EAI Endorsed Transactions
on Pervasive Health and Technology

Research Article EAI.EU

Towards Blockchain Empowered Trusted and Accountable Data Sharing and Collaboration in Mobile Healthcare Applications

Xueping Liang[1,*], Sachin Shetty[1], Deepak Tosh[2], Daniel Bowden[3], Laurent Njilla[4], Charles Kamhoua[5]

[1] Virginia Modeling Analysis and Simulation Center, Old Dominion University, Norfolk, VA
[2] Department of Computer Science, University of Texas at El Paso, El Paso, TX
[3] Sentara Healthcare, Norfolk, VA
[4] Cyber Assurance Branch, Air Force Research Laboratory, Rome, NY
[5] Network Security Branch of the U.S. Army Research Laboratory, Adelphi, MD

Abstract

Enabled by mobile and wearable technology, personal health data delivers immense and increasing value for healthcare, benefiting both care providers and medical research. The secure and convenient sharing of personal health data is crucial to the improvement of the interaction and collaboration of the healthcare industry. Faced with the potential privacy issues and vulnerabilities existing in current personal health data storage and sharing systems, as well as the blockchain integration concerns summarized in this paper, an innovative user-centric health data sharing solution by utilizing a decentralized but permissioned blockchain is proposed to protect privacy and enhance access management, with the help of channel formation scheme supported by the blockchain. By developing a web application for Personal Health Data Management (PHDM) systems, the individuals are capable of synchronizing sensor data from wearable devices with online account and controlling data access from any third parties. A mobile application is deployed to collect health data from personal wearable devices, manual input, and medical devices, and synchronize data to the cloud for data sharing with healthcare providers and health insurance companies. To preserve the integrity of health data, a proof of integrity and validation, is made available to each record, which is permanently retrievable from cloud database and is anchored to the blockchain network. Moreover, for scalable and performance considerations, a tree-based data processing and batching method is adopted to deal with large data sets of personal health data collected and uploaded by the mobile platform. To enable a trusted data access record, the Intel Software Extensions technology is utilized to ensure the accountability for data access and token based access control scheme is enhanced with the trusted hardware. Analysis shows that the proposed approach provides user privacy and accountability with acceptable overhead and scalability.

Keywords: Healthcare, eHealth, Privacy, Permissioned Blockchain, Access Control, Scalability, Integrity, Wearable Devices, Mobile Platform, Privacy Protection, Self-Sovereignty, Trusted Computing, Decentralization, Intel SGX, Accountability.

Received on 01 June 2018, accepted on 10 June 2018, published on 30 July 2018

doi: 10.4108/eai.24-7-2018.159338

*Corresponding author. Email: liangxueping2015@gmail.com

1. Introduction

In recent years, the rising of wearable technology and the Internet-of-Things (IoT) has brought great opportunities and challenges to the healthcare domain. Wearable technology

refers to networked devices embedded with sensors (e.g., heart rate, blood pressure, pedometer) which can be worn comfortably on the body to collect health data and tracking activities [1]. Enabled by cloud computing and big data analytics, those data collected from individual devices contributes to big health data and valuable insights can be derived. Hospitals and medical institutions can use these data to link with other Electronic Health Record (EHR) data, such as clinical notes, to facilitate health monitoring, disease diagnoses and treatment. Health insurance companies can make detailed and strategic policies according to individual characteristics, benefiting customers to choose flexible insurance plans according to their needs.

To handle sharing and integrating health data between institutions, a significant infrastructure is crucial. However, there are several challenges related to privacy, security, and interoperability facing current health care systems and cloud-based infrastructures. First, health data are highly privacy-sensitive, especially as more data are storing in a public cloud, raising the risks and concerns of data expose and leakage. Many current research and approaches targets at improving data providers' responsibilities to detect the data leakage activities, however, it is also crucial to protect data and reduce the data leakage risks in the infrastructure design. Second, current systems use centralized architecture, which increases the security risk footprint, and requires centralized trust in a single authority. Moreover, the effective integration of health data and the interoperability between healthcare systems remain a challenging task. Over 300 different EHR systems are in use today, but there are little or even no communication and cooperation among systems [2], resulting in the lack of a holistic and thorough view of personal health. Another challenge is that users have little control over their personal health data [3]. It is reported that 62% of insured adults rely on their doctors to manage their health records [2], which limits their ability to interact with other healthcare providers than their primary doctor. With the notion of Self-Sovereignty [4] concept and the increasing adoption of the mobile platform and wearable devices, it is urgent to build a new version of EHR systems with user- centric access control and privacy preservation.

To better bring this concept into reality, we adopt two novel technologies, Intel SGX and blockchain, to implement a patient-centric personal health data management system with accountability and decentralization. Intel SGX offers an anonymous key system (AKS) [5] that can generate an anonymous certificate which will then be transmitted to a certification platform for validation. The Intel SGX enabled hardware layer can provision a trusted execution environment in the cloud, and generate data access tokens for reliable data storage and process.

Blockchain technology originated from Bitcoin [6], where data are stored in a public, distributed and immutable ledger, which are maintained by a decentralized network of computing nodes, providing the robustness against failure and attacks, as well as functions for data provenance [7] and access control [8]. The metadata describing each transaction is available to everyone on the system, but that does not mean the data stored within the blockchain is readable [9]. Blockchain relies on pseudoanonymity (replacing names with identifiers) and public key infrastructure (PKI), keeping the privacy of the users. The workshop [10] co-held by the Office of the National Coordinator for Health IT (ONC) and the National Institute for Standards and Technology (NIST) focused on the blockchain usage in healthcare and research, and presented several papers from healthcare providers and academic researchers which clarified the implications of blockchain technology as an infrastructure for healthcare industry in different perspectives and use cases including privacy preservation for predictive modeling, increasing interoperability between institutions at a large scale, immutability of health records, health insurance claim process improvement, health information exchange, healthcare delivery models with artificial intelligence, identity management, monetization strategies and data provenance requirements.

In this paper, we first present a review of blockchain integration concerns existing in current blockchain development and applications, and then dive into the healthcare scenario and adopt the perspective of mobile users, take into consideration the emerging usage of wearable devices, and propose a mobile user controlled, blockchain-based system for personal health data sharing and collaboration, in the sense of IoT scenarios. In the implementation, we build our system on Hyperledger Fabric [11], which is a permissioned blockchain requiring the network nodes to validate, making a step closer to a privacy preserving personal healthcare system with a broader coverage of the healthcare ecosystem from the end device to the cloud, as well as the emphasis of the user ownership for health data. Our main contribution lies in the following objectives that we seek to fulfill.

- **Self-Sovereign Data Ownership.** Adopt the idea of user centric architecture to control data access and issue permissions. It is the information owner that decides who can access the data and whether to make the data public or private, as well as how to validate the data.
- **Scalable Data Processing.** The volume of health data collected from wearable devices and user input scales greatly which requires a high-speed processing capability of the system. The scalability and efficiency of data processing affects the complexity of system integration.
- **Permanent Data Record with Integrity.** Collect data records and submit an abstract of each record to the blockchain network.

The record should be included in a block. The integrity of the record is guaranteed by the consensus mechanism used in the block mining process.

- **Decentralized Privacy and Access Control.** For personal health data, each of the data access request should be processed to get a permission from the data owner with a decentralized permission management protocol. We propose a decentralized permission management protocol to deal with each personal health data request. The access control policies should be stored in a distributed manner which ensures stability. The data access records are stored to provide traceable logs, using blockchain to preserve immutability.
- **Trusted Accountability.** The trusted execution environment provisioned by Intel SGX is utilized to generate a fingerprint for each data access. For medical treatment and insurance enforcement, every action is traceable. Once data leakage is detected, the malicious entity can be identified for investigation.

The rest of the paper is organized as follows. Section 2 presents a review of blockchain integration concerns in current blockchain adoptions. Section 3 introduces the overall system design, including the architecture, system entities, key establishment and system procedures. We describe the system implementation in Section 5 and give a performance evaluation and security analysis in Section 6. Section 7 presents some related work, while Section 8 concludes the paper and talks about future work.

2. Blockchain Integration Concerns

2.1 Risks with Blockchain vs. Centralized Database

The standard method for maintaining identities in industries is to use a network connected and secure centralized database. The centralized database likely would have system backups in case something were to go wrong the system could be rolled back to a stable or correct state. Blockchain solutions for managing identification or data offers a more distributed redundant system but also offers more points of attack. With a centralized database solution, a company only needs to focus on maximum security and protection to one point. A distributed ledger system or blockchain would require a level of protection and security to many different nodes.

With more copies of the data in the wild there are more opportunities or ways that intruders could potentially access the data stored on the ledger in a malicious way. Depending on the type of implementation the system could be made up of non-uniform nodes providing not just new nodes of access but potentially additional attack surfaces on each node. This would require additional work to ensure each different node is secured properly to maintain a consistent level of privacy among all participating nodes.

2.2 Cryptographic Vulnerabilities

The foundational component of the Blockchain technology is the cryptosystem. State-of-the art blockchain's cryptography systems utilize public key algorithms such as, Elliptic Curve Cryptography and message digests, such as, SHA-256. In a typical bitcoin application, an Elliptic Curve key pair which contains a public key and private key is generated based on Secp256K1 curves. The private key has the traditional usage of being kept secret and utilized to sign transactions. For instance, in the bitcoin use case, when a user exchanges bitcoins with another user, the user will sign the transaction with their private key prior to announcing to the network. Once the transaction is signed, the miners in the network will use consensus algorithms to verify the validity of the transaction signature and validation is achieved through consensus. Upon successful validation, the distributed ledger in Blockchain is updated and the transaction is committed. At the same time, the public key is used to generate the bitcoin address and serves as the conduit to receive bitcoins. In addition to the public key, an added security is provided to the bitcoin's address by computing the hash using SHA-256 and RPIEMD-160. In addition, a byte with network identifier is prepending to the hash and checksum is computed with SHA-256 twice.

The generation of bitcoin wallet requires the following additional steps. First, a digital representation of your public key is computed using SHA-256 followed by RIPEMD-160. Second, a byte with network id is prepended to this string. Third, a checksum of this string is computed by performing SHA-256 twice. From these results the first 4 bytes are appended to the string produced in second step. This string is encoded in Base58 which results in the eventual bitcoin wallet address.

Elliptic Curve Digital Signature Algorithm (ECDSA) is used to authorize transactions in the blockchain. ECDSA's primary components are elliptic curve and a cryptographic hash function. The elliptic curve function, secp256k1 and SHA 256 are currently used by popular blockchain platforms, such as, Ethereum. The strength of elliptic curve cryptography is derived on the premise that the discrete logarithm problem on an elliptic curve is computationally hard to solve. However, there are vulnerabilities in implementation of ECDSA. Specifically, the vulnerabilities are associated with the processes for generation for curve, key, signature and verification. The process involved in generating elliptic curves is susceptible to adversarial attack. Specifically, the parameters in the elliptic curve secp256K1 function can be manipulated to generate weak curves. Secp256K1 is susceptible to side channel attacks which leak information. The twist curve vulnerability in Secp256K1

allows an attacker to generate curves similar to the original curve and leads to the attacker extracting the private key. The formulae used for scalar operations used on an elliptic curve might deviate causing errors. The emergence of Quantum Computing poses the greatest threat to the Blockchain cryptosystem. Quantum computers use subatomic particles and multiple states to conduct scalable operations. Quantum computers can exploit ECDSA by the use of qubits to conduct hash collision attacks.

2.3 Infrastructure and Technological Modifications

Infrastructure today is mostly developed to support centralized systems where bottlenecks exist with many nodes fighting for access to a central server. Today's infrastructure works to ensure the pipeline to these central servers are as large as possible. Infrastructure for decentralized systems will not require such large network access for one unit, but many smaller connections to many units. This type of infrastructure is already being developed in support of IoT. Blockchain technology will conveniently grow with the infrastructure that is being developed for the distributed nature of IoT. IoT will require systems that are not only built to work with distributed systems but function as a distributed system. As the infrastructure adapts to a more decentralized IoT infrastructure the system will coincidentally be adapting to a blockchain system.

One challenging area that need to be built on is the ability to monitor these distributed systems. Automated and controlled auditing will need to be developed in order to ensure the validity of the system. How these systems function will be a topic of research and future development. As shown in figure 1, the auditing system will likely monitor the consensus system or the validating nodes to ensure that everything is functioning the way it is intended to.

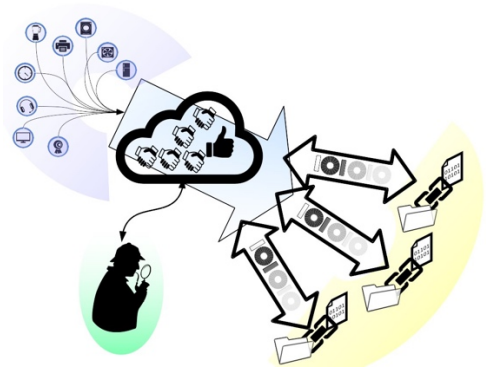

Figure 1. The blockchain system uses auditing processes to periodically monitor the operations ensuring the validity of the system.

2.4 Public vs. Private Blockchain

Original development of blockchain technology was cryptocurrency based implementation of bitcoin. This was a very public implementation of blockchain where the membership was available and free to everyone. The openness offered a level of security and attributed to its ability to be highly distributed. Public block chains as shown in the bitcoin implementation offer some of the best data integrity because of the larger number of copies and redundancy of the ledger. Public blockchains however, bring challenges to security concerns in scenarios where data needs to be better protected or not open in a public sense. In cases such as this private blockchains are becoming more popular, but are implementations that steer the development of blockchain away from a decentralized system to a more centralized system. The original intent of blockchain was to avoid centralized systems.

The public vs private blockchain dilemma is the balancing of the needs per scenario to best understand how much centralization is needed in order to implement a successful blockchain solution. IoT in a global sense will likely need a public blockchain solution but for some use cases private blockchains will be preferred. Public blockchain is a system with decentralization with no middle man or organization managing or controlling IoT that interact in the public space. IoT in a business sense will likely need a private blockchain solution, where in this case there is more centralization. Businesses will likely prefer this to protect their IoT devices and data associated with them. A Private blockchain will still allow some redundancy and decentralization but is not open to the public realm offering more security. Research is progressing in this area in order to identify the best amount of centralization to add to blockchain implementations with the focus on improving security and performance but recognize that as private blockchains move towards more centralization they stray away from the values that make blockchain so valuable [12].

2.5 Smart Contract Adoption Issues

Smart contracts offer a more efficient way for transaction occur. They do this by removing the man in the middle. Using automated services and algorithms to handle and ensure the transaction can happen, users will have to rely on the invisible happenings of the smart contract. These types of systems just like most automated systems involved with critical situations will take time for the general public to gain trust in them. Implementations of smart contracts are already being used and implemented in automatic toll collection and with the existing implementation of bitcoin. As the public uses these systems and future systems more their trust will grow and the benefits of smart contracts will be clear. Security concerns will always exist in automated systems but with research and newly developed tools this concern can be minimized and mitigated. The concerns with IoT and blockchain are described in the prior sections of this paper but systems are being developed to address these vulnerabilities.

2.6 Application Domain (Healthcare)

In recent years, the rise of wearable technology and the IoT has brought great opportunities and challenges to the healthcare domain. Wearable technology refers to networked devices embedded with sensors (e.g., heart rate, blood pressure, pedometer) which can be worn comfortably on the body to collect health data and tracking activities. Hospitals and medical institutions can use these data to link with other Electronic Health Record (EHR) data, such as clinical notes, to facilitate health monitoring, disease diagnoses and treatment. Health insurance companies can make detailed and strategic policies according to individual characteristics, benefiting customers to choose flexible insurance plans according to their needs. To handle sharing and integrating health data between institutions, there is a need for a secure and trusted data sharing infrastructure. However, there are several challenges related to privacy, security, and interoperability facing current health care systems and cloud based infrastructures. First, health data are highly privacy sensitive, especially as more data are storing in a public cloud, raising the risks and concerns of data exposure and leakage.

Current state-of-the-art approaches focus on improving the responsibilities of data providers to detect the data leakage activities, however, it is also crucial to protect data access and provide immediate notifications of data leakage risks. Second, current systems use centralized architecture, which increases the security risk footprint, and requires centralized trust in a single authority. Moreover, the effective integration of health data and the interoperability between healthcare systems remain a challenging task. Over 300 different EHR systems are in use today, but there are little or even no communication and cooperation among systems, resulting in the lack of a holistic and thorough view of personal health. Another challenge is that users have little control over their personal health data. It is reported that 62% of insured adults rely on their doctors to manage their health records, which limits their ability to interact with other healthcare providers than their primary doctor. With the notion of self-sovereignty concept and the increasing adoption of the mobile platform and wearable devices, there is an urgency to develop a new version of EHR systems with user-centric access control and privacy preservation.

As more physical devices such as mobile phones, wearable devices, and medical instruments are connecting to the Internet through embedded systems and sensors, large amounts of data can be collected and sent to the cloud computing system to conduct data analysis for a better and faster decision making. Moreover, these devices can perform commissions and tasks that humans cannot easily accomplish. However, as IoT grows, the connectivity is increasing, and the computing infrastructure will become more complex, opening up more vulnerabilities for the cyber-attacks. Some of the physical devices are located in unsecured environments and easily tampered by hackers. More of the data and the operation commands traveling over through wireless sensor network to the Internet, an untrusted communication channel, are likely to be modified. Therefore, device authorizations and data provenance would be a critical issue. Moreover, many existing IoT systems rely on centralized communication models to connect to servers or cloud computing that support processing and data storage. The problem is that the server will become a bottleneck and a new target for cyber-attack, as well as a point of failure that will disrupt the entire network and impact the data integrity. Meanwhile, many devices require regular and effective maintenance to operate correctly and meet their design specifications. The consequences of ineffective maintenance can be huge, especially in healthcare systems concerning patient care, personnel morale and management time [13]. Such consequences are often overlooked or miscalculated because device breakdowns are not just a cause of lost time but have a direct effect on patient throughput, efficiency and thus waiting lists. Therefore, the importance of effective maintenance to reduce the occurrence of such incidents cannot be overstated. Considering the emerging applications of IoT devices and the increasing complexity of both the software and hardware infrastructure, how to build a truly trusted and integrated environment to support this connected devices and computing infrastructure to transfer data, and to detect rogue IoT devices in a timely manner, remains challenged. The management of single or few devices is relatively simple. However, most hospitals have thousands of medical devices and to correctly maintain these devices can be difficult or impossible without a formalized computerized scheduling system (database). Therefore, it is crucial to adopt a structured approach to planned preventive maintenance projects and implement a computerized system, or audit, refine and improve the effectiveness of the existing implemented system. Moreover, to handle sharing and integrating health data between institutions, there is a need for a secure and trusted data sharing infrastructure between devices and device owners.

Considering the large number of nodes enrollment when faces with blockchain integration into IoT devices, the scalability and the performance such as network latency and throughput plays an important role. Consortium blockchains, also known as permissioned blockchains, which involve BFT protocols, serve an alternative to de- centralized cryptocurrencies and to characterize

the performance cost that decentralized blockchains incur by distributing trust. The technique to be adopted for improving the scalability of the consensus is to shard it by splitting up the task of consensus among concurrently operating sets of nodes, with the aim of improving throughput and reducing per-node processing and storage requirements. The second technique is to adopt the delegation of trust and a hierarchy of sidechains. Sidechains can potentially have a lower degree of decentralization than the top-level blockchain. Sidechains may also run non-proof-of-work consensus protocols, such as BFT. One sidechain structure, proposed in [14], permits transactions to move funds among independent chains. A miner coordination entity is involved to maintain the sidechains and to reduce the responsibility of the main chain while minimizing the side chain influence on the mining power of the main chain. Node bootstrapping is time consuming especially for large scale environment so in the system implementation, different types of bootstrapping strategies should be provided to dynamically adopt appropriate protocols. There is a trade-off between system robustness and fast node bootstrapping. To optimize the bootstrapping efficiency, minimum robustness requirement should be met. Rogue device detection is also challenging for large scale device clusters. The key system feature should be able to detect rogue devices with acceptable success rate but a low latency.

A mobile user controlled, blockchain-based system for personal health data sharing and collaboration is a good use case. The perspective of mobile users is taken into consideration the emerging usage of wearable devices. The system can be implemented on Hyperledger Fabric, a permissioned blockchain requiring the network nodes to validate, and realizes a privacy preserving personal healthcare system with a broader coverage of the healthcare ecosystem from the end device to the cloud, as well as the emphasis of the user ownership for health data.

3. System Design

3.1 System Overview

A three-layer architecture for accountability and privacy preservation is designed for the PHDM system. The data sharing layer provides users with entire control over their personal health data and handles data requests from third parties. The Intel SGX enabled hardware layer provisions a trusted execution environment in the cloud, generates data access tokens and is responsible for reliable data storage and process. The blockchain network layer, which is distributed and untrusted, records data operations and various data access requests for immutability and integrity protection. Figure 2 is a general scenario for the PHDM system. Personal wearable devices collect original health data, such as walking distance, sleeping conditions and heartbeat, which may be synchronized by the user with their online account associated with the cloud server and cloud database. Every piece of health data could be hashed and uploaded to the blockchain network for record keeping and integrity protection. The original data is maintained in the cloud database hosted on trusted platform enabled by Intel SGX. The user owns personal health data, maintains access tokens and is responsible for granting, denying and revoking data access from any other parties. As discussed in [27], Human Data Interaction (HDI) is challenging the existing IT design and practices, and we need to be aware that our data is being collected and be aware of the data themselves and the implications of that data. So in this paper, for personal health data access request from healthcare provider and health insurance company, a permission from the data owner is needed with a decentralized permission management scheme. Besides, each of the access request and access activity should be recorded on the blockchain for further auditing or investigation.

Key Establishment

In the patient centric data management system, users are required to register an online account to be involved in the system, and generate data encryption key pairs to encrypt their cloud data for confidentiality. For key management, we assume the system developers adopt a secure wallet service. The description of each key established is as follows.

- **User Registration Key KUR**. The user needs to create an online account to store health data collected from wearable devices and other sources in the cloud database. We denote the user registration key as KUR. Every time user wants to operate on their cloud health data, the registration key is needed. This key is generated from the platform identity key using Intel SGX anonymous key system and is thus bounded to the user. Even if the user's registration key is stolen or compromised, it could not be used elsewhere without the user authentication. Similarly, the registration key for healthcare provider and healthcare insurance company is KHR and KIR, respectively.
- **Data Encryption Key KDE**. After registration, the user generates an encryption key KDE to encrypt all the health data stored in the cloud database. When a health data entry is created, user has the option to encrypt the data entry, which limits the data access only to the key owners, and the hashed data entry will be uploaded instantly to the blockchain.
- **Data Sharing Public/Private Key Pair ($PKDS$, $PRDS$)**. For health data sharing, a public/private key pair will be generated, denoted as ($PKDS$,

PRDS). In some cases that the data sharing activity is to be recorded on the blockchain, the private key is used to generate a signature from the user to indicate the health data ownership, while the public key is used by others to verify the ownership. When users want to share their health data with healthcare providers or insurance companies, they share the private key for data access and the corresponding tokens generated with this private key.

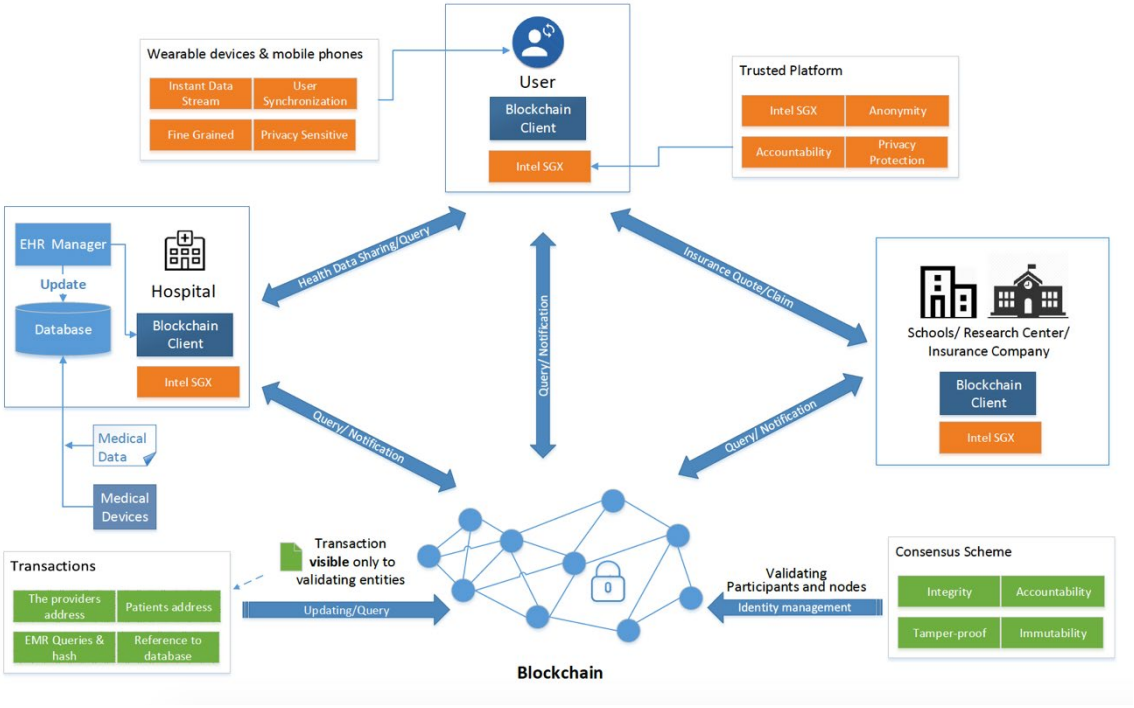

Figure 2. Patient Centric Personal Health Data Management System Scenario.

- **Platform Identification Key** *KPID* . Each trusted platform owns a platform identification key *KPID* , also generated from the platform identity key using Intel SGX anonymous key system. Every health data request and data access on a certain platform will generate an activity record signed by *KPID* for accountability while still with anonymity preserved. Different entity keys are noted as *KPIDu* for users, *KPIDp* for healthcare providers and *KPIDi* for insurance companies.

System Entities

User. System users collect data from wearable devices which monitor users' health data such as walking distance, sleeping conditions, and heartbeat. These data are then uploaded to the cloud database hosted on trusted platform via the mobile application. User is the owner of personal health data and is responsible for granting, denying and revoking data access from any other parties, such as healthcare providers and insurance companies. If the user is seeking medical treatment, the user would share the health data with the desired doctors. If the treatment is finished, the data access is revoked to deny further access from the doctors. Same scenario applies to user-insurance company relations. Besides, user can also record everyday activities according to a particular medical treatment such as medicine usage to share with the treatment provider for adjustment and better improvement.

Wearable Device. Wearable Devices serve to transform original health information into human readable format and then the data is synchronized by the user to their online account. Each account is associated with a set of wearable devices and possible medical devices. When a piece of health data is generated, it will be uploaded to the blockchain network for record keeping and integrity protection.

Healthcare Provider. Healthcare providers such as doctors are appointed by a certain user to perform medical test, give some suggestions or provide medical treatment. Meanwhile, the medical treatment data can be uploaded to the blockchain

network for data sharing with other healthcare providers under the user's permission. And the current healthcare provider can request access to previous health data and medical treatment from the user. Every data request and the corresponding data access is recorded on the blockchain.

Health Insurance Company. User may request a health insurance quote from health insurance companies or agents to choose a proper health insurance plan. However, users cannot hide or modify medical treatment history data since those data is permanently recorded on the blockchain network and the integrity and trustworthiness is ensured. Moreover, the insurance claims can also be recorded on the blockchain.

Cloud Database. The cloud database stores user health related data, data requests from the healthcare provider and insurance companies, data access record and data access control policy. Data access is accountable and traceable. Once data leakage is detected, the malicious entity can be identified.

Blockchain Network. The blockchain network is used for three purposes. For health data collected from both wearable devices and healthcare providers, each of the hashed data entry is uploaded to the blockchain network for integrity protection. For personal health data access from healthcare provider and health insurance company, each of the data access request should be processed to get a permission from the data owner with a decentralized permission management protocol. The access control policies should be stored in a distributed manner on the blockchain which ensures stability. Besides, each of the access request and access activity should be recorded on the blockchain for further auditing or investigation.

System Procedures

In the system, there are four phases for personal health data management including user registration, health data generation and synchronization (data generated from user, healthcare provider and insurance company), health data access management, health data access record uploading and health data access auditing.

3.2 Token-based Access Control

For anonymity and verification purposes, we adopt the token based access control mechanism to handle the data management process. As is shown in Figure 3, the cloud server is responsible for issuing and verifying tokens, and also maintaining both the data record database and data access log database. Users can request and share the access tokens to data requestors. Potential data requestors include healthcare providers, insurance companies and even system auditors. Each data and token operation is recorded in the blockchain and thus validated. After user registration, the cloud server can issue tokens based on the personal information provided by users. To access data, the required token will be presented to the cloud server and verified. The server issuance operation, the user token presentation and verification omit system logs which will be stored in the log database, as well as data requests and access from third parties.

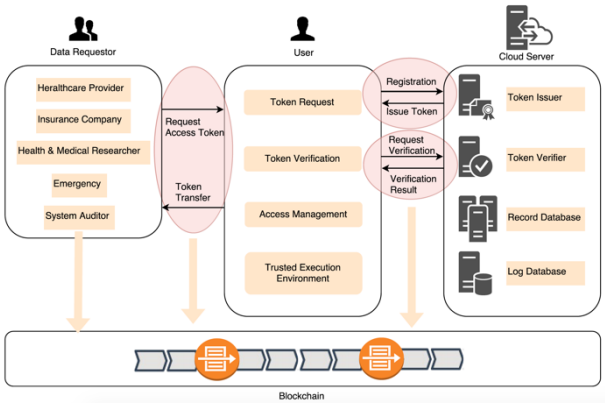

Figure 3. PHDM System Interaction.

U-Prove Based Token Generation

User registration is based on U-Prove [18], which is proved capable to be integrated into Trusted Platform Module 2.0 in [19]. U-Prove [20] includes three entities, namely issuer, prover and verifier. In our system, the issuer and the verifier is the same entity, that is, the cloud server. The user in our PHDM system is the prover entity in U-Prove model. During user registration phase, there are some parameter definitions for both prover and issuer.

- The value of the token information field (TI): $TI \in (0, 1)^*$
- The value of the token information field (PI): $PI \in (0, 1)^*$
- Application Attributes (AA): $(A1, ..., An)$, TI $(A1, ..., An)$ indicates n attributes from the application itself.
- Issuer Parameters (IP): $UIDp$, $desc(Gp)$, $UIDH$, $(g0, g1, ..., gn, gt)$, $(e1, ..., en)$, S
  $UIDp$ is an application-specific identifier for this particular IP, which is unique across the PHDM system and $desc(Gp)$ specifies the group (Gp) with an order of p which is used for discrete logarithm computation in the following verification steps. $UIDH$ is the identifier for the secure hash algorithm. $(g0, g1, ..., gn, gt)$ is the Issuer's public key. $(e1, ..., en)$ is generated from AA, indicating the format of each application attribute.
- The hash of the $IP(P)$: $P = H(IP)$

- Device-protected Boolean (DB): d
  This indicates whether the protocol is device pro- tected. PHDM adopts trusted execution environment so the value by default is *true*.
- Device Parameters (DP): gd, xd, hd
  The Device generator gd satisfies $gd \in Gq$. xd is device private key and hd is the public key.

With the above information provided, we choose the issuance protocol version number 0x01. The user platform identification key $KPIDu$ is used to generate the device private key. The token generation protocol during user registration is as follows.

The cloud server issues tokens to users with the signature $(\sigma_z{}^1, \sigma_c{}^1, \sigma_r{}^1)$. For privacy concerns, the application attributes are hashed for the generation of U-Prove based token. During some circumstances, the issuer is able to generate multiple tokens at one time for better performance.

Protocol 1 User Registration on the Cloud Server

Input:

$x_t = Hash(0x01, P, TI)$,
$x_i = Hash(A_i)$,
$\gamma = g_0 g^{x_1} \cdots g_n x_n h_d$
UID_p, random $\alpha, \beta_1, \beta_2, \omega$,
and issuer private key y_0

Compute:

$h = \gamma^{\alpha}$, $\sigma_z = \gamma^{y_0}$, $\sigma_z{}^1 = \gamma^{y_0}$,
$\sigma_a{}^1 = g_0{}^{\beta_1} g^{\beta_2} g^{w}$
$\sigma_b{}^1 = (\sigma_z{}^1)^{\beta_1} h^{\beta_2} \gamma w \alpha$
$\sigma_c{}^1 = Hash(h, PI, \sigma_z{}^1, \sigma_a{}^1, \sigma_b{}^1)$
$\sigma_r{}^1 = (\sigma_c{}^1 + \beta_1 \bmod q) y_0 + \omega \bmod q + \beta_2 \bmod q$

Output:

U-Prove token $T : UID_P, h, TI, PI, \sigma_z{}^1, \sigma_c{}^1, \sigma_r{}^1, d$
prover private key: α^{-1}

Token Presentation Protocol

A presentation proof of ownership of certain messages or attributes contained in the token is generated using the token private key and is required to access user data in the cloud database. Before accessing data, the data requestor needs to attest itself and convince the user that it is running on top of SGX enabled environment in an isolated enclave. The SGX attestation is launched by the data requestor which will send a signed quote to the data owner for verification using the platform dependent key. The remote attestation between the two platforms is performed with the assistance of the Intel Attestation Service [21]. After the verification, the user will request a one-time U-Prove token with a newly generated private key PR_{DS} and share it with the data requestor. The data requestor forwards the token to the verifier of the cloud database and will be granted access after the verification. Different decisions can be made by the user, such as to grant, deny and revoke access. The presentation proof serves two purposes. For one thing, it proves the integrity and the authenticity of the attribute values and for another, it establishes the confirmation of the ownership of the private key associated with the token itself, which will further prevent token replay attack.

4 System implementation

4.1 Personal Health Data Collection and Synchronization

Personal health data comes from wearable devices such as activity trackers or smart watches, and medical devices such as pacemakers or defibrillation, as well as manual user input for treatment tracking such as medicine usage and training. To synchronize personal data, the user first can register to the cloud service provider for an online account with enough storage capability.

4.2 Personal Health Data Integrity Protection and Validation

Figure 4 shows the basic data flow from the user device to the cloud server, finally anchored on the ledger with proof of integrity and validation. The health data comes from a variety of devices all day, resulting in a large number of data records. To facilitate scalable and efficient data processing and integrity protection, we develop a tree-based method for the integrity management of health data record. Some data records are batched to form a tree-based data struc- ture and handle dynamic data enrollment. The adoption of Merkle tree [15] realizes the scalability requirement, and most importantly improves the efficiency to validate the data integrity. Merkle tree is a binary tree structure where the input is a list of hashed data records. These records are ordered by the time when they are generated. Every two records are grouped together and the hashes of the two data records become two leaf nodes of the Merkle tree and consequently constitute a high level group node with the group hash generated by concatenating two hashes. Two group nodes will

follow the same way to generate a new higher level group node with a new hash. This step is repeated until there is a single hash which will become the tree root, that is, the Merkle root.

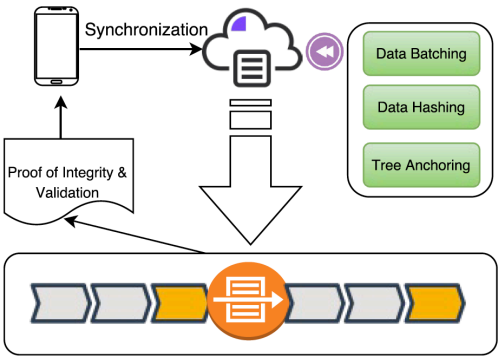

Figure 4. User Centric Personal Health Data Integrity Protection.

Chainpoint [17] is an open standard for creating a timestamped proof of any data, file, or series of events, which proposes a scalable protocol for publishing data records on the blockchain and generating a Merkle proof for each data record. In our implementation, we anchor a list of data records to multiple Fabric channels by binding the Merkle root to a blockchain transaction and verify the integrity and existence of data without relying on a trusted third-party. The hash of data records brings two advantages. For one thing, each Merkle tree can host a large number of records since only the hash of the data record is stored. For another, the hash is an effective measure to detect changes so that once a piece of data is modified, the action can be detected easily by traversing the tree.

The Merkle root, along with the tree path from the current node to the root node, serves as the proof of integrity and validation, that is, the Merkle proof. First, we need to identify the record location, the targetHashB. The target hash and the path to the Merkle root, that is, nodes in green, constitute the Merkle proof of the hashed data record, which is stored in a JSON-LD document that contains the information to cryptographically verify that the record is anchored to a blockchain. By calculating the hashes in different tree levels, it is easy and fast to obtain the root hash, which is anchored in the blockchain transaction, witnessed and maintained by some distributed nodes. It proves the data was created as it was at the time anchored. The Merkle root for each Merkle tree is related to one transaction in the blockchain network, which means a blockchain transaction represents a list of data records the Merkle hosts, enabling the scalability and effectiveness of data integrity protection and validation.

4.3 Personal Data Sharing and Healthcare Collaboration

The user can share data with healthcare providers to seek healthcare services, and with insurance companies to get a quote for the insurance policy and to be insured. When data sharing is detected in the system, there will be an event generated to record the data access request. The event record can be described using a tuple as {datahash, owner, receiver, time, location, expirydate, signature}. This record is then submitted to the blockchain network which is followed by several steps to transform a list of records into a transaction. A list of transactions will be used to form a block, and the block will be validated by nodes in the blockchain network. After a series of processes, the integrity of the record can be preserved, and future validation on the block and the transaction related to this record is available. Each time there is an operation on the personal health data, a record will be reflected to the blockchain. This ensures that every action on personal health data is accountable.

We implement an access control scheme by utilizing the Hyperledger Fabric membership service component and the channel scheme, as is shown in Figure 5. The CA is responsible for issuing transaction certificates for participating entities in the Hyperledger Fabric blockchain network and participating Fabric client, and generating the access control list during channel establishment according to user settings and operations. Different access type can be specified in the certificate, such as query and update operations for chaincode execution in the channel. Chaincode is a piece of code that is deployed to Hyperledger Fabric for enabling interactions between peers and the shared ledger. There are three operations on the chaincode, including deploy, invoke and query. A chaincode can be installed on a blockchain by executing a deploy transaction while a chaincode execution is launched by invoke transactions. Channel is formed to isolate individual activities among authorized parties. In Figure 5, we have two channels established for two users, respectively. A user may perform data collection and synchronization on the mobile platform, which will send web requests to the cloud server. Healthcare providers and insurance companies also communicate with the server to request or update health data and health insurance information. With the permission from users, these requests will be allowed to participate in a certain channel. The cloud server is configured with a Fabric client to communicate with the Fabric blockchain network peer. For different user activities, the data will be labeled with different channel ID to distinguish isolated domain. The query or update requests from the server will be forwarded to the Fabric network via Fabric client for confirmations. Distributed peers will validate the incoming requests and propose transactions by executing chaincode. The ordering service is responsible for checking transaction signatures and order them with channel IDs. For each channel, there is a subledger, part of the system ledger, to record all transactions in blocks.

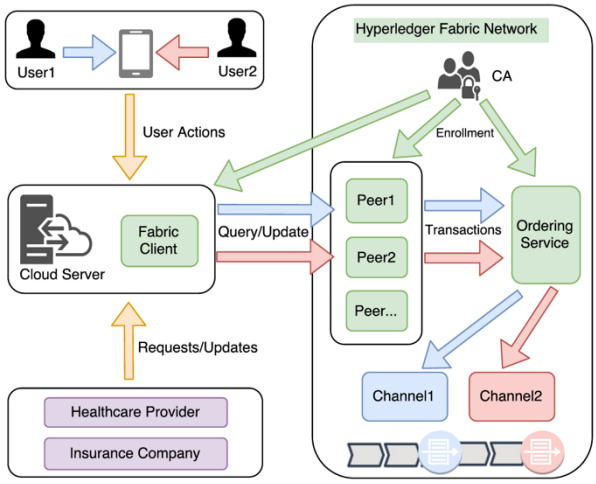

Figure 5. Personal Health Data Sharing and Collaboration Using Hyperledger Fabric and Channel for Mobile Users.

For privacy concerns, the user can selectively share health data with data requester, based on the necessity of how personal health data is required to assist the healthcare service. For example, a user's insurance history may not be important when the user is talking to a dentist. Similarly, the user's dental treatment is not necessary for skin testing or other treatment. To issue a specific certificate, the user can state clearly in the certificate what category of personal data is allowed access, whether read-only or read-write access is allowed. Moreover, in different channels, different grained information is shared. In this sense, our system provides a fine-grained privacy protection and access control policy.

5 System evaluation

Our system adopts a user-centric model for processing personal health data using blockchain network, ensuring the data ownership of individuals, as well as data integrity. The operations on the data records are highly inter-operable and compatible with current systems. By enforcing access control policies, users can handle their personal data without worrying about the privacy issues. Meanwhile, each request and update from healthcare providers and health insurance companies are recorded and anchored to the blockchain network, making actions towards personal health data accountable.

With all the security objectives proposed in Section 1 achieved, it is crucial to evaluate the system performance, regarding to the scalability and efficiency of the data integrity proof generation and data validation process. We test different numbers of concurrent records with a range from 1 to 10,000. Figure 6 and 7 shows the average time cost, respectively.

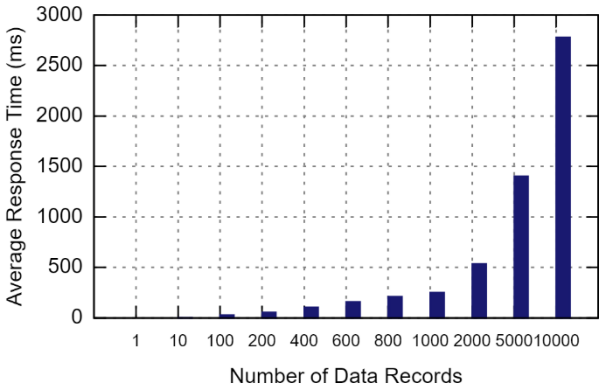

Figure 6. Average Time Cost for Data Integrity Proof Generation.

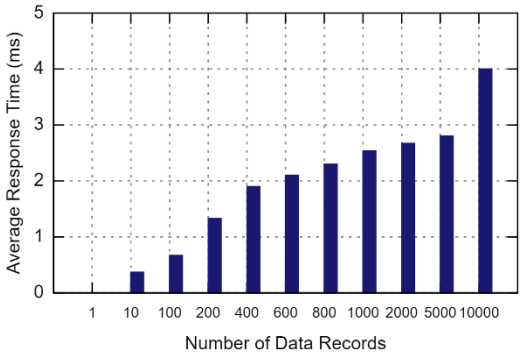

Figure 7. Average Time Cost for Data Integrity Proof Validation.

From these two figures, we can conclude that the system can handle a large dataset at low latency, which indicates the scalability and efficiency of the data process. By adopting Merkle tree method to batch data, we implement an algorithm with the computation complexity of $O(log2n)$. This is an important advantage when the data records are collected at a high frequency. In the future, we will take a deeper vision into the delay tolerance for healthcare data processing and improve the data collaboration procedures accordingly.

For U-Prove based token generation, we select five attributes predefined and involved in each token and two of them are required to obtain a data access token. During the token issuance, there are basically two cryptographic methods for digital signature including Subgroup and ECC. The evaluation results for token issuance and presentation with these two methods are shown in Figure 8 and Figure 9. It can be concluded that ECC-based token generation is more efficient than the subgroup-based method. This can be explained that ECC utilizes shorter key length for the elliptic curve than subgroups of equivalent security levels and computes faster with a small

field. Adopting the ECC-based U-Prove protocols for both token issuance and presentation, the average overhead brought to the system is 8.1% and 9.4%, respectively.

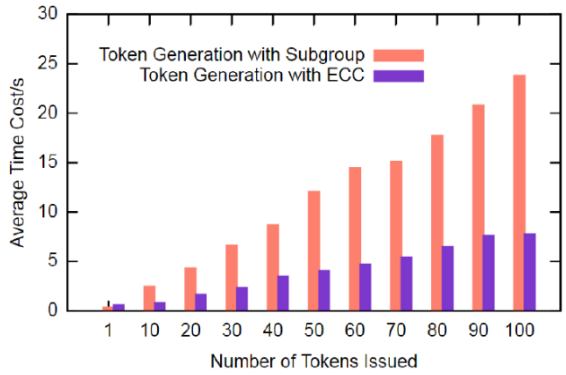

Figure 8. Average Time Cost for Token Issuance.

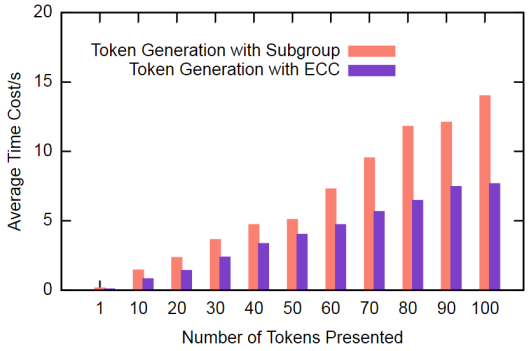

Figure 9. Average Time Cost for Token Presentation.

6 Related work

A simple mobile application is implemented in [22] for healthcare data sharing but is limited to patient and doctor communication. Some work has been done to integrate blockchain technology to the healthcare industry. [23] proposes a proof of interoperability to avoid the computation cost of proof of work but didn't mention the access control for personal health data. [24] addresses the adoption of blockchain in social network domain but not fully explores the benefits of the blockchain. Patientory is designed for healthcare storage network using Ethereum and adopts a token based access control model, but data privacy is highly dependent on the cryptography methods. MedRec [25] is a record management system focusing on EMRs using smart contract, resists against the single point of failure, but raises privacy concerns. [26] points out that MPC (Secure Multi-Party Computing) is a promising solution to enable untrusted third-party to conduct computation over patient data without violating privacy but the actual efficiency is not clear. [24] addresses the adoption of blockchain in social network domain but not fully explores the benefits of the blockchain. The security and privacy concerns in healthcare industry is fully discussed in [28], making research work on data protection in healthcare an urgent need.

7 Conclusion and future work

In this paper, we build a web based system for personal health data collection, sharing and collaboration between individuals and healthcare providers, as well as insurance companies, using blockchain and Intel SGX. The system can also be extended to accommodate the usage of health data for research purposes. By utilizing blockchain technology in the self-sovereign healthcare systems, we manage to distribute the responsibility of maintaining trusted records for data operation as well as token generations. Meanwhile, benefiting from the blockchain consensus scheme and the decentralized architecture, along with the trusted execution environment and the platform dependency provisioned by Intel SGX, the records are anchored with trusted times- tamping and redundancy, preserving both availability and accountability of the healthcare data and operations. We also propose a U-Prove based protocols for the permission management. We implement a prototype of the PHDM system and the evaluation shows that the performance is acceptable. The algorithm to handle data records can preserve both integrity and privacy at the same time. Meanwhile, we adopt the concept of channel supported by Hyperledger Fabric to deal with the isolated communication required by specific scenarios. In the future, we will integrate the PHDM system with the enhancement of a blockchain based access control scheme to provide better data protection and user privacy. Also we will explore how to combine both personal health data and medical data together and provide a better solution to cover a broader scenario in healthcare industries.

Acknowledgements

This work was supported by Office of the Assistant Secretary of Defense for Research and Engineering (OASD (R& E)) agreement FA8750-15-2-0120 and Air Force Material Command award FA8750-16-0301.

References

[1] A. D. Thierer, "The internet of things and wearable technology: Addressing privacy and security concerns without derailing innovation," *Richmond Journal of Law & Technology*, vol. 21, p. 1, 2014.

[2] "2016 connected patient report," https://www.salesforce.com/assets/pdf/industries/2016-state- of-the-connected-patient-pr.pdf.
[3] L. J. Kish and E. J. Topol, "Unpatients why patients should own their medical data," *Nature biotechnology*, vol. 33, no. 9, pp. 921– 924, 2015.
[4] J. H. Clippinger, "Why Self-Sovereignty Matters," https://idcubed.org/chapter-2-self-sovereignty-matters/, [Online; accessed 7-March-2017].
[5] N. Sarangdhar, D. Nemiroff, N. Smith, E. Brickell, and J. Li, "Trusted platform module certification and attestation utilizing an anonymous key system," May 19 2016, uS Patent App. 14/542,491. [Online]. Available: https://www.google.com/patents/US20160142212
[6] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
[7] X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat, and L. Njilla, "Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability," in *International Symposium on Cluster, Cloud and Grid Computing*. IEEE/ACM,2017.
[8] T. Hardjono and A. S. Pentland, "Verifiable anonymous identities and access control in permissioned blockchains."
[9] M. Mainelli, M. Smith *et al.*, "Sharing ledgers for sharing economies: an exploration of mutual distributed ledgers (aka blockchain technology)," *The Journal of Financial Perspectives*, vol. 3, no. 3, pp. 38–69, 2015.
[10] T. O. of the National Coordinator for Health IT (ONC), the National Institute for Standards, and T. (NIST), "Use of blockchain in healthcare and research workshop," 2016.
[11] C. Cachin, "Architecture of the hyperledger blockchain fabric," in *Workshop on Distributed Cryptocurrencies and Consensus Ledgers*, 2016.
[12] V. Gramoli, "On the danger of private blockchains," in *Workshop on Distributed Cryptocurrencies and Consensus Ledgers (DCCL16)*, 2016.
[13] C. Luca and R. Ciorap, "Software system for medical device man- agement and maintenance," in *International Conference on Advance- ments of Medicine and Health Care through Technology*. Springer, 2011, pp. 84–89.
[14] A. Back and G. Maxwell, "Transferring ledger assets between blockchains via pegged sidechains," May 9 2016, uS Patent App. 15/150,032.
[15] R. C. Merkle, "Protocols for public key cryptosystems," in *Security and Privacy, 1980 IEEE Symposium on*, April 1980, pp. 122–122.
[16] "Tierion api," https://tierion.com/app/api.
[17] "Chainpoint: A scalable protocol for anchoring data in the blockchain and generating blockchain receipts," http://www.chainpoint.org/.
[18] C. Paquin and G. Zaverucha, "U-prove cryptographic specification v1. 1," *Technical Report, Microsoft Corporation*, 2011.
[19] L. Chen and J. Li, "Flexible and scalable digital signatures in tpm 2.0," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, ser. CCS '13. New York, NY, USA: ACM, 2013, pp. 37–48. [Online]. Available: http://doi.acm.org/10.1145/2508859.2516729
[20] C. Paquin, "U-prove technology overview v1.1 (revision 2)," April 2013. [Online]. Available: https://www.microsoft.com/en- us/research/publication/u-prove-technology-overview-v1-1- revision-2/
[21] I. Anati, S. Gueron, S. Johnson, and V. Scarlata, "Innovative technology for cpu based attestation and sealing," in *Proceedings of the 2nd international workshop on hardware and architectural support for security and privacy*, vol. 13, 2013.
[22] H. Kim, H. Song, S. Lee, H. Kim, and I. Song, "A simple approach to share users' own healthcare data with a mobile phone," in *Ubiquitous and Future Networks (ICUFN), 2016 Eighth International Conference on*. IEEE, 2016, pp. 453–455.
[23] K. Peterson, R. Deeduvanu, P. Kanjamala, and K. Boles, "A blockchain-based approach to health information exchange networks," 2016.
[24] J. Zhang, N. Xue, and X. Huang, "A secure system for pervasive social network-based healthcare," *IEEE Access*, vol. 4, pp. 9239–9250, 2016.
[25] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: Using blockchain for medical data access and permission management," in *Open and Big Data (OBD), International Conference on*. IEEE, 2016, pp. 25–30.
[26] X. Yue, H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare data gateways: Found healthcare intelligence on blockchain with novel privacy risk control," *Journal of Medical Systems*, vol. 40, no. 10, p. 218, Aug 2016. [Online]. Available: https://doi.org/10.1007/s10916-016-0574-6
[27] T Werman, "Human Data Interaction (HDI): The New Information Frontier" Mar 2019. [Online]. Available: https://www.interaction-design.org/literature/article/human-data-interaction-hdi-the-new-information-frontier
[28] A. Chacko1 and T. Hayajneh1, "Security and Privacy Issues with IoT in Healthcare," EAI Endorsed Transactions on Pervasive Health and Technology, vol. 4, 2018.