Computer Networks 205 (2022) 108766

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

A survey of data minimisation techniques in blockchain-based healthcare

Rahma Mukta [a,*], Hye-young Paik [a], Qinghua Lu [b], Salil S. Kanhere [a]

[a] *School of Computer Science and Engineering, UNSW Sydney, Australia*
[b] *Data61, CSIRO, Sydney, Australia*

ARTICLE INFO

Keywords:
Privacy
Data minimisation
Healthcare
Blockchains

ABSTRACT

The push for digitising personal health records needs to occur with serious consideration of privacy in order to instill public confidence. However, the healthcare sector still experiences leakage of Personally Identifiable Information (PII) due to improper data protection practices and security failures by data custodians. Data minimisation refers to the practice of limiting personal data collection and usage to only what is required to fulfil a specific purpose, and is one of the directives in many privacy regulations and data protection acts. Blockchain technology provides a *neutral third-party platform* for healthcare applications on which trust can be built to increase confidence in all participating parties to create, store and share sensitive data. However, the aspects of design and implementation of data minimisation techniques within the blockchain context have not been systematically explored and no effort has been made to review and analyse the existing solutions so far. In this paper, we undertake a survey of data minimisation techniques in blockchain-based healthcare systems. We provide a broad definition of data minimisation and classify data minimisation approaches according to the different lifecycle phases of data processing workflows. We also present a comparative analysis on privacy properties achieved by these methods. This study offers a unique view of data minimisation from both data custodians and data owners' viewpoints, and suggests several areas of future research and development to improve privacy in healthcare through data minimisation.

1. Introduction

The recent push to digitise and streamline the management of personal health records needs to occur with serious consideration of privacy in order to instill public confidence. However, according to the recent notifiable data breaches report by US Department of Health and Human Services, the healthcare sector still experiences leakage of Personally Identifiable Information (PII) more than any other sector [1].

There are many causes of data breaches, but most point to improper data protection practices and security failures by data custodians [2]. Given that preventing data leakage is not always possible, we turn our attention to *data minimisation* as an antidote to minimise the possibility of leakage while sharing the data.

Data minimisation refers to the practice of limiting personal data collection and usage to only what is required to fulfil a specific purpose, and is one of the directives in many privacy regulations and data protection acts (e.g., GDPR,[1] HIPAA[2]). The discussions on data minimisation so far have been focused on the obligations of the data custodians and techniques that are implemented by them. However, the same issue also can be considered from the data owner's perspective, i.e., the technological solutions that are intended to give data minimisation controls to the data owners. This would mean the owners *exercising the control* to limit personal data exposure in any data sharing situation, and choosing with whom to share their data.

Our motivations to formulate this survey paper are as follows:

First, broadly speaking, at the heart of any regulated privacy practice, such as data minimisation, is the challenge of balancing the inherent competing needs of the data custodians and data owners. The data custodians would like to maximise the information collected about a patient (e.g., for a healthcare professional to provide an optimal care, for a pharmacist to offer some personalised products). The data owners would wish to minimise the information exposure, even to a point that anonymity may always be preferred (e.g., a pharmacist may not need to know the patient's PII as long as the presented prescription is genuine and the prescription holder can be verified). Examining currently available data minimisation solutions from both viewpoints could highlight the gaps in meeting this balance.

* Corresponding author.
E-mail addresses: r.mukta@unsw.edu.au (R. Mukta), h.paik@unsw.edu.au (H.-y. Paik), qinghua.lu@data61.csiro.au (Q. Lu), salil.kanhere@unsw.edu.au (S.S. Kanhere).

[1] GDPR Article 5, https://gdpr-info.eu/art-5-gdpr/.

[2] Health Information Privacy, https://www.hhs.gov/hipaa/.

https://doi.org/10.1016/j.comnet.2022.108766
Received 17 February 2021; Received in revised form 27 October 2021; Accepted 3 January 2022
Available online 15 January 2022

Second, it is also widely acknowledged that the *"balance"* in current privacy practices still is tilted towards data custodians and in many cases, users are left with little choice but to trust in the good faith of the custodians [3]. Especially in the healthcare sector, the lack of trust and confidence in data custodians have been a major barrier in adopting electronic health records (EHR) [4].

Recently, blockchain technology has gained a lot of attention in the healthcare domain. Because of the explicit trust properties in the architecture (e.g., traceability, verifiability, and integrity), blockchain provides a *neutral third-party platform* on which trust can be built to increase confidence in both parties. We argue that blockchain is a uniquely positioned technology that could enable the balance between the two parties with competing needs. This could support design and implementation of new data sharing techniques that promote data minimisation practices on *both* sides.

In this paper, we will explore the blockchain-based solutions for data minimisation applied in healthcare from both data owner and data custodian's viewpoints. Exactly *how* data minimisation is achieved is not explicitly prescribed, but there are common privacy properties that are widely utilised by the privacy research community as the evaluation criteria for systems that practice data minimisation [5–7]. Our survey will present a systematic analysis of the solutions according to these evaluation criteria.

Related surveys and our contributions. To the extent of our knowledge, most of the survey articles have focused on providing comprehensive overviews of the security issues and challenges of blockchain applications [8–11]. Some others have focused on privacy aspects in blockchain based health data sharing [12–15]. However, it is noted that, to the best of found knowledge, there is no survey study that focuses on data minimisation techniques. This paper is one of the first works that systematically categorises data minimisation techniques and evaluates them according to privacy requirements. Our study particularly looks into the practice of limiting health data sharing in applications based on blockchain technology.

Furthermore, our paper offers several attractive influences. First, it leads to understand the data management challenges in health data sharing from an end-to-end viewpoint covering the whole lifecycle of data. Second, it identifies existing data minimisation practices (e.g., encryption, data masking, access delegation). Finally, this survey clarifies the system design considerations and implementation options for data minimisation.

Paper organisation. The rest of the paper is organised as follows: prior to presenting the results of our review, we will first introduce our research questions to lay the foundation for the survey in the next section. Next, in Section 3, we will define data minimisation and the scope of data minimisation on different data processing stages in healthcare sector. In Sections 4–6 we will review the data minimisation techniques in the extant literature, and current evaluation methods to set up our own evaluation criteria and provide a comparative summary of the existing systems based on those criteria, respectively. In Section 7, we identify a number of potential research opportunities. Finally, we conclude the paper in Section 8.

2. Survey overview

The goal of the survey is to identify whether blockchain can act as a *neutral third-party platform* on which trust can be built to increase confidence from both custodian and owner side. We attempted to answer this query formulating following research questions. The task of identifying and analysing the relevant literature is informed by these questions.

- (Q1) What is data minimisation? This question helps us understand the commonly accepted definition of the term "data minimisation" in the literature, through which we can identify relevant data minimisation techniques. Section 3 presents the results of this question.
- (Q2) What technical solutions are proposed in literature to implement data minimisation? This question will answer the details of how data minimisation is achieved according to the current literature. Also, we make note of the cases where the use of blockchain is relevant to implementing the data minimisation technique. Section 4 categorises and describes the results of this question. We have divided this section into two subsections: Section 4.1 includes the solutions from data custodian perspective. Section 4.2, from data owner perspective.
- (Q3) How is the efficacy of data minimisation in a system evaluated? We address this question in two parts: Section 5 discusses the common approaches for evaluating data minimisation techniques and introduces our evaluation criteria based on some of the standard privacy properties considered to be relevant to data minimisation. We have applied the criteria to the techniques found in the literature. We looked through each solution and examined if the privacy property in question (e.g. anonymity) is present. In Section 6, we present the results of our evaluations. Again, we divide the section into two: data custodian perspectives and data owner perspectives.
- (Q4) What roles does blockchain technology play in enabling data minimisation? To address this question, we synthesise the survey findings from (Q2) and (Q3) to gather how blockchains are currently utilised in data minimisation solutions and if the needs of data owners and data custodians are met, drawing further research questions and directions on the topic. Section 7 shows the results of this question.

As the context of the survey is in healthcare systems, we limit our coverage to related work in the healthcare domain, with particular focus on blockchain-based solutions. The selected papers were categorised and analysed according to the survey questions. In presenting the findings, we use the term 'solution sets' to describe the various technical solutions in the context of blockchains. We discuss the solution sets from the data owner and custodian perspectives, before presenting our overall observations.

3. What is data minimisation

We first discuss the common definitions found in the literature before introducing our own view of data minimisation for this paper. Next, we present the evaluation criteria for assessing the efficacy of data minimisation solutions.

When defining data minimisation, many papers refer to the GDPR Article 5 where data minimisation is listed as a privacy principle that ensures "the processing of personal data that is adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed". In the literature, some of the common phrases found to introduce data minimisation are: "the purpose of the processing must be specified when the data is acquired", "the data should be deleted when no more required for the specified purpose" [16], "the amount of shared data strictly limited to the minimum necessary" [17], "to minimise collecting personal data" [6]. In privacy regulatory documents, we see similar phrases, For example, HIPAA [18] states "to limit the scope of the PHI (Protected Health Information) the health systems use, disclose or request to the minimum necessary".

The emerging themes here can be summarised as (i) minimising the collection of personal data and the data retention period to the minimum necessary and (ii) collecting, using and sharing data for its specified purpose. These definitions, however, mostly "repeat" the regulatory document definitions and are less instructive on how data minimisation can be achieved.

In considering how data minimisation could be interpreted in concrete terms for personal data processing systems, we observe the privacy concerns this principle aims to address. That is, the concerns stem from the fact that personal data often contains PII, and uncurbed collection and sharing of PII increases the likelihood of the data being correlated and reveal even more personal information. This increases the severity of the potential harm that a data breach incident can cause.

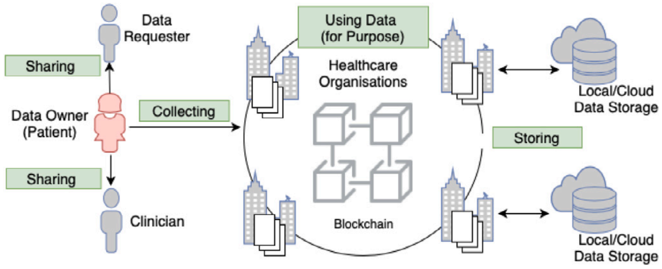


Fig. 1. Data minimisation in different data processing phases in healthcare system.

Indeed, the GDPR Article 5 articulates a set of principles relating to protecting personal data. Data minimisation applies to protection against excessive collection of PII. We believe this entails not only minimising unnecessary PII data collection but also properly guarding the data against improper access or sharing, and such protection should be in place throughout the life cycle of the personal data in the system.

Reflecting this, we examine the end-to-end data processing stages. For the context of this survey, we consider blockchain-based healthcare systems.

Fig. 1 shows a simplified view of a blockchain-based healthcare system in order to emphasise the main data processing stages in the system. Blockchains are typically present in healthcare systems as a data store, referred to as on-chain, which often holds metadata, hashed (when data stored off-chain) or encrypted EHRs or identity registry. It also acts as a third party authority that provides trust or security related operations (e.g., verification, authentication, provenance). The data flow depicts the healthcare clinics (data custodian) collecting the patient data, utilising it for certain purposes within the system, and storing it on local or cloud data stores (i.e., off-chain). Note that the patient can share the data with clinicians (for clinical purpose) or other data requester (for non-clinical purpose like research or marketing).

The boxed labels annotate the data processing stages. For each stage, we highlight the different types of data minimisation techniques may apply.

- Collecting stage and *collection minimisation*: during this stage, the custodian acquires personal data. *Collection minimisation* refers to the technical solutions that ensure the data collection is minimised (particularly, PII) and done with a specific purpose.
- Using Data stage and *purpose minimisation*: during this stage, the custodian utilises the data for specified purposes. *Purpose minimisation* refers to the solutions that ensure the collected data is not re-purposed for another use.
- Storing stage and *storage minimisation*: during this stage, the data is stored for later use/access. *Storage minimisation* refers to the solutions that ensure the data stored is securely protected from unauthorised access, and the data retention period is enforced (i.e., deleting data when no longer necessary).
- Sharing stage and *sharing minimisation*: this applies during data sharing phase to limit the disclosure of personal data to what is necessary. *Sharing minimisation* refers to the solutions where owner is able to granularly decide what information to share.

In the context of healthcare systems, data sharing is essential in providing treatments [19]. Sharing the right data at the right time can ensure proper treatment with reduced cost (e.g. no repeated tests). On the other hand, misinterpreted or delayed data can be life threatening. Therefore, the discussion on data minimisation techniques should include potential approaches to meet the balance between the privacy needs of a patient to be protected and the data collection/usage needs of the healthcare system to provide an optimal care.

4. Data minimisation technical solutions

To understand how the balance between the needs of the patients and data custodians in healthcare systems are met by the current solutions, we present representative works on data minimisation techniques from two viewpoints: data custodians and patients (as data owners). In deciding to which viewpoint a minimisation technique represented in a paper belongs, we consider where in the system the data resides when the technique is applied (i.e., if the data is on a server with the custodian, the technique is considered from the custodian point of view).

We use the different phases in the data flow (Fig. 1) to categorise the data minimisation techniques found in blockchain-based healthcare systems in the literature. We have grouped representative techniques in each category. We have considered multiple papers in each technique, and chose one or two works from each group based on the level of details available in the papers and to what extend the implementation and evaluation contribute to answering our survey questions. Finally, we chose 10 works for the custodian viewpoint and 11 works for the owner.

4.1. Data custodian's perspectives

First, we look at a range of data minimisation techniques from the custodian's viewpoint, which are categorised into the following three solution sets in accordance with the Collection stage, Using Data stage, and Storing stage of the data processing in Fig. 1.

4.1.1. Solution set 1 (data masking): Minimising collection

During Collection stage, the data custodian is responsible for collecting data legitimately (i.e., according to their purpose), while minimising inclusion of PII. The decision to collect minimised patient's data can be made at design time or at run time by custodians. For example, at design time a custodian's system may make some data entry fields optional, giving the patient the choice not to expose some data to the system [20]. For specifying the purpose, a general solution is to display the custodian's privacy policy stating the purpose of data collection, to which owners consent before submitting the data.

Given that the ultimate reason for limiting the collection is to minimise the risk of the patient data being correlated, a more secure approach would be to collect the data anonymously. In this set of solutions, we see a range of techniques that are designed to collect data while concealing the identity of the users. That is, data minimisation is achieved by the custodians not explicitly linking the data with the patient's identity.

- Dwivedi et al. [21] proposed a lightweight digital ring signature on EHR to obscure the signer's identity. Under the scheme, the data owner's signature is mixed with signatures of other users within the named ring, making it difficult to associate the data with a single identity. Blockchain network serves as an intermediary for managing the individual keys that are to be included in the formation of the ring signatures and also to manage the authentication of users. A signer can ask for a set of public keys from the blockchain ledger to produce a ring signature. No one can identify the real signer, because anyone within the authenticated group can be a signer. During verification, the receiver must be able to compute the signature from the correct constituent keys retrieved from the blockchain network. Shukla et al. [22] proposed a similar technique where owner's IoT device (i.e., data generator) signs the EHR using ring signature. For these schemes to work as intended, ensuring the authenticity of the public keys used to produce the ring signatures is essential. The blockchain provides such a trusted key exchange platform between the signer and receiver.

- Qianqian et al. [23] proposed a system named HS-BC (Block-chain Based Healthcare System) where hospitals maintain a permissioned blockchain together to store EHR. HS-BC uses Attribute Based Signature (ABS) scheme as a means to verify if the EHR being received from blockchain have been signed by an authorised role (e.g., doctor) in the system. The *attribute authorities* (AAs) generate attribute master-keys and update-keys for each user, based on their claimed attributes (e.g., for Jane, who is a doctor, may claim attributes such as "work-for Hospital A", "is-a Physician"). Jane can use a signature generated based on these attributes to sign an EHR instead of her real identity. Through the AAs, the EHR can be verified to have been signed by a doctor, without revealing Jane's identity. Therefore, blockchain works here as a trusted ledger to prove doctor's signature while concealing the identity information.
- In Hossein et al. [24], a combination of public key and private key are used as pseudonymous identifier of a EHR, thus masking the real identity. These pseudonymised EHRs are stored in an off-chain database. A private blockchain is used to store the hash of the EHRs (for checking the data integrity) and keeps track of the access to EHRs as logs. Chen et al. [25] adopted a privacy reserving technique, called K-anonymity to pre-process some sensitive data field (that may reflect owner identity) of the EHR, where a set of values (e.g., patient age indicates a range 20–30) are appeared at least K times in a table. Thus, the owner identity cannot be inferred directly. In BBEHR [26], the fingerprints, instead of identifiable personal attributes, are used as a pseudonym to authenticate the patient's identity and link to their respective EHRs. In all of these works identity information is disguised within a pseudonym and blockchain establishes the trusted link between the pseudonym and associated EHR.
- Thein et al. [27] have proposed a semi-trusted gateway server (i.e., honest but curious) to retrieve the encrypted EHR from a cloud storage and send it to the data custodian. In this type of transactions, the metadata which is stored on-chain in a private blockchain, can prove the authenticity of data being transferred by the gateway server. The idea is to discourage direct communication between the data owner and custodian to avoid identity disclosure during data sharing. The access logs stored in the blockchain can be used for auditing purposes if any suspicion about unauthorised access arises. Therefore, using a gateway server supports data minimisation by allowing the data owner to avoid identity disclosure. However, the tamper-proof characteristic of blockchain for maintaining the access logs is required to make the end-to-end communication trustworthy.

4.1.2. Solution set 2 (access delegation): Minimising purpose

During Using Data stage, the most appropriate technical solution is to enforce the specified collection purpose with the intended use. Once the data is collected for one purpose, the custodians need to acquire further consent from the owners before sharing with other parties for different purposes. However, in current implementations, consent management is "static" in that the user consent is sought at the start of data collection (for all anticipated purposes for the data) and is rarely updated. We found no custodian-side solution that contacts users to seek explicit consent if the data is to be re-purposed.

In a healthcare setting, frequent sharing of patient data may have to occur, sometimes for unforeseen reasons. When a doctor needs to discuss a patient's records with another clinicians, a patient's consent may be required. A technical means for managing user consent in these scenarios is managing access delegation. That is, Mark (a patient, delegator) can *give consent to delegating access* to Jane (a doctor, now delegatee). Jane then has an authority to share Mark's data with another party according to the delegation conditions.

In this set, we see examples that provide delegatable access control. We identify the following technical solutions:

- Pussewalage and Oleshchuk [28] proposed an extension of an attribute-based access control scheme which integrates access delegatability. That is, in this scheme, a user can delegate access to another user via delegating *access attributes*. Further, the scheme allows the user to limit the length of the chain of delegation. Blockchain is adopted to transparently record and confirm the assignment of access attributes for delegation as well as revocation. As a decentralised platform blockchain makes this system conducive to a multi-domain healthcare environment while simultaneously enforcing the specific collection purpose based on the access attributes. Therefore, in this scheme, blockchain is supporting purpose minimisation (i.e., delegatable access control) by enabling a confirmation of the delegation conditions (i.e., the access attributes) prior to granting access on health data.
- MedSecureChain [29] offered a similar delegated access to requesting EHR users according to the conditions of authorisation table generated by the owner. The table is used to define access levels to different categories of EHR users using some numeric access codes. Here, both EHR and authorisation table are stored in blockchain enabling the confirmation of access approval.
- Ahmed et al. [30] proposed an emergency access control mechanism where the owner is the delegator and the smart contract acts as a delegatee. In normal condition, data is frequently accessed by multiple participants, such as the patient, family members, and family doctor based on a pre-defined access list. When the purpose is 'emergency', the permissioned blockchain network only gives the health data to authorised emergency staff which have the granular access rights from the database according to the access permissions (patient's rules).

4.1.3. Solution set 3 (access control): Minimising storage and access

For Storing stage, data custodians will employ various access control schemes to minimise the number of users who can access sensitive data. The delegatable access control schemes discussed earlier have overlapping functions for this purpose. In this solution set, we will point out other access control schemes that do not involve delegation. During this stage, another concern will be data retention management where data is removed from the system after a specified retention period. Ideally, such removal should be automatically enforced by the system. We identify the following technical solutions:

- ACTION-EHR [31] is a permissioned blockchain-based system for EHR data integration and sharing. Each healthcare centre provides a blockchain node connected with its own EHR system to form the blockchain network. A patient can initiate a record-sharing request to his caregiver and defines the sharing permissions to the caregiver's healthcare centre within the network. A patient then needs to specify the category of the data to be shared, and for how long the sharing permission should remain valid. The transaction recording this permission is generated automatically based on the information provided by the patient (via a web application) and is broadcast to the blockchain network. Here, blockchain plays the role of an immutable store of these permission records. Simultaneously, the encrypted patient data is uploaded to the cloud according to the permission.
- MediBChain [19] also uses a permissioned blockchain to store EHR. In their access control scheme, a registered data owner will share encrypted EHR with blockchain through the private accessible unit (a secure channel to connect with blockchain), which is then stored on-chain. Each such transaction in the blockchain returns a transaction identifier (block number). This transaction identifier is associated with the owner's identity and the owner can solely use the identifier to access the on-chain data in the future. Therefore, blockchain infrastructure enforces the access permission to valid users (e.g., owner and owner permitted authorised person).

We did not find any work that particularly addresses EHR retention management where data is removed after a specified retention period on the custodian side. Firstly, removing EHR from a blockchain-based system is challenging due to the immutability of the underlying platform. Secondly, within the realm of custodian control, owners have limited power to access and remove the data after the desired retention period. However, the authors in [32] proposed a solution where an owner can send a request to the data custodian to delete old EHRs stored on the server. The following needs to be done: (1) the owner (patient) of the EHRs sends a delete request to the server; (2) the server sends the information for deleting the EHR as part of the transaction of the next block; (3) once a new block is mined and verified, the server changes smart contracts to unlink the reference to the EHR to be deleted; (4) the server deletes patient's EHRs; and (5) the server sends a notification to the patient to indicate that the EHRs are successfully deleted.

4.2. Data owner's perspectives

Now we look at a range of data minimisation techniques from the owner's viewpoint. We categorised the following three solution sets concerning the Sharing stage, Using Data stage, and Storing stage from Fig. 1. The solutions discussed below allow owners to hold and manage their own data (e.g., an owner can selectively share data on a need-to-know basis).

4.2.1. Solution set 4 (selective disclosure): Minimising data sharing

During Sharing stage, data minimisation concerns are on exposing as little data as necessary and avoid revealing PII. We looked to *selective disclosure* techniques where the data owner can choose a subset of the data they want to share (e.g., share email address, but not the phone number in a patient registration record).

- Sun et al. [33] proposed a decentralised attribute-based signature scheme (called DABS) where the data owner is assigned a series of attributes issued by different attribute authorities which are combined to form their global identity (GID). The owner can obtain the corresponding signature key for each attribute from the authorities. The owner can then choose to share only a subset of the attributes (i.e., minimised sharing) by signing the selected attributes with their corresponding keys. When sharing the health data, the owner broadcasts the address of the data to the blockchain along with the signature. The receiver then verifies the attributes through the associated signature. Blockchain here plays the role of a trusted ledger for managing signatures and their verification.
- Jingwei et al. [34] designed a selective content extraction algorithm for selective disclosure of sensitive data. To avoid illegally extracting content, the data originator (clinician) defines extractable data fields from the content when signing it (e.g., from a medical record for a patient, a clinician may only define age, while address and diagnosis may be extracted). Next, patient extracts their records accordingly and generates extraction signature. The encrypted extraction signature along with corresponding EHR is stored in the cloud with an index reserved in a temper proof consortium blockchain. Daniel et al. [35] proposed to apply redactable signatures in context of medical records where the owner can redact PII from the records. The extracted/redacted data are proved as authentic using blockchain, who stores the extraction/redaction signature and the associated keys (to verify the stored signature).
- Tomaz et al. [36] and Sharma et al. [37] utilise a selective disclosure technique known as *zero knowledge proof (ZKP)*. In ZKP, one can prove the possession/knowledge of X without revealing X itself. A practical example of ZKP is proving that someone is over 18 years without showing their DOB. In [37], the authors proposed a data sharing system that integrates a health insurance claim process with a patient's EHR. They introduced a ZKP-based patient identity authentication system where a pin code or passphrase is used to prove the patient's nation-wide digital identity. The hash of passphrase is stored on a smart contract. Patient needs to generate the knowledge of pre-image of the hash for authentication rather than disclosing the detail identity. The authors in [36] proposed a mobile health (mHealth) system.[3] To ensure that only a legitimate wearable device connects to the official mobile health application, a ZKP-based device authentication scheme is used where a prover (wireless device) must send a proof of its identity (public key) with encrypted (using a shared secret key) health data in a single message to the verifier (mHealth smartphone App).

4.2.2. Solution set 5 (consent management): Minimising purpose

Similar to the custodian's viewpoint, for Using Data Stage, we consider the consent management as a technical means to achieve purpose minimisation. The consent mechanisms here can be more "dynamic" compared to what is available from the custodian side, in that as long as the data resides with the user, the access request can be made every time the data is needed and the consent can be given on a per-request basis (i.e., per use). For each request, its data access purpose can be checked and consented to by the user. We identify the following examples in the category:

- Xueping et al. [38] proposed a personal health data management system, especially for the data collected through wearable sensors. The data is uploaded onto a cloud server where the user has full control over issuing data access permission per request. This allows the user to manage the purpose (and consent) of data sharing in a fine-grained manner. For instance, a token-based verification scheme allows the user (i.e., data owner) to issue one-time-only access to a data custodian. Each time there is a data request, the data owner generates an access token for that particular requester. The requester then presents the token to the cloud server, which verifies it before allowing access. Each token operation is recorded on the blockchain to avoid double spending (i.e., multiple uses of the same token). Additionally, blockchain stores the data access logs (hashed) which can be used to trace data access activities or detect unusual access patterns (e.g., pointing to a potential information leakage). In summary, the token-based access control allows data owners to issue a token per access request, which can be tailored to the purpose of access. The blockchain strengthens the scheme by allowing it to keep track of each token usage (and its intended access purpose) in immutable logs. A repeated attempt to use the same token can be automatically disregarded.
- Hylock et al. [39] addresses the issue of consent revocation as part of managing data sharing rights. The system uses a proxy re-encryption technique which enables the delegation of data decryption rights by a delegator to a delegatee through an intervening proxy. However, once the re-encryption key (which gives a decryption right) is exposed to a delegatee, it is difficult to revoke it (i.e., the key can be reused by the delegatee against the owner's wishes). In the proposed scheme, named 2PD (two party decryption), the said problem is addressed by adding an "intermediary" to the protocol. In 2PD, decryption requires keys from delegator, delegatee plus the intermediary who manages re-encryption keys for the delegatees. A re-encryption key alone does not allow the delegatee to decrypt the data, practically revoking the key after each use. The system utilises smart contracts in a permissioned blockchain to transparently implement the intermediary's behaviour.

[3] Medical assistance systems that are supported by mobile devices and wireless sensors.

- BCHealth [40] proposes the usage of separate chain to keep the access policies private to the owner. The system uses a permissioned blockchain called 'data chain' to store the hash of EHR. At the same time, to control access over EHR, owner stores the desired access policies in another private blockchain called 'access control (policy) chain'. This separation ensures that the access policies will remain immutable and private to the owner, and access to the EHR data will be controlled as expected. In the proposed system, the owner stores the access policies in chain as a type of transaction, named "Policy Transaction". There is a binary value that determines the validity of the transaction, value 1 for valid and value 0 for invalid transactions. If the owner needs to revoke permission from a specific user, he creates the same policy transaction with value 0.

4.2.3. *Solution set 6 (access control): Minimising storage and access*

For this solution set, we consider a range of options found in the literature for providing owner-centric data storage and access control architectures. We mean "owner-centric" when the system design allows the data to reside completely on the owner side (e.g., the owner's own cloud storage, personal devices). The user consent related schemes discussed in Solution Set 5 have overlapping functions with this solution set (e.g., access control). Here, we discuss generic software architectural options for implementing owner-centric data stores. We identified the following technical solutions.

- In GuardHealth [41], data owners can encrypt personal data from IoT/smart devices and health clinics and upload them to their own cloud storage. A consortium blockchain and smart contracts are used to facilitate data sharing. The indexes of the owner's data are generated and stored on the blockchain and the owner can manage and enforce access policies as smart contracts. After the permitted time period, smart contracts withdraw the access permissions automatically.

  A similar architecture is presented in Nguyen et al. [42] where a private cloud is used for storing data (by the owner). A private Ethereum blockchain and smart contracts are used to encode and enforce data sharing policies.
- The authors in [43,44] discuss the concept of self-sovereignty in patient identity (i.e., total control over managing their own identities and identity data) and the role of blockchains. Decentralised identity management models such as Self Sovereign Identity (SSI) remove the need for central control in identity management. SSI uses a peer-to-peer model where identity is established and verified through a neutral data registry. The data registry is often implemented as a blockchains. Although real-world implementations that offer self-sovereignty of patient identity are in early development [45–47], the recent W3C standard schemes on digital identity (e.g., DID (Decentralised Identifier))[4] and identity data model (e.g., Verifiable Credential (VC))[5] provide practicable means to realise the concept in healthcare.
- Yue et al. [48] addresses the issue of managing data retention. The proposed EHR management system has a data storage layer (a private blockchain) and a data management layer. The data management layer is designed as a set of mobile apps called *HDG* (Healthcare Data Gateway) that are independently running, but communicate with each other. Data requests and access do not bypass HDGs. In their access control model, a data requester (e.g., a doctor) enters a request into their HDG which includes the requester's desired retention period (e.g., 60 days from the collection). The data owner's HDG processes each incoming request and sends a copy of the data to the requester's HDG. After the retention period, the requester's HDG will delete the data automatically. Blockchain is here to ensure the data authenticity sent at requester's HDG.

Tables 1 and 2 present the summary of the existing works. We consider the performance and limitations of the main ideas with regard to data minimisation. We investigated the role of blockchain specifically in each work. In almost all of the works, blockchains are used as a trusted middleware to confirm EHR integrity and to establish a trust relationship between the owner and requester. Considering the limitations of each work, the issues listed in Tables 1 and 2 show the further need of privacy preservation (e.g., can a user remain anonymous). This gives the basis for evaluating the data minimisation techniques according to the privacy properties as detailed in Section 5.

5. On evaluating data minimisation techniques

This section presents the first part of the response to (Q3) (in Section 2), where we discuss how data minimisation techniques are currently evaluated in a system, and detail our evaluation criteria.

When it comes to assessing the effectiveness of data minimisation, we find that there is a lack of commonly-accepted approaches that are specifically designed for evaluating data minimisation techniques. We also find that some privacy or security solutions that are not explicitly presented in the context of data minimisation, can also achieve some aspects of data minimisation, but are not evaluated through the data minimisation lens.

For example, a health record encryption technique aimed at ensuring data integrity can also achieve data minimisation when it is used to restrict data sharing with only legitimate users (i.e. *sharing minimisation*). Using a data masking technique such as ring-signatures can hide PII, so it can be considered as a form of *collection minimisation*. An access delegation scheme that delegates access authorisation to only a party with a specific purpose can be seen as *purpose minimisation*. The literature where these techniques are used however consider evaluation methods that are mainly aimed at demonstrating system performance or scalability [33,41]. Some works use security threat models or mathematical analysis to show how security properties such as integrity and availability are satisfied [35,38].

Although the said methods may be appropriate for evaluating their respective solutions from a technical point of view, none of them explicitly analyse efficacy of those solutions from the data minimisation view-point. Therefore, for this survey, we need a consistent framework to evaluate the techniques identified in the literature through the data minimisation lens.

5.1. *The LINDDUN privacy properties for data minimisation*

Pfitzmann and Hansen [7] proposed a particular privacy terminology for data minimisation which includes 12 privacy properties (e.g., *anonymity*, *unlinkability undetectability*, *identifiability*).

Based on this work, Deng and Wuyts et al. [49] introduces a comprehensive list of privacy properties and a privacy threat analysis model, named the LINDDUN methodology (Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of information, content Unawareness, policy and consent Noncompliance). Both the properties and threats categories are derived from studies of privacy policies and regulations, including the healthcare domain. The authors separate so-called "hard privacy" from "soft privacy". The inclusion of hard privacy properties in a system means the system expects as little data as possible from the data owner and employs a high-level data protection scheme under the assumption that adversaries are always motivated to breach privacy. The hard privacy therefore refers to strict personal data protection *including data minimisation principles*. Our choice of privacy properties for evaluating data minimisation techniques is informed by the hard privacy concepts developed in the LINDDUN methodology.

The soft privacy concerns privacy properties that are relevant to monitoring, policy enforcement and audit, such as *Content awareness* (the data owner can view their own data), *Policy compliance* (the custodian has a privacy policy in place), which we considered not

[4] DID, https://w3c-ccg.github.io/did-primer/

[5] VC, https://w3c.github.io/webpayments-ig/VCTF/primer/

Table 1
Custodian perspective: summary of existing works with blockchain role. —refers to 'not mentioned'.

Scheme	Main ideas	Blockchain type	Blockchain role	(+) Performance (-) Limitations
[21]	To obscure signer's identity	Private	Trusted intermediary for user authentication	(+) EHR owner can sign data anonymously (-) did not consider malicious activity from internal user
[22]	To obscure signer's device identity	Private	Trusted intermediary for device authentication	(+) secure data transmission (-) access control is not considered
[23]	To confirm that the EHR are signed by authorised role	Permissioned	Trusted ledger is used to prove identity without disclosing	(+) ABS with attribute revocation (-) EHR is stored directly on chain
[24]	To preserves privacy of the patients	—	Trusted ledger to ensure data integrity	(+) pseudonymity of patients (-) pattern detection of pseudonyms
[25]	To anonymise owner identity	Consortium	Middleware to perform as the trust layer among users	(+) EHR pre-processing for privacy preservation (-) may suffer from background knowledge attack
[26]	To disguise owner identity	Private	Create a trusted link between EHR and owner pseudonym	(+) biometric identity to ensure recoverable access to EHR (-) pattern detection of the hash values of biometric data
[27]	To avoid identity disclosure during data sharing	Private	Ensure trustworthy communication among the users	(+) reduced data sharing responsibility to owner (-) gateway server is semi-trusted
[28]	Multilevel access delegation based on user purpose	Public	To record and confirm the validity of access attributes	(+) controlled access delegation (-) pseudo identities are included in plain in public blockchain
[29]	To allow limited access to EHR	Private	To store EHR and authorisation rules	(+) authorised access to EHR (-) no access revocation mechanism
[30]	To define the rules for emergency access	Permissioned	Immutable storage of access permissions	(+) emergency access management (-) Only emergency purpose is considered
[31]	Patient specified sharing permission	Permissioned	Trusted auditing service	(+) limited and authorised access on EHR (-) update operation of sharing permission needs to update the ledger
[19]	To store EHR data in blockchain	Permissioned	Store EHR and enforce access permission	(+) authorised access (-) encrypted EHR is stored on chain

directly relevant to the technology-focused solutions that we plan to evaluate.

The mapping of privacy properties and associated threats are summarised in Table 3. We consider the privacy threats and properties referenced in the "hard priacy" category for evaluation. In the following, we introduce a brief description of each privacy property.

5.2. Data minimisation properties

Anonymity: Anonymity is the state of an individual of not being identifiable within a set of individuals. Anonymous data cannot be re-identified, as anonymisation is the process of removing personal identifiers, both direct (e.g., name, address) and indirect (e.g., postcode, personal income).

Pseudonymity: Pseudonymity refers to the use of pseudonym (i.e. identifier other than real name) as identifiers. The idea of pseudonymisation is to process the personal data in such a manner that the personal data can no longer be attributed to a specific individual without the use of additional information. A typical pseudonym scheme will then store such additional information separately and apply technical and organisational policy measures to ensure that personal data is protected from being linked to a natural person. The distinction with anonymisation is that, pseudonymous data can still be identifiable with the use of such additional information.

Unlinkability: This property describes the inability of an adversary to sufficiently distinguish whether two items of interest within a system are related or not. For instance, given some EHRs containing PII (e.g., the social security number, date of birth, gender) and medical diagnosis details of each person, a system with unlinkability would ensure that an individual who happens to know DOB and gender of a particular person, cannot associate it with the person's medical diagnosis.

Plausible deniability: This refers to the ability for a user to "deny having knowledge, or having performed an action that other parties can neither confirm nor contradict [49]". From the privacy threat point of view, this property means an attacker can indeed prove that the user has the knowledge or has performed the action. In some situations, plausible deniability is desirable over non-repudiation (e.g., off-the-record medical consultations).

Undetectability: Undetectability is defined as "the inability for an attacker to sufficiently distinguish whether a data is the item of interest (IOI)" (e.g., a message in transit is not discernible from random noise) [7]. Anonymity and unlinkability are about the relationship between IOI and the identity of the owner, undetectability concerns protection of IOI itself.

Unobservability: This property needs to satisfy both: (1) the inability of an attacker to sufficiently distinguish the presence of an item of interest (IOI) (i.e., undetectability) and (2) anonymity of the identities involved in the IOI.

Confidentiality: This implies controlled access and protection against unauthorised access to personal data. Typically in a healthcare context, confidentiality between clinician and patient is based on ethical and legal usage of data. However, when patient data is to be shared with other data custodians (except clinician), access control is necessary to ensure confidentiality. In addition, according to Saunders et al. [50], "Confidentiality refers to all information that is kept hidden from

Table 2
Owner perspective: summary of existing works with blockchain role.

Scheme	Main ideas	Blockchain type	Blockchain role	(+) Performance (-) Limitations
[33]	To share some selected attributes signed by the attribute key	Consortium	Trusted ledger for verifying the shared attributes	(+) sharing a subset of information (-) overhead of managing the attribute keys
[34]	Selective content extraction from EHR	Consortium	Trusted ledger to verify the extracted content	(+) selective sharing (-) overhead of storing extraction signature
[36]	Device authentication using zkp	Public	Proof verification	(+) sharing the proof without disclosing the device identity (-) applicable for limited sets of data (e.g. numeric data)
[37]	User authentication using zkp of passphrase	Public	Smart contract to proof hashed passphrase	(+) authenticated access to EHR using zkp (-) applicable for limited sets of data (e.g. numeric data)
[38]	Token generation per access request	private	Keeps track of token usage	(+) token based access control considering the usage purpose (-) token pattern analysis based on owner signature may reveal access information
[39]	To manage consent using proxy re-encryption	Permissioned	Smart contract as a delegatee to manage re-encryption key	(+) support consent revocation (-) update of re-encryption key requires to reproduce the contract
[40]	A separate private chain is used to store the user access policies	Permissioned and private	As a decentralised controller to store the hash and access policies of EHR	(+) chain with access policies is only accessible to the owner (-) data accessing overhead due to multiple chain
[41]	Owner manages limited access on EHR	Consortium	Smart contract to enforce data sharing policy	(+) supports access permission revocation (-) did not consider internal malicious node
[42]	To use smart contract to keep a list of grantee	Private	Trusted intermediary to enforce data sharing policy	(+) supports dynamic access management (-) smart contract needs to be updated each time a new grantee is accepted
[44]	Self-sovereignty in owner identity	Public	Acts as a trusted central controller of identity system	(+) owner can manage their identities (-) increased responsibility to EHR owner
[48]	Gateway for retention management	Private	Ensure data authenticity	(+) shared data is only accessible at gateway (-) limitation of gateway in large file handling

Table 3
Mapping privacy properties to threats; [49].

Privacy properties	Privacy threats
(Hard)	
Unlinkability	Linkability
Anonymity/Pseudonymity	Identifiability
Plausible deniability	Non-repudiation
Undetectability/Unobservability	Detectability
Confidentiality	Disclosure of information
(Soft)	
Content awareness	content Unawareness
Policy/consent compliance	Policy/consent Non-compliance

everyone except the primary requester". Sharing encrypted data is thus the widely implemented solution to maintain data confidentiality.

Note: even though *anonymity* and *pseudonymity* are related, we discuss them separately in our evaluation. In many cases, complete *anonymity* can be considered a ‘double-edged sword’ as it allows a dishonest data owner to remain undetected in the system. This could be particularly problematic in the healthcare sector where having the linkability of medical history could be paramount in providing good care. The current Opioid crisis [51], for example, partly blames the inability of the healthcare systems to track and monitor a patient’s medical history who may be illegally collecting prescriptions. Moreover, anonymisation eradicates the possibility of contacting the data owner again if reuse of their data requires their consent. Contacting the owner is necessary for purpose minimisation [52]. Thus, the healthcare sector requires an alternative solution that lies in the middle of complete anonymity and no anonymity. This is where *pseudonymity* evolves as a probable solution in the healthcare literature.

6. Evaluation results

To evaluate the data minimisation solutions, we studied each paper in terms of technical description of the proposed solution. We then determine if and how the privacy properties are considered in their design and implementation.

- For anonymity, we examined if the technical solutions specifically remove identifiable information from the data before sharing or storing it. Some works use known techniques such as ring signatures [21], while others design their own techniques such as a token-based data sharing scheme (completely removing PII) [38].
- For pseudonymity, we observed if the technique applies any new or known pseudonym-based identifiers in the data. For instance, use of bio-metric data for identity management [26,53] or digital key signature schemes [31] has been considered.
- For unlinkability, we examined the technical solutions to determine if there is a potential risk of linkability. Some works specifically mention unlinkability in their solution.
  For solutions that use the same key signature to sign different records, we determined that there is no unlinkability because the records associated with the same signature can be correlated to reveal unintended information. Some other solutions that provide allow the user to create multiple key signatures or transform the same key signature with hashing techniques are defined as showing unlinkability.

- For plausible deniability, we looked at the technical designs to determine whether there is any related metadata (e.g., hash of data with timestamp, approval signatures) or if audit trail logs are stored on-chain. This information potentially could be used to confirm or deny if a certain event occurred.
- For undetectability, we tried to identify if there is any technique employed to disguise stored and shared data (e.g., encryption techniques) both in storage or in transit.
- For unobservability, according to the definition in Section 5.2 unobservability reveals only a subset of data, that could be revealed from lack of anonymity and undetectability. Thus, we looked at the system design to find any probable data leakage due to lack of anonymity solutions.
- For confidentiality, as per the definitions outlined in the previous section, we focused on the following two aspects in the solutions: (i) whether the custodians or owners can exercise access control over the data, (ii) if encryption techniques are used to store or share data (i.e., undetectability).

We summarise the evaluation results according to data custodian and data owner perspectives. In Tables 4 and 5, for each property examined, we determine that:

- the solution exhibits the property (●).
- the solution partially exhibits the property (◑), e.g., for confidentiality, a solution may allow the data owner to have control over access to her data, but provide no encryption.
- the solution does not exhibit the property (○).
- the solution exhibits the property because it is implied by other properties (⦿). For instance, [7] considers that unobservability is achieved if undetectability and anonymity are present.
  We also argue that for the solutions [34,44] anonymity can be attained, if pseudonymity and unlinkability are present in the system. These works used public key as pseudonyms (pseudonymity) and eliminate the possible linkability of the public keys (i.e., each participant generates a unique account and each health record transaction on the blockchain can be signed with a randomly generated public key) to the real identity (unlinkability). This leads them to achieve anonymity (real identity is hidden using multiple pseudonyms) [54].

It is noted that when a solution achieves anonymity (e.g., use of ring signatures), we considered pseudonymity is *not applicable* and marked the property with (—).

6.1. Data custodian perspectives

Table 4 summarises the evaluation of the solutions from the data custodian's view point. Besides what we have summarised in Table 4, below, we include some remarks on each property.

ANonymity (AN). Using a ring signature allows a signer to sign data anonymously, that is the signature is mixed with other groups (the named ring), and no one (other than the original signee) knows which member/device signed the message [21,22]. Such a technique may still be vulnerable to inferences or pattern detection.

To mitigate such risks, Thein et al. [27] aimed to achieve relationship anonymity (hide correlation between sender and recipient). This is done through supporting indirect communications between the data owner and custodian via a trusted gateway server.

PSeudonymity (PS). Different de-identification techniques can hide the real identity of a user. Several implementations use public key as the pseudonym [19,24,28,39] to hide the user's real identity. ActionEHR [31] stored EHRs on a blockchain in a "key–value" pair form, where key is the pseudonym of a patient. However, multiple interactions with a pseudonymous identity can be problematic as the relation between the real identity and a pseudonymous identity can be exposed. Chen et al. [25] adopted k-anonymity to convert the sensitive data fields (that may identify the owner) into a range of values (e.g., zip code is between [2010-2019]), so that the table gets multiple values with similar range. However, k-anonymity cannot guarantee anonymity since homogeneity attack and background knowledge attack may reveal the owner's real identity [55]. In the past, many systems have been designed using innocuous identifiers like biometric identity [26,53] to reduce the computation overhead associated with the use of private keys. However, as biometric data is regarded as personal data, new privacy concerns may emerge. In [23], clinician sends the generated data to blockchain. To achieve sender anonymity, clinician must only share the required attributes (e.g. hospital name, department, title) for verifiability, but not the real identity.

UnLinkability (UL). Most healthcare systems examined from the data custodian's perspective implement pseudonymous identity management systems. Identity is typically obscured behind a public key, but other attributes of data are publicly shared [28]. This is problematic for health data. First, basic demographic information (e.g. age, race, ethnicity, gender, marital status, income, education, and employment) can identify people [56], and if an individual's pseudonymous public key is matched to their real identity, all transactions associated with that public key can be linked to that individual's identity. Depending on the role of the individual in the system, such a leak may reveal information at different granularity (e.g., a psychologist's signatures on prescriptions may reveal more specific information than what a general practitioner's could). One possible solution is not to disclose the public key while signing a data transaction. Ring signature endorsed by [13] can achieve this, where a signer selects several users to form a group including himself, then signs a message on behalf of the group using the group of public keys (from group members). By verifying the signature, the verifier is sure that signature was generated by someone in the group, but he cannot find the public key of the real signer from the group of keys. Thus ring signatures ensure unlinkability, because the verifier knows nothing about the real signer [57].

While linkability is a known problem on a public blockchain, it is also problematic on a private blockchain as an individual may not want all of the members of the private blockchain to have access to the same data, or they may want to revoke authorisation to their data at a later point in time, both of which are not possible once their identity is linked to their public key. Thus, blockchain implementations that allow for selective disclosure of private information (e.g. such as Zcash) and rely on approaches such as zero knowledge cryptography to provide verification of transactions with a high degree of privacy over the underlying data will be needed within the healthcare industry.

Plausible deniability (PD). In many implementations, we observe the use of digital signatures [21–24,27,28,31] associated with the EHRs, some directly stored on the blockchain, some indirectly linked through metadata. In some cases, biometrics such as a patient's fingerprints are used for identification (e.g., [26]). All of these traces, combined with the immutability of blockchain records, may deny the plausible deniability from an actor who participated in the record creation.

UnDetectability (UD). Most healthcare solutions use encryption (in storage, and in transit). The schemes proposed in [23,26,29,53] are an exception, where a permissioned blockchain is used to store raw data directly on-chain. This is under the assumption that only the clinician and patents will access the blockchain.

UnObservability (UO). Practices such as using a gateway or third party contact to transfer data [24,31,53], pseudonymous identities stored on public blockchain [28], claimed attributes shared in attribute-based signature generation (e.g. a clinician claiming a attribute set (Director, Hospital A, Physician) may reveal her identity if there is only one director in hospital A). The work in [23] can disclose some data which can be inferred to reveal unintended information.

Table 4
Techniques of blockchain-based EHR systems supporting solution sets from custodian's perspective and comparative analysis based on privacy properties. The schemes read as follows: ●: exhibits the property, ◑: partially exhibits the property, ○: do not exhibits the property,—refers to 'not applicable'.

Ref.	Used techniques	AN	PS	UL	PD	UD	UO	CF
[21]	1. anonymous transaction using ring signature 2. decentralised data storage 3. cryptographic functions to protect owner's data	●	—	●	○	●	●	●
[22]	1. fog nodes to handle communication among multiple healthcare IoT devices 2. Signature-Based Encryption algorithm for device identification 3. ring signatures are used to transfer keys between IoT device and fog nodes	●	—	●	○	●	●	●
[23]	1. integrity detection of EHR data 2. attribute-based signature scheme with attribute revocation	○	●	○	○	○	○	◑
[24]	1. Public key as pseudonym 2. reduce the size of transactions to be light for transmitting over BC	○	●	○	○	●	○	●
[25]	1. K-anonymity disguises the owner identity 2. searchable encryption figures out the required EHR 3. smart contract implements the attribute-based access control mechanism	○	●	○	○	●	○	●
[53]	1. biometric authorisation 2. protect the privacy and immutability of sensitive data using blockchain	○	●	○	●	○	○	◑
[26]	1. utilise hash of patient biometric as identities to ensure recoverable access on EHR in case of lost secret key 2. detect modification and activity log on EHRs	○	●	○	○	○	○	◑
[27]	1. gateway server (semi-trusted) to verify the authenticity of actions inside the system 2. proxy re-encryption for access control	○	○	○	○	●	○	●
[28]	1. Delegatable attribute-based access control 2. secure and restricted access with chain of delegation 3. access delegation are limited with owner consent 4. user generated pseudo-identities	○	●	○	○	●	○	●
[29]	1. EHR are stored in blockchain 2. OAuth like delegated access to different categories of EHR users	○	○	○	●	○	○	◑

(*continued on next page*)

Confidentiality (CF). Most systems use encryption techniques. As mentioned in undetectability, prior work [23,26,29,53] stores the raw data on permissioned blockchains. All systems employ access control schemes. Some used smart contracts as an access manager to validate all read/write requests [26,30]. Others used proxy re-encryption [27] for access control, attribute based access control [25,28] and identity-based access control [19,24,31].

6.2. Data owners perspectives

Table 5 summarises the evaluation results on the solutions of data owners'. Besides what we have summarised in Table 5, below, we include some remarks on each property.

ANonymity (AN). In [34,44] each participant generates a unique account and each health record transaction on the blockchain can be signed with a randomly generated public key. Therefore, we see that anonymity is present in these solutions by virtue of their use of pseudonymity (i.e., public key instead of real identity) and unlinkability (i.e., different public keys for different transactions).

Some solutions [38] removed PII sensitive information such as name and location from the health data. A redactable signature is used to remove certain identifying attributes in order to anonymise a health credential whilst retaining its authenticity [35]. In healthcare, such anonymisation is also desirable when unbiased second opinions are required to solve complex medical problems, while having access to relevant information, e.g. anamnesis, lab-results, etc. All of these schemes therefore stored metadata on-chain for verification purposes.

GuardHealth [41] proposes a scheme where trustworthiness of the data generator and sharer (e.g., cloud data provider as custodian, patient as owner) are calculated without having to know any PII sensitive information present within the data. The blockchain performs trust assessment, where the patient rates (ranges from 0 to 1) the trustworthiness of the cloud service provider (data custodian) and

Table 4 (*continued*).

Ref.	Used techniques	AN	PS	UL	PD	UD	UO	CF
[30]	1. predefined access permission rules for different purpose (i.e. normal/emergency) 2. limited access time using smart contract 3. access confirmation by consensus	○	●	○	●	●	○	●
[31]	1. hybrid data management approach 2. public key infrastructure-based asymmetric encryption and digital signatures to secure shared EHR data 3. access permission is stored on blocks corresponding to an ID (e.g registered clinician) for a specific time frame 4. adding new permission needs to update the ledger with another time frame	○	●	○	○	●	○	●
[19]	1. data access with owner's consent 2. data safety by encrypted health data (e.g., Elliptic Curve Cryptography (ECC)) 3. secure channel for user interaction	○	●	○	○	●	○	●

similarly the custodian can attest the reliability of the data provided by the patient.

PSeudonymity (PS). Pseudonyms-based schemes require a secure place to manage the mapping of pseudonyms and real identity. For instance, in [33,39,40,42], clinicians store signed and encrypted EHR data (including PII) in an off-chain storage (e.g., private cloud). Pseudonyms for patients are generated, and mapped with the corresponding EHR. This mapping information is stored on-chain so that its access can be managed transparently. All of these works have considered blockchain as a decentralised, single source of truth, thus they do not have to interact with third parties for verification of shared data.

UnLinkability (UL). The schemes described in [33,39] cannot avoid linkability, as the block header (on-chain) includes the signature of the doctor. In the scheme described in [38], an access token containing the owner's signature is used to share data. If the owner uses the same access token in multiple data sharing scenarios, unlinkability is not guaranteed due to pattern detection.

In contrast, a redactable signature adopted in [35] is unlinkable, because the signature generation depends on the data content [58]. ZKP is another solution [36] to ensure anonymous authentication, where small pieces of unlinkable information are accumulated to verify the assertions without disclosing the content [59]. For example, during passport authentication, name on the passport which is signed by the issuing government are disclosed and verified, but the remaining attributes (e.g., date of birth, passport number) can remain undisclosed. In this case, the globally unique signature value is kept hidden to avoid the linkability. Therefore, no one can establish the link between multiple showing of same data (i.e., passport name signed by government X).

Plausible deniability (PD). In most blockchain-based systems [33–35, 37,40–42], the originating data provider (either patient or clinician) signs health transaction records, which may make it difficult to deny the action later on.

Verifiable credentials in [44] are signed using the public keys associated with the relevant identifier, therefore the signer may not be able to deny the action if the public key to the identifier mapping is revealed.

UnDetectability (UD). In all the solution sets from the owner's side, the data is encrypted and stored on a private storage. Valid custodians can use the decryption key to access this data.

UnObservability (UO). In some solutions, even if undetectability is present, not having anonymity could lead to not having unobservability. For instance, in [33], for each EHR created on the blockchain, the signature of the doctor is included. The doctor uses the same signature for all of her EHR. In [39,40], the EHR includes patient's identity. In both cases, analysing access logs (which are stored on-chain) can reveal the identities relating to EHRs (e.g., the same doctor's signature can reveal the identity).

Confidentiality (CF). From the owner's perspective, since all of the systems employ encryption, the main aspect to consider is access control. The access control of encrypted data is essentially achieved through managing access to the decryption key [33,36,38] (i.e. decryption key is shared with a custodian to grant access to the data). Proxy re-encryption (PRE) is one technique for access control which includes delegation of data sharing rights [37,39,41].

BPDS [34] used attribute based encryption (ABE)-based access control. In [42], the data owner can use a smart contract and access policy list to grant access. The policy list contains the public keys of the access grantees that the owner wishes to share data with. The smart contract then verifies the public keys and grants access accordingly.

7. Summary and research directions

This section presents the response to (Q4) where we synthesise the overall summary of survey findings, our observations from the survey and propose research directions.

7.1. Summary of data minimisation techniques

In this subsection, we summarise the findings from Section 4. We organise the solution sets from Section 4 in Fig. 2 according to the data processing stages. We have depicted the processing stages in a tree structure both from custodian side (i.e., when data resides with custodians) and owner side (i.e., when data resides with owners).

In Section 4.1, we presented data custodian perspective in collection, using (purpose), storage/access stages. In Collection stage, we identified some data minimisation techniques that are designed to collect legitimate data and minimise the inclusion of PII. Next in Using Data stage, we observed whether a consent action is manifested (e.g. owner gives consent to his clinician to share EHR for different purposes) in the system. We noted that the consent management at the custodian side is 'static' in that owners provide consent of all

Table 5
Techniques of blockchain-based EHR systems supporting solution sets from owner's perspective and comparative analysis based on privacy properties. The schemes read as follows: ●: exhibits the property, ◑: partially exhibits the property, ○: do not exhibits the property, ⊙: exhibits the property, because it is implied by other properties, —refers to 'not applicable'.

Ref.	Used techniques	AN	PS	UL	PD	UD	UO	CF
[33]	1. attribute based signature for effective verification of signer attribute without exposing the identity of the signer 2. attributes are associated with owner's global identity 3. verification of data authenticity and signer's identity	○	●	○	○	●	○	●
[34]	1. consortium blockchain stores data indexes with the list of authorised data users 2. content extraction algorithm to hide sensitive information (e.g., patients' name or ID number) from health data 3. separate pseudonym on each transaction	⊙	●	●	○	●	●	●
[35]	1. anonymity by redacting identifiable information 2. source authenticity of original data source is ensured without any interaction with the original signer	●	—	●	○	●	●	●
[36]	1. ZKP based device authentication system to identify legitimate wearable device 2. lightweight ZKP to run on mHealth devices with minimum resources 3. owner provides the decryption key to legitimate custodian	●	—	●	○	●	●	●
[37]	1. proxy cloud server for managing re-encryption to delegate access to the custodian 2. automated insurance claims using smart contracts 3. ZKP-based patient identity authentication system	○	●	○	○	●	○	●
[38]	1. owner issues permission token per access request 2. key management using wallet service 3. token-based verification to grant one time access to data 4. generate a fingerprint for each data access to make each action traceable	●	—	○	○	●	●	●

(*continued on next page*)

anticipated use of the data once at the start of collection and future consent for possible re-purpose is rarely sought by the custodians. We believe this lack of schemes that track and manage consent is one of the major oversights in current privacy practices. We introduced a few examples of managing consent for access delegation. For Storing stage, we highlighted some examples that use permissioned blockchain to form a trusted consortium of health clinics to store and share EHR. Although, we did not find any solution that addressed EHR retention management from the custodian side, we included a solution that allowed removal of EHR at the request of the owner.

In Section 4.2, we have presented data owner perspective in sharing, using (purpose), storage/access stages. In Data Sharing stage, we identified solutions that allow owners to decide what to share and with whom. Existing solutions include atomic credential generation, selective content extraction and zero knowledge proof to implement *selective disclosure* from EHR. In these solutions, blockchains act as a reliable and trustworthy intermediary between owners and custodians. During Using Data stage, we considered the issue of consent management. Unlike the custodians' view, when the data is with the user, consenting per access request (i.e., per data use/purpose) is possible. For instance, a token-based access allows the owner to grant access in a fine-grained manner (e.g., one-time only). Storing the access traces on blockchain can help track and analyse the access patterns to detect problems (e.g., unintended sharing). Blockchain can also play an intermediary role in delegating access as a third party "re-encryption key manager" where the shared re-encryption key to a delegatee can be revoked after one-time use through the manager. For Storing stage, we identified the solutions where owners store EHR on their personal storage (e.g. private cloud), but sharing a copy of the data is managed through blockchains and smart contracts. An index of data (i.e., available EHRs) and public keys of custodians approved for access for particular EHR are stored on the blockchain. We discussed self-sovereignty in patient identity. Emerging Web standards such as Decentralised Identifier (DID) and Verifiable Credentials, often combined with blockchains as a trusted mediator of information sharing, can provide a decentralised identity and credential management framework. Such a framework removes the need for a central authority, giving patients full control over their identity. As a solution to manage data retention, we described how a set of independently running data access gateway applications may enforce removal of data after the specified retention period.

Table 5 (*continued*).

Ref.	Used techniques	AN	PS	UL	PD	UD	UO	CF
[39]	1. Controlled access mechanism using smart contract 2. adoption of proxy re-encryption for granting and revoking data access rights 3. Redactable patient blocks, by way of chameleon hashing to minimise data fragmentation, allow for in-place editing, and reduce resource consumption 4. patient block contains plaintext metadata, encrypted patient data and smart contract	○	●	○	○	●	○	●
[40]	1. Data integrity by storing hash of EHR in a permissioned blockchain, 2. another private chain 3. owners have control over their data as they can define and manage the access policies over their data	○	●	○	○	●	○	●
[41]	1. Proxy Re-encryption, to dynamically allow requester to access data and revoke permissions easily 2. smart contract to achieve secure and efficient data storage and sharing 3. revocability of access consent by the owner 4. trust assessment mechanism to improve the reliability of anonymous data	●	—	●	○	●	●	●
[42]	1. decentralised IPFS storage on a mobile cloud platform 2. role-based access control 3. smart contracts can identify, validate request and grant access permissions by triggering transactions or messages.	○	●	○	○	●	○	●
[44]	1. self sovereign identity (SSI) model for data management 2. check and change the status of issued credentials through smart contract 3. high storage throughput 4. different identifiers for different transactions protect personal data	⊙	●	●	○	●	●	●
[48]	1. owner can manage their EHR securely through their own gateways 2. MPC to allow computation on encrypted data without sharing raw data to untrusted requester 3. replicas of data is shared to data requester that can be deleted after retention period	●	—	●	○	●	●	●

7.2. Limitations and potentials of blockchain in healthcare

In this subsection, we discuss the limitations and potential usage of blockchain in the context of data minimisation particularly in healthcare domain.

The benefits of using blockchains for healthcare data management, compared to conventional methods of using cloud and database systems, include: i) high availability of the systems through decentralisation, ii) tamper-proof storage of data for improved security and trust, and (iii) traceability of data for improved data provenance and auditability. All of these are critical in improving the security and privacy of EHR-based healthcare systems. We observed that, with blockchain technology, EHRs will be owned and controlled by the true owner (i.e., the patient). The owners will have the right to decide who can and cannot access their EHR and for what purpose and blockchain can provide the assurance of transparency and auditability in data access by the custodians. However, there are still several open challenges.

Firstly, because the blockchain transactions are immutable and permanent, there are potential issues with managing data. For instance, take the issue of data quality. Human errors can occur in any application (e.g., wrong input), but in blockchain, correcting errors could be more costly, sometimes not practical. Complying to data retention regulation may be another issue. A privacy regulation such as GDPR may require that data to be removed from the system after a certain period. In this regard, there are some solutions in other domains. In supply chain management, the data quality issue is addressed by reputation management techniques [60]. Also, in terms of correcting errors, there are other possible solutions such as editable blockchain protocol, where edit operations are performed once approved by the blockchain policy (e.g., voted by the majority, validated by the miners etc.) [61,62].

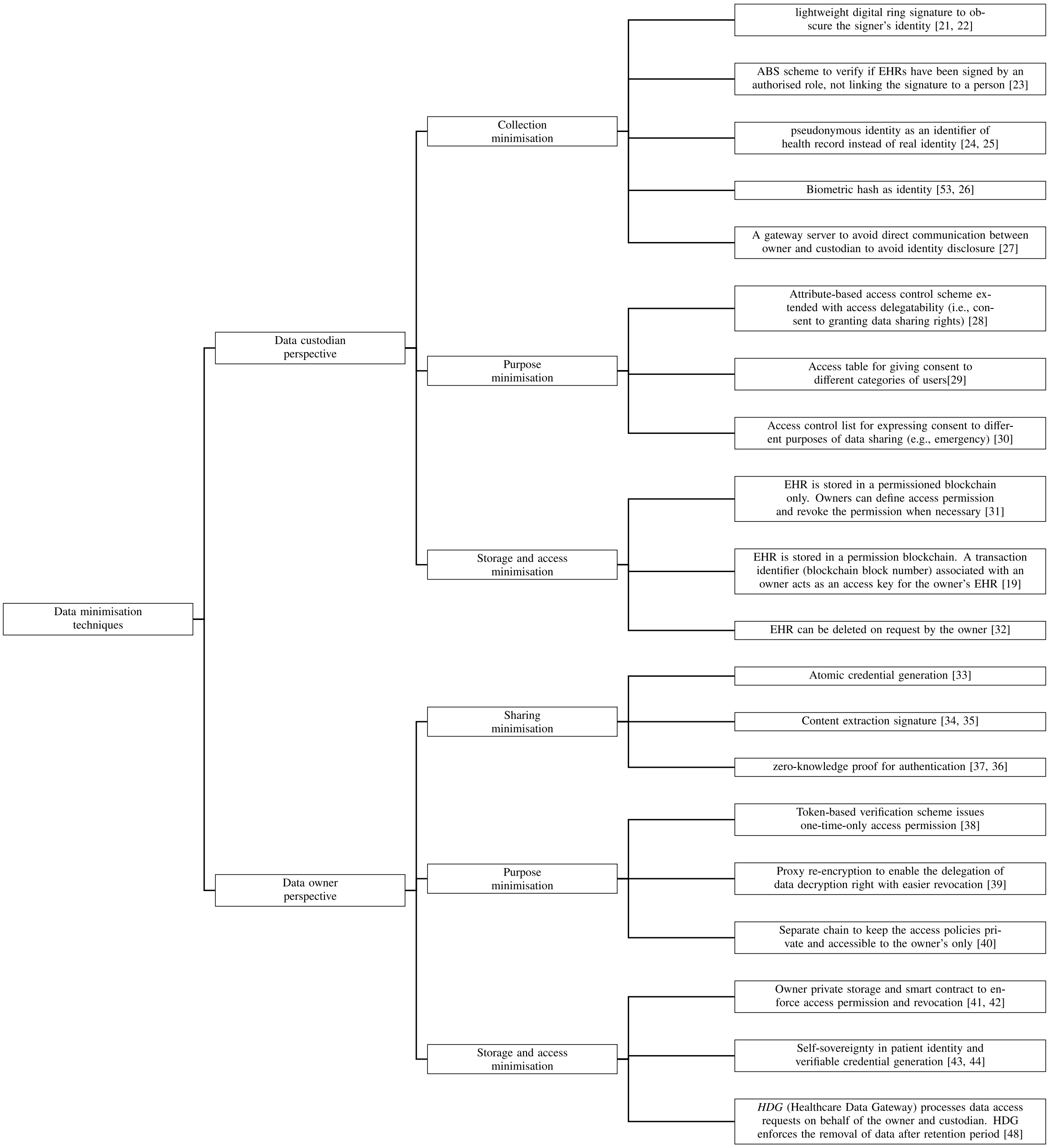

Data minimisation techniques
Data custodian perspective
Collection minimisation
lightweight digital ring signature to obscure the signer's identity [21, 22]
ABS scheme to verify if EHRs have been signed by an authorised role, not linking the signature to a person [23]
pseudonymous identity as an identifier of health record instead of real identity [24, 25]
Biometric hash as identity [53, 26]
A gateway server to avoid direct communication between owner and custodian to avoid identity disclosure [27]
Purpose minimisation
Attribute-based access control scheme extended with access delegatability (i.e., consent to granting data sharing rights) [28]
Access table for giving consent to different categories of users[29]
Access control list for expressing consent to different purposes of data sharing (e.g., emergency) [30]
Storage and access minimisation
EHR is stored in a permissioned blockchain only. Owners can define access permission and revoke the permission when necessary [31]
EHR is stored in a permission blockchain. A transaction identifier (blockchain block number) associated with an owner acts as an access key for the owner's EHR [19]
EHR can be deleted on request by the owner [32]
Data owner perspective
Sharing minimisation
Atomic credential generation [33]
Content extraction signature [34, 35]
zero-knowledge proof for authentication [37, 36]
Purpose minimisation
Token-based verification scheme issues one-time-only access permission [38]
Proxy re-encryption to enable the delegation of data decryption right with easier revocation [39]
Separate chain to keep the access policies private and accessible to the owner's only [40]
Storage and access minimisation
Owner private storage and smart contract to enforce access permission and revocation [41, 42]
Self-sovereignty in patient identity and verifiable credential generation [43, 44]
HDG (Healthcare Data Gateway) processes data access requests on behalf of the owner and custodian. HDG enforces the removal of data after retention period [48]

Fig. 2. Classification of minimisation techniques in blockchain-based healthcare systems.

Secondly, cross-border sharing of EHR where different and often conflicting regulations may hinder the benefit of blockchain data sharing. Indeed, the expectation of an individual's privacy varies from one country to another based on government regulations, even within the same country (e.g., cross different states). Therefore, future research on regulation, standardisation, and cross-border EHR retrieving policies including retention and usage purpose are duly urgent.

Thirdly, in the healthcare literature, private and permissioned blockchain are mostly used to avoid public visibility, where all the blockchain users are known to one another and users must be authorised to view and add to the blockchain [19,23,24,27,30,31,39,42,48]. Further, some users may have approved access to only a portion of the ledger. In consequence, it may become difficult to query expected data within a blockchain, limiting clinical, statistical and research uses of data.

Finally, mobile health applications are becoming more commonplace [36]. The problem arises because of high heterogeneity in communication protocols, medical devices, and platforms. Securely managing access control across these varying protocols and platforms could present a challenge. Blockchain can provide a single source trusted mediator that enables seamless communication amongst the heterogeneous platforms.

This study found that most published research on the use of blockchain in the health sector presents theoretical frameworks, architectures, or models with few technical details. However, there is seldom a prototype or pilot implementation to learn from. Some of the prototype implementations performed qualitative security analysis to evaluate their system [19,21,24,26,28,29,31,34,48]. The others evaluated their system using different metrics [22,23,25,27,30,33,36–42,44]. Among them, most papers discussed transaction latency and throughput. Also, due to the variation in the test environments, observation points could be varied. We see that implementation of blockchain systems in settings that are more close to the practical scenarios, and perhaps more real-world use cases are necessary so that we can perform careful assessment of the potential benefits and cost of implementing and running blockchain-based systems in healthcare applications.

7.3. Research directions

Here, we discuss some future research challenges.

Complete anonymity vs. Data utility. In health data storage, it could be argued that complete anonymisation may hinder the benefits of data analysis in some situations (e.g., detecting the actual person who may be spreading any contagious disease if his movement is not restricted immediately). It is challenging to achieve a good trade-off between privacy preservation and data utility.

Continuous purpose tracking and consent management. According to the GDPR article 5(1)(c), organisations that process personal data (i.e., data custodians) must practice *purpose limitation*. This principle is recognised as an ethical standard for data custodians. To the best of our knowledge, there is no technical solution for continuously tracking the purpose of data use in healthcare after its initial collection. As a consequence, it is not always possible to confirm that the personal data is used in accordance with the collected purpose. Porsdam et al. [63] demonstrated that a machine learning-based consent and access analysis using the traces generated by blockchain/smart contracts can detect if the data accessed is relevant for the consent given for that data. Additionally, artificial intelligence can be applied to design automatic generation of smart contacts to enhance secure and flexible data access operations. However, in case of healthcare research, sometimes it is not possible to fully identify the purpose of personal data processing at the time of collection [64].

Blockchain data privacy. Because the blockchain data is immutable, we must define which type of data can and should be stored in blockchain. Firstly, for proper functioning, blockchain architectures require that the identifiers (i.e. public keys) of participating entities must be visible. Thus the identifiers cannot be minimised further or stored in an off-chain storage. Secondly, EHRs could be large in volume and it may not be scalable to store them directly on the blockchain. Some solutions store EHRs off-chain and only the proof of data (e.g., certain claims about the data, a hash of EHRs) is stored on-chain. This allows the participants to check the authenticity and accuracy of the off-chain EHRs [65]. Thirdly, if it is required to store some data on-chain, encrypted data or hashed data without the key can be stored in blockchain [66]. However, this may generate another challenge, as metadata associated with EHRs is also potentially PII [67].

Responsible self-governance. The solutions that gives the owners more control over their data can create new challenges in terms of the owners practising responsible self-management of data (e.g., not generating fake or adversarial data). Furthermore, self-governance requires the owner to manage their private/public key pairs in order to provide cryptographic signatures and authorise access to their medical data. However, the underlying complexity of managing the keys should be hidden behind a sufficiently user-friendly interface, likely deployed as a web and/or mobile application. However, this also opens the door to potential security threats associated with the key management mechanisms. Another challenge with self-governance is when the patient is unable to grant the necessary access permits. The reasons could range from simply losing the private key, an illness that incapacitated the person (e.g., Alzheimer), to an acute medical emergency. To deal with such cases, the patient must setup and manage a precise and transparent chain of access delegation and responsibility.

8. Conclusion

Our study showed that blockchain has a great potential in transforming the conventional healthcare systems into fully digitised and streamlined data sharing platform with increased privacy protection, and more control for the patients (data owners). In this paper, we reviewed and discussed data minimisation techniques from these two opposing sides: data custodians and owners. Then, we identified a number of potential research opportunities. This analysis can serve as a primer for adopting data minimisation in the healthcare sector and offer insights into architectural design choices that need to be considered. Moreover, we hope this review will contribute to further insight into the development and implementation of the next generation healthcare systems, which will benefit the owners in terms of maintaining personal privacy while simultaneously meeting the custodian's needs.

CRediT authorship contribution statement

Rahma Mukta: Conceptualization, Investigation, Methodology, Writing – original draft. **Hye-young Paik:** Conceptualization, Methodology, Data curation, Writing – original draft. **Qinghua Lu:** Conceptualization, Writing – review & editing. **Salil S. Kanhere:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported through an ''Australian Government Research Training Program Scholarship'' funded by ''Commonwealth Department of Education and Training''.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.comnet.2022.108766.

References

[1] J. Davis, UPDATE: The 10 Biggest Healthcare Data Breaches of 2020, So Far, Tech. Rep., Health IT Security[online], 2020, URL https://healthitsecurity.com/news/the-10-biggest-healthcare-data-breaches-of-2020-so-far.

[2] A.H. Seh, M. Zarour, M. Alenezi, A. Sarkar, A. Agrawal, R. Kumar, R.A. Khan, Healthcare data breaches: Insights and implications, Healthcare 8 (2) (2020).

[3] S. Haas, S. Wohlgemuth, I. Echizen, N. Sonehara, G. Müller, Aspects of privacy for electronic health records, Int. J. Med. Inform. 80 (2) (2011) e26–e31, (Special Issue: Security in Health Information Systems).

[4] A. Grando, J. Ivanova, M. Hiestand, H. Soni, A. Murcko, M. Saks, D. Kaufman, M.J. Whitfield, C. Dye, D. Chern, J. Maupin, Mental health professional perspectives on health data sharing: Mixed methods study, Health Inform. J. 26 (3) (2020) 2067–2082.

[5] S. Karagiannis, E. Magkos, Decentralized internet privacy: Towards a blockchain framework for healthcare, in: 11th Mediterranean Conference on Information Systems, 2017.

[6] Q. Ramadan, D. Strüber, M. Salnitri, J. Jürjens, V. Riediger, S. Staab, A semi-automated BPMN-based framework for detecting conflicts between security, data-minimization and fairness requirements, Softw. Syst. Model. (2020).

[7] A. Pfitzmann, M. Hansen, A Terminology for Talking About Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management, Tech. Rep., TU Dresden and ULD Kiel, 2010.

[8] M.A. Ferrag, M. Derdour, M. Mukherjee, A. Derhab, L. Maglaras, H. Janicke, Blockchain technologies for the internet of things: Research issues and challenges, IEEE Internet Things J. 6 (2) (2019) 2188–2204.

[9] M.A. Ferrag, L. Shu, The performance evaluation of blockchain-based security and privacy systems for the internet of things: A tutorial, IEEE Internet Things J. (2021) 1.

[10] H.-N. Dai, Z. Zheng, Y. Zhang, Blockchain for internet of things: A survey, IEEE Internet Things J. 6 (5) (2019) 8076–8094.

[11] K. Peng, M. Li, H. Huang, C. Wang, S. Wan, K.-K.R. Choo, Security challenges and opportunities for smart contracts in internet of things: A survey, IEEE Internet Things J. 8 (15) (2021) 12004–12020.

[12] J.J. Hathaliya, S. Tanwar, An exhaustive survey on security and privacy issues in Healthcare 4.0, Comput. Commun. 153 (2020) 311–335.

[13] X. Zhu, J. Shi, C. Lu, Cloud health resource sharing based on consensus-oriented blockchain technology: Case study on a breast tumor diagnosis service, J. Med. Internet Res. 21 (7) (2019) e13767.

[14] I. Abu-elezz, A. Hassan, A. Nazeemudeen, M. Househ, A. Abd-alrazaq, The benefits and threats of blockchain technology in healthcare: A scoping review, Int. J. Med. Inform. 142 (2020) 104246.

[15] H. Jin, Y. Luo, P. Li, J. Mathew, A review of secure and privacy-preserving medical data sharing, IEEE Access 7 (2019) 61656–61669.

[16] V. Ferrari, EU Blockchain Observatory and Forum Workshop on GDPR, Data Policy and Compliance, Tech. Rep., Amsterdam Law School, 2018, Research Paper No. 2018-23, http://dx.doi.org/10.2139/ssrn.3247494.

[17] D. Chadwick, D. Longley, M. Sporny, O. Terbu, D. Zagidulin, B. Zundel, Verifiable credentials implementation guidelines 1.0, 2019, https://www.w3.org/TR/vc-imp-guide/#data-minimization.

[18] Office for Civil Rights, Collection, Use and Disclosure Limitation, Tech. Rep., The Nationwide Privacy and Security Framework for Electronic Exchange of Individually Identifiable Health Information, 2003.

[19] A. Al Omar, M.S. Rahman, A. Basu, S. Kiyomoto, Medibchain: A blockchain based privacy preserving platform for healthcare data, in: G. Wang, M. Atiquzzaman, Z. Yan, K.R. Choo (Eds.), Security, Privacy, And Anonymity In Computation, Communication, And Storage, Springer International Publishing, 2017, pp. 534–543.

[20] A. Senarath, N. Arachchilage, A data minimization model for embedding privacy into software systems, Comput. Secur. 87 (2019) 101605.

[21] A. Dwivedi, G. Srivastava, S. Dhar, R. Singh, A decentralized privacy-preserving healthcare blockchain for IoT, Sensors 19 (2) (2019) 326.

[22] S. Shukla, S. Thakur, S. Hussain, J.G. Breslin, S.M. Jameel, Identification and authentication in healthcare internet-of-things using integrated fog computing based blockchain model, Internet Things 15 (2021) 100422.

[23] Q. Su, R. Zhang, R. Xue, P. Li, Revocable attribute-based signature for blockchain-based healthcare system, IEEE Access 8 (2020) 127884–127896.

[24] K.M. Hossein, M.E. Esmaeili, T. Dargahi, A. khonsari, Blockchain-based privacy-preserving healthcare architecture, in: IEEE CCECE, 2019, pp. 1–4.

[25] H.Z. Y. Chen, G. Xue, A blockchain-based medical data sharing mechanism with attribute-based access control and privacy protection, 2021, pp. 1–12,

[26] V. Ramani, T. Kumar, A. Bracken, M. Liyanage, M. Ylianttila, Secure and efficient data accessibility in blockchain based healthcare systems, in: GLOBECOM, 2018, pp. 206–212.

[27] T.T. Thwin, S. Vasupongayya, P. Gope, Blockchain-based access control model to preserve privacy for personal health record systems, Sec. Commun. Netw. (2019).

[28] H.S. Gardiyawasam Pussewalage, V.A. Oleshchuk, Blockchain based delegatable access control scheme for a collaborative E-health environment, in: IEEE IThings and IEEE GreenCom and IEEE CPSCom and IEEE Smart Data, 2018, pp. 1204–1211.

[29] T. Rathee, P. Singh, Medsecurechain: Applying blockchain for delegated access in health care, in: R. Kountchev, R. Mironov, S. Li (Eds.), New Approaches For Multidimensional Signal Processing, Springer Singapore, Singapore, 2021, pp. 153–163.

[30] A.R. Rajput, Q. Li, M. Taleby Ahvanooey, I. Masood, EACMS: Emergency access control management system for personal health record based on blockchain, IEEE Access 7 (2019) 84304–84317.

[31] A. Dubovitskaya, F. Baig, Z. Xu, R. Shukla, P.S. Zambani, A. Swaminathan, M.M. Jahangir, K. Chowdhry, R. Lachhani, N. Idnani, M. Schumacher, K. Aberer, S.D. Stoller, S. Ryu, F. Wang, ACTION-EHR: Patient-centric blockchain-based electronic health record data management for cancer care, J. Med. Internet Res. 22 (8) (2020) e13598.

[32] H. Yang, B. Yang, A blockchain-based approach to the secure sharing of healthcare data, in: Norwegian Information Security Conference, 2017.

[33] Y. Sun, R. Zhang, X. Wang, K. Gao, L. Liu, A decentralizing attribute-based signature for healthcare blockchain, in: 27th ICCCN, 2018, pp. 1–9.

[34] J. Liu, X. Li, L. Ye, H. Zhang, X. Du, M. Guizani, BPDS: A blockchain based privacy-preserving data sharing for electronic medical records, in: GLOBECOM, 2018, pp. 1–6.

[35] D. Slamanig, S. Rass, Generalizations and extensions of redactable signatures with applications to electronic healthcare, in: Communications and Multimedia Security, Springer, 2010, pp. 201–213.

[36] A.E.B. Tomaz, J.C.D. Nascimento, A.S. Hafid, J.N. De Souza, Preserving privacy in mobile health systems using non-interactive zero-knowledge proof and blockchain, IEEE Access 8 (2020) 204441–204458.

[37] B. Sharma, R. Halder, J. Singh, Blockchain-based interoperable healthcare using zero-knowledge proofs and proxy re-encryption, in: COMSNETS, 2020, pp. 1–6.

[38] X. Liang, S. Shetty, J. Zhao, D. Bowden, D. Li, J. Liu, Towards decentralized accountability and self-sovereignty in healthcare systems, in: Information And Communications Security, Springer International Publishing, 2018, pp. 387–398.

[39] R.H. Hylock, X. Zeng, A blockchain framework for patient-centered health records and exchange (HealthChain): Evaluation and proof-of-concept study, J. Med. Internet Res. 21 (8) (2019).

[40] K. Mohammad Hossein, M.E. Esmaeili, T. Dargahi, A. Khonsari, M. Conti, BCHealth: A novel blockchain-based privacy-preserving architecture for IoT healthcare applications, Comput. Commun. 180 (2021) 31–47.

[41] Z. Wang, N. Luo, P. Zhou, GuardHealth: Blockchain empowered secure data management and Graph Convolutional Network enabled anomaly detection in smart healthcare, J. Parallel Distrib. Comput. 142 (2020) 1–12.

[42] D.C. Nguyen, P.N. Pathirana, M. Ding, A. Seneviratne, Blockchain for secure EHRs sharing of mobile cloud based E-health systems, IEEE Access 7 (2019) 66792–66806.

[43] B. Houtan, A.S. Hafid, D. Makrakis, A survey on blockchain-based self-sovereign patient identity in healthcare, IEEE Access 8 (2020) 90478–90494.

[44] S. Matteo, T. Andrea, Pistis: a credentials management system based on self-sovereign identity, in: Computer Science and Engeneering (Master thesis), Politecnico Di Milano, 2019-20.

[45] Gem, Health — Gem, 2017, https://enterprise.gem.co/health/. (Accessed 15 February 2021).

[46] Medrec, What is MedRec? 2016, https://medrec.media.mit.edu/technical/. (Accessed 15 February 2021).

[47] L. Hendren, K. Kuzmeskas, Health nexus, 2018, https://crushcrypto.com/wp-content/uploads/2018/03/HLTH-Whitepaper.pdf. (Accessed 15 February 2021).

[48] X. Yue, H. Wang, D. Jin, M. Li, W. Jiang, Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control, J. Med. Syst. 40 (10) (2016) 1–8.

[49] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, W. Joosen, A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements, Requir. Eng. 16 (1) (2011) 3–32.

[50] K.C. Saunders B, Anonymising interview data: challenges and compromise in practice, Qual. Res.: QR 15 (5) (2015) 616–632.

[51] T.C. Buchmueller, C. Carey, The effect of prescription drug monitoring programs on opioid utilization in medicare, Am. Econ. J.: Econ. Policy 10 (1) (2018) 77–112.

[52] N. Mamo, G.M. Martin, M. Desira, B. Ellul, J.-P. Ebejer, Dwarna: a blockchain solution for dynamic consent in biobanking, Eur. J. Human Genet.: EJHG 28 (5) (2020) 609–626.

[53] D. Dhagarra, M. Goswami, P. Sarma, A. Choudhury, Big data and blockchain supported conceptual model for enhanced healthcare coverage: The Indian context, Bus. Process Manag. 25 (7) (2019) 1612–1632.

[54] M. Arapinis, T. Chothia, E. Ritter, M. Ryan, Analysing unlinkability and anonymity using the applied pi calculus, in: Computer Security Foundations Symposium, CSF, IEEE, US, 2010, pp. 107–121.

[55] A. Machanavajjhala, J. Gehrke, D. Kifer, M. Venkitasubramaniam, L-diversity: privacy beyond k-anonymity, in: 22nd International Conference On Data Engineering, ICDE'06, 2006, p. 24.

[56] L. Sweeney, Simple demographics often identify people uniquely, 2000.

[57] R.L. Rivest, A. Shamir, Y. Tauman, How to leak a secret, in: C. Boyd (Ed.), Advances in Cryptology, Springer, 2001, pp. 552–565.

[58] O. Sanders, Efficient redactable signature and application to anonymous credentials, in: A. Kiayias, M. Kohlweiss, P. Wallden, V. Zikas (Eds.), Public-Key Cryptography, Springer, Cham, 2020, pp. 628–656.

[59] R. Henry, Efficient Zero Knowledge Proofs and Applications (Ph.D. thesis), University of Waterloo, 2014.

[60] S. Malik, V. Dedeoglu, S.S. Kanhere, R. Jurdak, Trustchain: Trust management in blockchain and IoT supported supply chains, in: IEEE ICBC, 2019, pp. 184–193.

[61] D. Deuber, B. Magri, S. Thyagarajan, Redactable blockchain in the permissionless setting, 2019, CoRR arXiv:1901.03206.

[62] A. Dorri, S.S. Kanhere, R. Jurdak, MOF-BC: a memory optimized and flexible BlockChain for large scale networks, 2018, CoRR arXiv:1801.04416.

[63] S.P. Mann, J. Savulescu, P. Ravaud, M. Benchoufi, Blockchain, consent and prosent for medical research, J. Med. Ethics (2020).

[64] D.J. Hand, Aspects of data ethics in a changing world: Where are we now? Big Data 6 (3) (2018) 176–190.

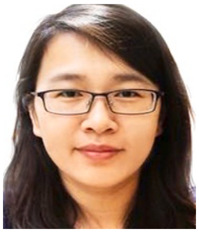
[65] E. Politou, F. Casino, E. Alepis, C. Patsakis, Blockchain mutability: Challenges and proposed solutions, 2019, CoRR arXiv:1907.07099.

[66] M.T. de Oliveira, L.H.A. Reis, R.C. Carrano, F.L. Seixas, D.C.M. Saade, C.V. Albuquerque, N.C. Fernandes, S.D. Olabarriaga, D.S.V. Medeiros, D.M.F. Mattos, Towards a blockchain-based secure electronic medical record for healthcare applications, in: IEEE ICC, 2019, pp. 1–6.

[67] F. Karegar, C. Striecks, S. Krenn, F. Hörandner, T. Lorünser, S. Fischer-Hübner, Opportunities and challenges of CREDENTIAL - towards a metadata-privacy respecting identity provider, in: Privacy And Identity Management, in: IFIP Advances in Information and Communication Technology, Springer, 2016, pp. 76–91.

Rahma Mukta is a Ph.D. student at University of New South Wales (UNSW), Sydney, Australia. Her research interest includes blockchain applications, decentralised identity management, verifiable credentials, and privacy challenges in e-health.

Hye-young Paik received the Ph.D. degree in computer science from the University of New South Wales (UNSW), Sydney Australia. She is currently a Senior Lecturer in the School of Computer Science and Engineering, UNSW. Her research interests include process automation, distributed software systems and architectures, privacy, and security in distributed systems. She is a member of the IEEE and the ACM.



Qinghua Lu is a senior research scientist at CSIRO's Data61, Australia. Before she joined Data61, she was an associate professor at China University of Petroleum. She formerly worked as a researcher at NICTA (National ICT Australia). She received her Ph.D. from University of New South Wales in 2013. Her recent research interest includes software architecture, blockchain, and software engineering for machine learning. She has published 100+ academic papers in international journals and conferences. She is an IEEE senior member.

Salil S. Kanhere is a Professor in the School of Computer Science and Engineering at UNSW Sydney, Australia. He received his MS and Ph.D. in Electrical Engineering from Drexel University, Philadelphia. His research interests include pervasive computing, Internet of Things, cyber–physical systems, blockchain, cybersecurity and applied machine learning. He has published over 300 peer-reviewed articles and delivered over 50 keynote talks and tutorials on these topics. He has received 8 Best Paper awards. He has co-authored a book on Blockchain for Cyber–physical Systems published by Artech House in 2020. He is a contributing research staff at CSIRO's Data61 and has held visiting positions at the Institute of Infocomm Research Singapore, Technical University Darmstadt, University of Zurich, and Graz University of Technology. Salil is an ACM Distinguished Speaker, an IEEE Computer Society Distinguished Visitor and Senior Member of the IEEE and ACM. He received the Friedrich Wilhelm Bessel Research Award (2020) and the Humboldt Research Fellowship (2014) from the Alexander von Humboldt Foundation in Germany. He is Editor in Chief of Ad Hoc Networks and Area Editor for IEEE Transactions on Network Management and Service, Pervasive and Mobile Computing, Computer Communications, and International Journal of Network Management. He regularly serves on the organising committee of several IEEE and ACM international conferences (examples include PerCom, MobiSys, CPS-IOT Week, WoWMoM, LCN, MSWiM, ICBC).