Public Health and Informatics
J. Mantas et al. (Eds.)

A Blockchain-Based Healthcare Platform for Secure Personalised Data Sharing

Juliana BOWLES[a,1], Thais WEBBER[a], Euan BLACKLEDGE[b] and Andreas VERMEULEN[a,b]
[a] *University of St Andrews, School of Computer Science, UK*
[b] *Sopra Steria, Edinburgh, UK*

Abstract. To facilitate personalised healthcare provision across Europe, we envision solutions that enable the secure integration and sharing of medical health records. These solutions should address privacy concerns, such as granular access control to personal data, establishing what should be accessible when and by whom, whilst complying with collective regulatory frameworks such as the European General Data Protection Regulation (GDPR) and adhering to international standards on how to manage information security. The proposed healthcare system design integrates technologies such as blockchain and scalable data lakes with adequate system routines to guarantee the secure access of confidential data. In this paper, we present the essential architectural components for the secure integration of medical records in a blockchain-based platform. We present a patient-centric data retrieval approach which incorporates a structured format to compose access rules.

Keywords. Healthcare system, medical records, blockchain, data lake, access rules

1. Introduction

Healthcare systems must adapt to modern digital transformation practices [1],[2]. Seamless integration of medical records and patient experience form the basis of improved health and social care services [3]. In Europe, an expectation to facilitate cross-border transfer of medical records and enhance healthcare provision would require robust guarantees of trust in the healthcare data sharing system [4]. On the one side it must allow patients to define their own data-sharing agreements, and on the other side it must comply with multiple, possibly conflicting, legislative frameworks and healthcare policies, which in the European Union (EU) include the General Data Protection Regulation (GDPR) [5].

Recent work on the subject of personal data management and regulations [4],[5] raised concerns to improve healthcare provision in European countries under GDPR. Several emerging healthcare platforms focus on the integration of next-generation technologies such as Blockchain and Big Data solutions altogether [1],[2], empowering patients to control personal data and customise access rules. The ultimate goal is the improvement of end-user experience, with adequate measures to address security and privacy concerns relating to integrated medical records that will be accessible anywhere, at anytime.

[1] Corresponding Author: Juliana Bowles, School of Computer Science, University of St Andrews, KY16 9SX, St Andrews, UK. E-mail: jkfb@st-andrews.ac.uk.

Our proposed platform [3],[6],[7] known as the *Serums Smart Health Centre System* (SHCS), comprises of two key components: a *hyperledger fabric blockchain* utilising smart contracts and *individual data lakes* for each healthcare provider. Smart contracts allow for business logic to be stored on a distributed network. This ensures that any privacy or data sharing rules, which have been put in place by either patients or healthcare providers, remain consistent regardless of where the data is accessed from. Furthermore, in a distributed network we avoid a single rule controller and allow for redundancy across the network to address possible temporary connectivity failures. Each data lake follows a precisely defined schema with clearly defined boundaries between the processing layers that have been applied to the medical data sets [3]. Their primary role is to format the data, at time of request, into a data vault structure, allowing for multiple sets of distinct data to be joined into a single record known as *Smart Patient Health Record* (SPHR) [7].

In this paper, we present the overall SHCS design which incorporates a blockchain solution to enable secure and recorded access to medical records. Our focus is on advancing the pursuit towards a secure patient-centric solution for personalised sharing of heterogeneous health data. The processes and context of data deployment within the SHCS are dynamically designed to allow for compliance with GDPR, both in its current form and henceforth, to withstand the non-static nature of the governmental regulations over time and adhere to international standards for privilege management and access control.

2. Access Control Design

Our healthcare platform [6],[7] aims to give patients the means to share their medical data with an oversight of who has access to a particular subset of their data and for how long. We are designing the SHCS system in such a way that it balances usability with strict adherence to international and national laws, as well as healthcare standards and individual care protocols that each healthcare provider has in place for data protection. The benefit for patients, likely to increase their trust in our developed solution, lies in facilitating their understanding of who has access to different parts of their data, and how they are able to control such access by explicitly allowing/disallowing data access. To achieve this goal, the platform incorporates three core components: data tags (on data stored in the data lake), rules and the blockchain.

Figure 1 depicts the SHCS high-level interactions initiated by different possible users (e.g., patient, healthcare professional) involving different modules in the system such as the blockchain and the data lakes. Further details on the integration design can be found in previous published work [6],[8]. The SHCS interaction with the blockchain module checks access rules on the requested medical data. This enables the SPHR retrieval process in distributed data lakes. Consequently, an authorised user within the platform can only visualise the data she/he is authorised to access, according to the created and verified access rules.

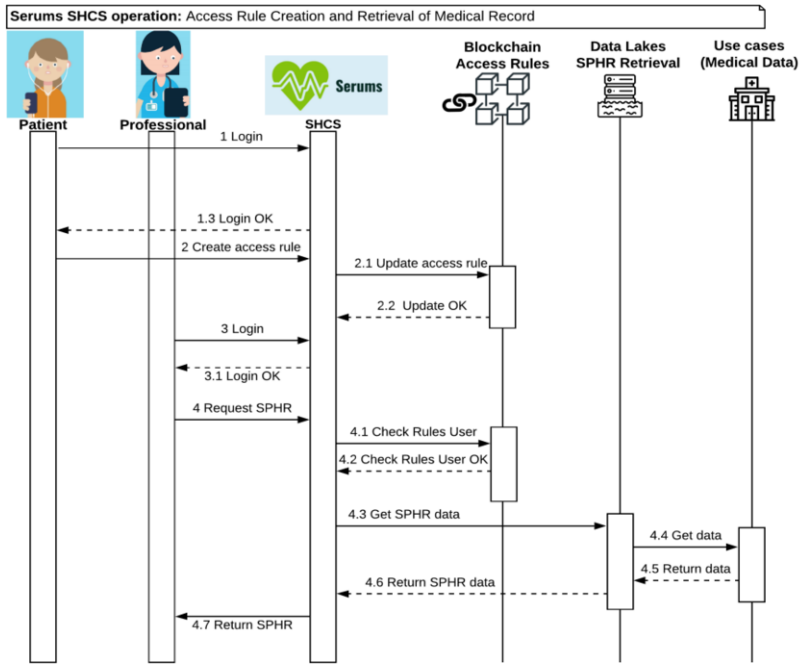

Figure 1. SHCS high-level operation and system modules' interactions.

2.1. Tags

Tags are a set of predefined keys that have been selected to act as generic names for subsets of data available in the data lake. Possible keys include elements such as *appointment*, *medication*, *operation*, *device*, just to name a few. The exact same set of tags is supplied to each of the authorised healthcare providers, i.e., medical data sources. It is then up to individual healthcare providers to decide what data should be assigned to each predefined tag. In practice, the tag definitions are stored as *JSON objects* within the data lake for the healthcare provider that defines them.

Tags can be used in flexible ways. For instance, a healthcare provider does not have to use all of the tags, only the ones that accurately describe their data. Furthermore, a single tag can be used to capture data from multiple sources across different healthcare systems. This could be the selection of more than one table within a database or even the selection of data from multiple systems within a single hospital's storage network.

2.2. Rules

While the use of tags defines *which* data can be shared, rules define *how* the data will be shared. Each rule describes: whose data the rule relates to; who the rule is applicable to (such as a specific doctor or group of specialists); whether to grant or deny permission to the professional(s) the rule is applied to; how long the rule is applicable; which tag(s) the rule covers. Rules can be generated by both healthcare providers and patients. Rules generated by healthcare providers follow a typical pattern of being the default rules applicable to all their patients enrolled in the Serums SHCS. This ensures that healthcare

provider data governance rules are enforced and any basic level of data sharing the institution requires is in place automatically, essentially mimicking their existing data systems. Patient defined rules, by contrast, allow end users to create custom access rules. Underlying a user-friendly interface to the rules definition within the SHCS, we propose a formal representation for the rules as well as a mechanism to further detect and solve rule conflicts automatically in production versions.

In the following, let T be a set of tags, Act be a set of actions, Id_S denote a set of identifiers indexed by a sort in S, where sorts correspond to granters and grantees, that is, S is a disjoint union where $S = S_G \uplus S_R$.

An *access rule* r is a tuple $r = (g, R, \alpha, D, \Gamma)$ where:

- $g \in Id_G$ is a granter, such as *patient*, *healthcare professional* and so on,
  - $R \subseteq Id_R$ is a subset of grantees (i.e., healthcare professionals) where necessarily $g \notin R$,
  - $\alpha \in Act$ is an action, such as *allow* or *deny*,
  - $D = (d_1,d_2) \subseteq \mathbb{N}\times\mathbb{N}$ is the time interval indicating when the rule is valid where necessarily $d_1 \leq d_2$, and
- $\Gamma \subseteq T$ is a subset of tags (e.g., appointment, operation, medication, device).

One example of a possible rule is:

$$r_1 = (p_1, \{m_1, m_2\}, \text{allow}, (d_1,d_2), \{operation, medication\})$$

where patient $p_1 \in Id_p$ allows doctors $m_1,m_2 \in Id_m$ to have access to all operations and medication records that p_1 received between dates d_1 and d_2.

When rules are defined for the same grantee, their combined effect can be given as a new rule, or conversely, be separated into multiple rules for different granters. For instance, if $r_1 = (p_1, R, allow, D_1, \Gamma_1)$ and $r_2 = (p_1, R, allow, D_2, \Gamma_2)$ are two access rules for $p_1 \in Id_G$ and $R \subseteq Id_R$, then naturally we write $r = (p_1, R, allow, D_1 \cup D_2, \Gamma_1 \cup \Gamma_2)$.

Assume the complete set of rules given by $\mathscr{R}$. We define a *priority function* over $\mathscr{R}$ by pf: $\mathscr{R} \to \mathbb{N}$, such that for arbitrary rules $r_1, r_2 \in \mathscr{R}$, if $pf(r_1) < pf(r_2)$ then r_2 is a rule with higher priority. A set of rules $R \subseteq \mathscr{R}$ for grantee g is correct if and only if there are no rules in R that contradict each other. Conflict can arise when different actions are used, for instance simultaneously allowing and denying access over the same data to the same granter for intersecting time periods. When rules are in conflict, the rule with the highest priority takes precedence. If their priority is the same, the disallowing rule takes precedence. To check rule consistency automatically we can apply efficient constraint solvers (cf. [6]).

2.3. Blockchain

The Blockchain module provides the utility of smart contracts to store and execute transactions and agreements between parties, i.e., patient and healthcare providers. It operates in an immutable and distributed database allowing easy audit trails logging the events taking place within SHCS.

At their most basic level, smart contracts provide an access flag as to whether a healthcare professional has access to particular patient data, and if so, what can they see. In order for a healthcare professional to visualise patient data via the SHCS, a rule must

exist on the blockchain which grants this. Without an explicit rule, access is denied and the request comes to an end with no data being retrieved.

The basic logging that SHCS implements covers all the events associated with the smart contracts such as when a new contract is created or updated, or when a healthcare professional makes a request for data. In addition, the blockchain tracks the lineage and provenance of any data that is returned via the SHCS. Data within a data lake is removed almost as soon as it has been delivered as part of a request. As such, it is important that each request leaves a permanent record of everything that took place during this process. This record contains no personal data, rather it describes what has happened in a way that allows system admins to test that a particular rule was executed correctly.

3. Conclusions

The use of a structured approach to define access rules enables us to tackle and conform to important security issues such as access control to personal medical data and governing policies. We have built a high-level model of access control to capture the particular requirements of the SHCS such as the definition to support collective and individual rules in such a way users can easily define who is allowed to access what when introducing the concept of tags in data lakes, from whom (which user, i.e., patient), and when (rule's validity). In future work, we aim to integrate a user-friendly interface within SHCS for defining rules as well as coding European use cases, first to enhance the set of proposed tags, and then to test the efficiency of the computational structures of the blockchain and data lake modules altogether. Moreover, our platform comes with a data access scheme that can easily be expanded into a global health and social care framework for big data processing, or even machine learning (ML) application at scale.

Acknowledgements: This research is funded by the EU H2020 project SERUMS (grant 826278).

References

[1] Gavrilov G, Vlahu-Gjorgievska E, Trajkovik V. Healthcare data warehouse system supporting cross-border interoperability. Health informatics journal. 2020;26(2):1321–1332.
[2] McGhin T, Choo KKR, Liu CZ, He D. Blockchain in healthcare applications: Research challenges and opportunities. Journal of Network and Computer Applications. 2019;135:62 – 75.
[3] Bowles J, Mendoza-Santana J, Vermeulen AF, Webber T, Blackledge E. Integrating Healthcare Data for Enhanced Citizen-Centred Care and Analytics. Studies in Health Tech & Inf. 2020;275:17–21.
[4] Mohan J, Wasserman M, Chidambaram V. Analyzing GDPR compliance through the lens of privacy policy. In: Heterogeneous Data Management, Polystores, and Analytics for Healthcare. Springer; 2019. p. 82–95.
[5] Truong NB, Sun K, Lee GM, Guo Y. GDPR-Compliant Personal Data Management: A Blockchain-Based Solution. IEEE Trans on Information Forensics and Security. 2020;15:1746–1761.
[6] Bowles J, Mendoza-Santana J, Webber T. Interacting with next-generation smart patient-centric health-care systems. In: UMAP'20 Adjunct: Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization; 2020. p. 192–193.
[7] Janjic V, et al. The Serums tool-chain: Ensuring Security and Privacy of Medical Data in Smart Patient-Centric Healthcare Systems. In: 2019 IEEE Int. Conf. on Big Data; 2019. p. 2726–2735.
[8] Webber T, Mendoza-Santana J, Vermeulen A, Bowles J. Designing a patient-centric system for secure exchanges of medical data. In: Int. Conf. on Computational Science and Applications (ICCSA 2020). vol. 12254 of LNCS. Springer; 2020. p. 598–614.