MEDINFO 2017: Precision Healthcare through Informatics
A.V. Gundlapalli et al. (Eds.)

doi:10.3233/978-1-61499-830-3-45

Enabling Patient Control of Personal Electronic Health Records Through Distributed Ledger Technology

James Cunningham and John Ainsworth

The Health eResearch Centre and The Farr Institute, Division of Informatics, Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester

Abstract

The rise of distributed ledger technology, initiated and exemplified by the Bitcoin blockchain, is having an increasing impact on information technology environments in which there is an emphasis on trust and security. Management of electronic health records, where both conformation to legislative regulations and maintenance of public trust are paramount, is an area where the impact of these new technologies may be particularly beneficial. We present a system that enables fine-grained personalized control of third-party access to patients' electronic health records, allowing individuals to specify when and how their records are accessed for research purposes. The use of the smart contract based Ethereum blockchain technology to implement this system allows it to operate in a verifiably secure, trustless, and openly auditable environment, features crucial to health information systems moving forward.

Keywords:

Electronic Health Records; Health Information Systems; Technology

Introduction

Medical data, both routinely collected and specifically studied, is increasingly being recorded, represented, and stored electronically [1]. Having access to these resources in electronic form is immensely beneficial to researchers, allowing novel research methods to be applied over large volumes of data in a way that would not have been possible even a decade ago [2, 3]. This switch to the digital does, however, bring with it problems relating to the physical and ethical security of medical data [4]. Data stored electronically is much easier to copy, distribute, and mine for confidential information. Breaches of security and the loss or misappropriation of data negatively impact the public perception of medical research and threaten to bring down regulatory restrictions that will prevent or hinder future research [5]. The current state of information security within the medical informatics domain makes controlling and identifying such breaches difficult [6].

Further, there is a gap between the ideal needs of the research community and ethico-legal restrictions on the use of personal medical record data [7]. The research community's desire for essentially unfettered access to data is checked by the legal responsibility of data owners to guarantee the privacy of the patients whose data they protect. There is also a recognized need for public involvement in the research process [8] and even where patient data is de-identified and used in the aggregate, granting of explicit consent for data use has been identified as the preferred model for half of all patients, with increased awareness of Electronic Health Records (EHRs) impacting positively on willingness to consent to research [9]. The best way of achieving and maintaining this balance between trust, security and admittance of public participation is an open question [10] — how can medical data be shared in a way which still, at its core, guarantees as far as possible the security and integrity of the data being shared? Addressing these challenges will be of crucial importance to health informatics moving forwards.

We present a proof-of-concept system for enabling fine-grained specification of access control policies as pertaining to third-party access to electronic health records on an individual level, which goes some of the way towards tackling these issues. Giving patients control of access to their own records whilst giving research organizations, possibly from the private commercial sector, the ability to directly reach out to patients and request access to medical data, opens up a series of issues around trust and security. We have implemented this system using the Ethereum platform, a modern, smart contract based, distributed ledger system [11]. This choice of platform not only allows for a natural expression of a solution to the problem we address but also explicitly addresses and solves the underlying complex issues of trust and security.

Methods

Enabling direct patient involvement in controlling the use of medical data, and doing so in open and secure manner will enhance both uptake and acceptance of medical informatics platforms aimed at enabling access to research data. We aimed to develop a core Application Programming Interface (API), for use as a service within a wider platform, that would enable a permission system through which patients could both specify who could access their records and to review the uses to which their data have been put. We identified this as the most fundamental building block of any system that would purport to enhance patient control in a research-oriented informatics system.

The requirements for the design of the API were gathered through an analysis existing systems for sharing and reuse of medical data deployed within the North of England [12]. This gave us an overview of the core functionality required by such a system.

The design of the API proceeded first with the identification of the key class of actors who would interact with the system. A series of core requirements based around ensuring requisite levels of security, trust environment, and transparency were then developed. Following this development, a series of use cases were produced that specified the ways in which the system actors would be able to interact with the API in order to achieve the overarching goal of enabling patient control of access requests to EHR data.

In implementing the system to address these use cases we chose distributed ledger technology ('DLT') as the underlying technical implementation. DLT is the mechanism, in terms of

data structures and associated computational methods, underlying the Bitcoin cryptocurrency, in which the specific instance of DLT is the Blockchain [13]. The driving use case behind the design of the Blockchain was the desire to allow for a decentralised transaction ledger; the provision of a canonical global account balance for all holders of the Bitcoin currency, without reliance on a trusted third party as a point of control [14]. The Blockchain implementation further provides both a public key–based infrastructure for account identification and control, and a 'mining' mechanism (a computational competition which both incentivizes hosting of the peer-to-peer Bitcoin network and solves the so-called 'Byzantine Generals problem' — that of guaranteeing consensus amongst distributed network nodes containing potential bad actors) [15]. Since its conceptualisation DLT found extended use across a range of application areas [16] and has been extended with the addition of 'smart contracts', a means of adding distributed computational processing to the underlying transactional ledger [17].

The core features of DLT were identified to meet the specified system requirements. Specifically, we chose the Ethereum Platform [11] to host the API given an assessment of its capabilities in producing a system to fulfil the outlined use cases. Ethereum itself provided both the core DLT blockchain capabilities allowing for hosting within a trustless, secure environment, and a smart-contract implementation allowing for the programmatic implementation of the API directly on the platform itself.

Results

Presented in the following section are the results of our design and implementation in terms of the actors we identified, the core requirements, the driving use cases and a specification of the API.

The core actors within the system were identified as follows:

- **Public Participant** — someone for whom associated EHRs are accessible within some system who will be granted control of a permission system for accessing those records.
- **Research Organization** — Representing an organizational entity that wishes to request access to EHR data for research purposes.
- **Data Custodian** — An organizational entity with ultimate control of source electronic medical data

We identified the following underlying requirements for the system:

- **Trustless** — Reliance on a third party to maintain control of some or all of the system functionality decreases trust and adds a single point of failure to the system.
- **Incentivized** — With use of a distributed technology, it is required that participants within the system should share the burden of hosting the system itself. A mechanism is needed to ensure this happens.
- **Secure** — The system needs to be secure in that it must prevent actions being performed by entities or actors not specified as being allowed to perform those actions.
- **Identifiable** — Actors need to be strongly and verifiably identifiable within the system. Recognizing that key management by the lay public is a difficult issue, formal identity management may be delegated to other actors (for example Data Custodians performing actions within the system on behalf of users)
- **Transparent** — Perceived trust in the system is crucial so all actions that take place within the system need to be publicly auditable.

The following use cases were identified as the drivers of the API design:

- A **Research Organization** can publish a request for data in the form of a **Research Proposal**. The **Research Proposal** will outline what medical data is required for the research, the form of that data and the limits of its use.
- A **Data Custodian** can vet a **Research Proposal** and decide whether it is published to **Public Participants** whose data that custodian safeguards.
- A **Public Participant** can set a general preference for how their medical data should be used. This will allow options for allowing or denying all requests, or granting permissions on a proposal-by-proposal basis.
- A **Public Participant** can view lists of **Research Proposals** which would utilize their private medical data.
- A **Public Participant** can specify an option against a **Research Proposal** indicating whether or not they will grant permission for the use of their data within that proposal.
- A **Data Custodian** can request a list of patients who have consented for a particular **Research Proposal**

We chose to implement the code for enacting these use cases using DLT. Through creating a distributed ledger instance, records of what proposals have been offered and which participants have granted, either implicitly through global options, or explicitly against a particular proposal, are recorded within the 'blocks' that form the ledger. These use cases were then implemented directly in the form of smart contracts deployed within the blockchain instance. A smart contract is an executable piece of code that references the current state of the ledger and can write back to it. Smart contracts are authored in Solidity, a Turing complete programming language, broadly similar in syntax and structure to javascript. These were compiled into Ethereum compatible byte-code and hosted on the ledger instance. An illustrative extract of such a smart contract is given below.

```
contract Participant{

    address admin;

    function Participant  (address owner_address) public {
        owner = owner_address;
    }

  function Owner () constant returns (bool confirmation){
    return msg.sender == owner;
  }

  enum Option { Grant, Deny }

  struct Participant {
    uint identifier; // every participant has an id number
    address participant_addr;
    address id_custodian; // the id of this patient's corresponding data custodian
    string record;
    uint status_;
    uint[] institution;
  }

  ...
}
```

Figure 1– A Portion of a Smart Contract

Accounts for actors within the system (patients, research organizations and data custodians) were either created directly as user accounts within the Ethereum system (with associated public/private key pairs), or, in the case of patient accounts,

managed via a data custodian account -- that is, public participants controlled their accounts via functions exposed by the data custodian entity. This mitigated the difficulty of requiring public participants in the system to manage their own public/private key pairs.

The hosting environment for the API consisted of a private instantiation of the Ethereum platform -- one entirely distinct from the canonical, publicly accessible peer-to-peer instance. Further, node hosts were set up inside in virtual machines and firewall rules put in place to restrict peer-to-peer connections to other known hosts. Access to the system in its entirety is then ultimately controlled through network-level security.

Discussion

We chose distributed ledger technology as the underlying technical implementation for this service as it provided, as an integral part of the technology itself, a secure framework that enables deployment within a network not reliant on a central point of trust. As stated, establishing a balance between the desires of the research community, the ethical and legal obligations of data custodians, and the ultimately decisive needs as well as the perceptions of the public is of key importance in establishing sustainable medical informatics frameworks enabling the use of electronic health data. The nature of distributed ledger technology, in that it provides an auditable, accountable framework removed from a single point of trust (and hence possible failure or compromise), makes it a natural fit for such systems. Identity management is a fundamental requirement of any such system — inability to establish and ensure identity across a system immediately invalidates any claim to security. Again DLT is built around the concept of identity and public key management, providing a base on which to build systems where identity is a crucial component. Smart contract functionality is provided by the Ethereum platform.

The blockchain data structure is designed for redundancy across a network of peers and smart contract mechanisms deliberately replicate computation of programatic structures across all nodes in the network. Whilst this provides benefit, as mentioned, in the areas of trust and security it does introduce computational inefficiencies. These raise potential issues in terms of the use of the technology in a real-world setting, particularly one at scale. However, the use cases that the design of the system was based upon do not require absolute real-time instantiation across the system. The preferences set by users and the requests for research data do not need immediate processing or acting on by external services. As such this mitigates to a degree the natural inefficiencies in the system. The reliance of current distributed ledger technology on essentially arbitrary computation (token mining) again introduces inefficiencies into the system. The original aim of the mining mechanism in the Bitcoin scheme — that of incentivizing distributed participation and as a means of introducing and distributed currency — does not translate entirely into an environment with a higher degree of trust between participants. Such mechanisms can still be leveraged as a means of ensuring fair participation in the ecosystem; for example with tokens, rather than providing economic value, it can be used as a proof-of-participation and required for ongoing use of the system.

Distributed ledger technology, in particular Ethereum, are new and hence relatively untested technologies. Whilst their primary use case as financial tools with intrinsic economic value incentivizes both secure implementation and testing through real world exposure, bugs and security flaws are do persist. Further, the rapid pace of innovation and evolution of the platforms will carry increased risks of potentially harmful design flaws becoming apparent.

Constant evaluation and auditing of the underlying technologies in terms of monitoring and addressing security flaws will therefore be crucial in the future use of this technology. the current system we have developed exists as a stand-alone API with the explicit design goal of acting as a component in a wider service oriented medical informatics platform. Future work will address this as we aim to roll out a test deployment of the system within an infrastructure for enabling actual access to data within a real-world setting. Also, given the potential issues surrounding public perception of distributed ledger technology, both in respect to it being a nascent and relatively untested technology, and its association with negative reporting and use within black markets, work will be done on gauging public attitude towards the use of the technology within a healthcare environment.

For any medical informatics system the key driver of its implementation, and the ultimate gauge of its success, will be measured in the benefit it brings to medical care and practice. In moving forwards the state of the art in terms of security, privacy and accountability, distributed ledger technology has the potential to add a significant degree of trust to the functionality of medical informatics systems. With this enhanced trust will come the ability to utilize medical data within richer settings and with a wider range of participants — something that will ultimately improve medical care and practice moving forwards.

Conclusion

We have identified a series of key requirements for enabling enhanced patient control of EHRs and have developed a proof-of-concept API meeting these requirements. Distributed ledger technology was chosen specifically to meet what we see as the fundamental issues in health informatics, namely trust and security. Whilst problems do exist, particularly with respect to the fundamental inefficiency and current immaturity of the technology, distributed ledger technologies offer unique solutions in the health informatics domain and will inevitably see increasing use in the field in future.

References

[1] A. K. Jha, D. Doolan, D. Grandt, T. Scott, and D. W. Bates, "The use of health information technology in seven nations," *International Journal of Medical Informatics*, vol. 77, no. 12, pp. 848 – 854, 2008.

[2] T. Murdoch and A. Detsky, "The inevitable application of big data to health care," *JAMA*, vol. 309, no. 13, pp. 1351–1352, 2013.

[3] R. Bellazzi, "Big data and biomedical informatics: a challenging opportunity," *Yearbook of medical informatics*, vol. 9, no. 1, p. 8, 2014.

[4] M. Wiesenauer, C. Johner, and R. Ro ̈hrig, "Secondary use of clinical data in healthcare providers-an overview on research, regulatory and ethical requirements," *Stud Health Technol Inform*, vol. 180, pp. 614–8, 2012.

[5] H. K. Patil and R. Seshadri, "Big data security and privacy issues in healthcare," in *2014 IEEE international congress on big data*, pp. 762–765, IEEE, 2014.

[6] T. Sahama, L. Simpson, and B. Lane, "Security and privacy in ehealth: Is it possible?," in *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on*, pp. 249–253, IEEE, 2013.

[7] G. Chatellier, V. Varlet, C. Blachier-Poisson, *et al.*, "big data and open data: What kind of access should researchers enjoy?," *The ́rapie*, vol. 71, no. 1, pp. 107–114, 2016.

[8] L. Doyal, "Informed consent in medical research. journals should not publish research to which patients have not given fully informed consent–with three exceptions.," *BMJ: British Medical Journal*, vol. 314, no. 7087, p. 1107, 1997.

[9] F.Riordan,C.Papoutsi,J.E.Reed,C.Marston, D. Bell, and A. Majeed, "Patient and public attitudes towards informed consent models and levels of awareness of electronic health records in the UK," *International Journal of Medical Informatics*, vol. 84, no. 4, pp. 237 – 247, 2015.

[10] N. C. Lea, J. Nicholls, C. Dobbs, N. Sethi, J. Cunningham, J. Ainsworth, M. Heaven, T. Peacock, A. Peacock, K. Jones, *et al.*, "Data safe havens and trust: Toward a common understanding of trusted research platforms for governing secure and ethical health research," *JMIR Medical Informatics*, vol. 4, no. 2, 2016.
[11] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, 2014.
[12] J. Ainsworth and I. Buchan. "Combining health data uses to ignite health system learning. " *Methods Inf Med* 54.6 pp. 479-87, 2015.
[13] R. Ali, J. Barrdear, R. Clews, and J. Southgate, "Innovations in payment technologies and the emergence of digital currencies," *Bank of England Quarterly Bulletin*, p. Q3, 2014.
[14] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system." https://bitcoin.org/ bitcoin.pdf, 2008. [Online; accessed 28Nov-2016].
[15] A. Miller and J. J. LaViola Jr, "Anonymous byzantine consensus from moderately-hard puzzles: A model for bitcoin." http://nakamotoinstitute.org/research/anonymous-byzantine-consensus, 2014. [Online; accessed 28-Nov-2016].
[16] M. Pilkington, "Blockchain technology: principles and applications," *Research Handbook on Digital Transformations, edited by F. Xavier Olleros and Majlinda Zhegu. Edward Elgar*, 2016.
[17] S. Omohundro, "Cryptocurrencies, smart contracts, and artificial intelligence," *AI matters*, vol. 1, no. 2, pp. 19–21, 2014.

Address for correspondence

James Cunningham, james.a.cunningham@manchester.ac.uk