof of genuine of each record. Many proposed medical record blockchain implementations only include the digital signature of each record or file. The use of digital signature can avoid the record or file being tampered. However, it cannot avoid the deletion of specific file or record without any notice from system. A modified blockchain can be implemented to further enhance the integrity of all the records/files encapsulated inside the chain. A modified blockchain can include randomly selected hash values of blocks from the chain. This can constantly check the integrity of each record/file from the chain. In turn, it ensures the existence of each record/file in the long run [9].

Data security can be implemented with ease using state of art IT security and control technology. Tokenization can be utilized to send reading authorization to whoever medical personnel who needs reading specific report. Token can contain specific authorization command without directly containing specific record/file.

Forever increasing number of records or files can be a headache in the long run to storage, operation and archiving. To copy the whole blockchain from old storage device to new storage device is not difficult but tedious. However, if any of the file format is going to be outdated and needs conversion, it is going to be chaos. Not only the file needs conversion, but also the digital signatures and hash values need recalculation too. To change or delete record/file, one can neither delete nor purge specific record/file in blockchain. This is one of the underlying criteria of blockchain technology. The owner can only invalidate the old record/file and add a new one to overlay on top of the old record/file which cannot be deleted or purged. However, the old record/file will be there forever. Similar to bitcoin transaction settlement, old transaction will be there forever and new transaction will replace the old one. To update the old record is similar to spending specific electronic currency note. Instead, owner must spend the old record and create a new record with different format. From storage point of view, records/files must be transferred to new storage from time to time. Old records/files must be converted to new reading formats regularly. In turn, both hash values and digital signatures must be regenerated and the old records/files must be invalidated at some point of time regularly. This could further increase the overall size of the blockchain exponentially. Therefore, we must decide at the very beginning that we have to use certain reading formats forever. Besides, we may also need to create a separate blockchain to facilitate migration purpose.

Although forever increasing number of records poses a big challenge, there may be an alternative design [10]. Blockchain itself cannot be deleted partly. The whole thing is built forever. We can only delete or purge the whole chain at the same time. If we want to delete or purge some records or files or reports in the near future, we must

design it at the very beginning. Since patients will die one day, after certain period of time, the passed away patients' records should be cleared forever or archived in another storage for later retrieval for research. One possibility is to build many blockchains for different groups of patients. The grouping can base on age, or geographic location, or hospitals/clinics, etc. Then, at certain point of time, when all patients within one blockchain are deceased, we can remove or archive the whole blockchain directly without hesitation. Otherwise, we must be prepared for forever increasing storage in the long run.

Reading right for patient requires either a pre-written report for patient only or a universal standard interpretation of all reports written by medical doctors. Since all reports are written by human medical doctors, all the wordings can be very much different for the same thing. Therefore, one possibility is to use big data technique to re-interpret all reports for patients. The actual original reports must have original medical doctor's signed consent before being released to patients due to lawsuit concern from medical doctors. However, big data analytic can give a very brief statistical summary of each report of specific patient. Big data technique can also display the patients' medical conditions over a time line in a graphical view. This is essentially important for patients and other medical doctors in the long run. Other medical doctors may need such systematic time line for diagnosis and treatment.

For big data analytic to work properly, all records must not be encrypted. Therefore, it is important to strip off the identity of patient from medical record or file before use. A separate database for big data analytic is needed for further isolation. However, we can still use certain serial number for backward association with specific patient if needed.

For retiring medical doctors, medical report ownership can be assigned to another professional or to the system for further reading right management. Although it may be not preferable from patients' point of view, it is a second best option. Many other arrangements can be developed thereafter.

Translation from Chinese medical doctor's report into western medical doctor's reading style, or vice versa, is a very difficult challenge to be solved. It is that the underlying philosophy of both western medical doctor and Chinese doctor are very much different. It is necessary to do a thorough research on terms interpretation to have an accurate translation back and forth. However, if there are reports for similar diagnosis and treatment coming from both western medical doctors and Chinese medical doctors, big data can still make an approximate association of terms coming from both sides. Then it may be possible to make translation using big data analytics. The more reports the system has, the more accurate association it can make. It will be ongoing research and tedious work. However, big data will provide an immediate viable approach to such difficulty.

Although it is possible to detect duplicated pieces of work, it is not so easy to prevent similar, but different, pieces of work. With digital signature of every piece of document, we can easily detect duplication. However, even a slight change in specific document, a different digital signature will be generated and cannot be associated with the original document. Fortunately, we can use big data analytic to detect similarity between two documents in a large volume of works. Although big data analytic is not fool proof technique, its approximation is more than enough.

Blockchain is a very good choice for present and previous works by others which concentrate on the integrity of medical reports without worrying about many other factors. However, in the long run, blockchain related application development will hit a solid wall very soon when number of medical data records or files hits a larger volume. They will need to fix a lot of challenges mentioned in previous section. It is a lot better to prepare for the worst cases to come. Blockchain can protect the integrity. However, it cannot guarantee the existence or the interpretation of the reports. It cannot control the viewing right. It cannot solve the challenges of outdating formats. It cannot handle forever increasing storage size. It cannot handle deletion or purge. Blockchain is not everything. We need other technologies to supplement any design.

Hong Kong government has its own electronic health system (eHR). The system philosophically is similar. However, the data types are not comprehensive enough. It does not consider other important factors. It may not be very useful for most patients and medical doctors. Technologically, eHR does not use blockchain technology and is a centralized system. Patients cannot contribute their data to the system either. The system eHR is still a very primitive design. It needs much further improvement.

Finally, the system should implement a reading log on all patient's medical records. This log can be a database table or can also be a blockchain which will avoid log record tampering or deletion.

3 Conclusion

In this paper, the actual design of sharing medical records system is explored. Every aspect of medical records is studied briefly. The characteristics of blockchain are explained on its pros and cons. Big data analytic has its own capabilities and is explained briefly in introduction. Medical records sharing system's integrity, deletion and purge, record sharing, storage, format conversion, data migration, interpretation and standardization are examined. Different approaches to overcome all human related challenges are suggested in this paper with confidence. The use of blockchain is very important. However, other technologies, big data analytics and tokenization, are very much needed to supplement the original design. Although actual implementation is not put into production, the approaches should be practical enough. Each suggestion gives some details and reasons behind the technical direction. Hope that this will stimulate further research and development for the benefit of patients as well as general medical community.

References

1. Hong Kong Government. Electronic Health Record Sharing System (2016). http://www.ehealth.gov.hk/en/home/index.html
2. Franco, P.: Understanding Bitcoin – Cryptography Engineering and Economics. Wiley, Hoboken (2015)

3. Stallings, W.: Cryptography and Network Security – Principles and Practice, 6th edn. Pearson, London (2014)
4. Milne, J.S.: Elliptic Curves, Kea Books (2006)
5. Chaum, D.: Blind signatures for untraceable payments. In: Advances in Cryptology: Proceedings of CRYPTO 1982 (1982)
6. Forde, B.: MedRec: Electronic Medical Records on the Blockchain, 2 July 2016. https://www.pubpub.org/pub/medrec
7. Everington, J.: Du to use blockchain for health records, 30 May 2016. http://www.thenational.ae/business/technology/du-to-use-blockchain-for-health-records
8. Scott, M.: The Future of Medical Records: Two Blockchain Experts Weigh In, 1 May 2016. https://btcmanager.com/news/the-future-of-medical-records-two-blockchain-experts-weigh-in/
9. Ekblaw, A., Azaria, A., Vieira, T., Lippman, A.: MedRec: Medical Data Management on the Blockchain, 11 August 2016. https://medium.com/mit-media-lab-digital-currency-initiative/medrec-electronic-medical-records-on-the-blockchain-c2d7e1bc7d09#.wj7tr2fvq
10. Yessi Bello Perez (@yessi_kbello). Medical Records Project Wins Top Prize at Blockchain Hackathon, 10 November 2015. http://www.coindesk.com/medvault-wins-e5000-at-deloitte-sponsored-blockchain-hackathon/