Freie Universität Berlin

Secure Data Exchange Using Distributed Ledger Technologies

*A **thesis exposé** presented for the degree of **Bachelor of Science**.*

Lucas Antelo, Freie Universität Berlin, Germany

Matriculation number: 5094454

antelo.lucas@fu-berlin.de

20th March 2019

Supervisor:

Nicolas Lehmann[1], Freie Universität Berlin, Germany

Reviewers:

Prof. Dr. Agnès Voisard[2], Freie Universität Berlin, Germany

Prof. Dr. Matthias Wählisch[3], Freie Universität Berlin, Germany

Version: 0.8

[1]Dept. of Computer Science and Mathematics, Databases and Information Systems Group
[2]Dept. of Computer Science and Mathematics, Databases and Information Systems Group
[3]Dept. of Computer Science and Mathematics, Internet Technologies Group

1 Introduction

The part *Motivation* discusses distributed ledger technologies (DLTs) and its current implementations in the healthcare setting. Furthermore, the part dives into the problems of data silos in several healthcare systems worldwide and how DLTs solve the connections among them improving the availability and completeness of electronic healthcare records and consequently the quality of care. The part *Outline of Contribution* provides an overview over the contributions of this thesis and their implications for the current German healthcare system.

1.1 Motivation

At the Cebit 2018 in Hanover, a DLT named blockchain caught a lot of attention and it seems that it may solve all our problems. And currently, the healthcare sector seems to have many of those problems due to shortages of healthcare professionals, outdated IT infrastructures and the burden of an aging society. And the number of problems are estimated to increase in the next two decades. Thus, the question arises if blockchain technology is able to at least solve some problems in this sector?

Unfortunately, actual DLT implementations by several companies remain quite obscure in the healthcare sector and inquiries over specific implementations inconclusive. Nevertheless, the hype around blockchain and healthcare does not wane and the fear of a bubble is not negligent. As in 2019, Estonia remains the only country in the world to have implemented DLT on a national scale. Estonia implemented this technology for the Estonian Electronic Healthcare Records in collaboration with Nortal, Helmes, Guardtime and INTELSYS [16, 17] in 2016, which allows patients to access their electronic health records and check for any inconsistencies. This makes tampering with the data more difficult to accomplish.

In the future, electronic health records will be the norm in the US due to political and societal pressures [12] and it is expected to be similar in Germany in the next two decades. The next step is to coordinate these data silos held by each healthcare facility such that patients and further healthcare providers have access to a complete history of health records improving the availability and completeness of their medical history and consequently the quality of care. These electronic health records form the basis for reimbursements, clinical research, teaching and legal disputes to name a few. And soon this data will form the training data in order to train artificial intelligence applications [29] improving the efficiency and quality of care as well as provide enhanced tools for future healthcare professionals.

In general, DLTs efficiently, reliably and persistently store records of data and data transactions in a network among different parties. DLT guarantees data integrity by making these records immutable. On the other hand, by making these records publicly available confidentiality concerns arise over the storage of data and transactions.

Fortunately, there exist several solutions how to ensure privacy on a public ledger [2, 7, 9, 28] using DLTs. Nevertheless, the General Data Protection Regulation and health laws impose further restrictions like the right of correction and deletion of personal information, which is not trivial when using an immutable public ledger as data structure.

This thesis analyzes the environment for data exchanges for highly sensitive data in different public ledger settings and focuses on the question if DLT allows secure data exchanges between healthcare providers and patients in Germany.

Based on these findings, appropriate solutions for imposing confidentiality and fulfilling the restrictions given by the General Data Protection Regulation (GDPR) and further relevant laws are examined systematically. Finally, a prototype fulfilling the given constraints is presented and analyzed.

In the current German healthcare system, health records are stored electronically or in a paper format. These records are usually incomplete and not easily transferable between healthcare providers making the access of such information difficult to accomplish. Such circumstances interfere with the quality of care and moreover their effectiveness. Beyond that, patients are less reliant on healthcare professionals compared to the last century such that the business model is switching from a doctor-centered to a patient-centered business model, where the patient decides over the next steps. Consequently, DLT is the next trivial step where the information is distributed according to the data owner: the patient.

Example 1: *Mr. Nakamoto is 21 years old and has recently suffered a pneumonia. His general practitioner Dr. Ripple treated him with the appropriate antibiotic and anti-inflammatory drugs. During the treatment, Mr. Nakamoto decides to visit his girlfriend in Zug, Switzerland. During his stay in Zug, he looses consciousness and develops an acute respiratory failure. The ambulance brings him to the emergency at the Zuger Kantonsspital where healthcare professionals start to treat him based on the current clinical findings. Unfortunately, his health condition deteriorates and Mr. Nakamoto is put under artificial life support as his respiratory system is unable accomplish its function. His girlfriend was unaware of his current health situation nor does she know who his current general practitioner is. This situation could have been prevented if Dr. Ripple had made specific information available and thus avoiding the administration of a medication on which Mr. Nakomoto was allergic to.*

This example shows a realistic scenario that happens on a daily basis worldwide in many healthcare facilities. Unfortunately, health records remain highly inaccessible in a highly mobile population causing fatal results. This asks for an appropriate infrastructure for the healthcare setting that makes critical information available to care takers improving the quality of care and the clinical outcome of their patients. DLTs enables the storage of such information making it easily accessible where needed. On the other hand, security and privacy concerns arise when using such a technology, that need to be addressed.

This thesis addresses the main problem in the current healthcare setting by showing how to make these data silos accessible with DLT. In theory, this technology facilitates the completion of individual health records, that allows healthcare professionals access to critical information in life-threatening situations.

1.2 Outline of Contribution

This thesis provides a solution of a secure data exchange platform for highly sensitive data between healthcare providers. The solution is based on an appropriate DLT and allows healthcare professionals and patients alike, to access and complete individual health records in order to improve the quality of care and consequently the clinical outcome. This thesis scrutinizes the hype around the blockchain technology [33] in the healthcare setting and presents such an implementation and its respective blueprint.

2 Background

The second section *Background* introduces the definitions, methodology, related systems and works for the scientific analysis of this thesis. Healthcare systems worldwide already use well implemented protocols in order to exchange highly sensitive data between healthcare providers, insurers and patients. Nevertheless, these protocols remain highly unsatisfactory and do not conform to different requirements that all the parties demand for: openness, reliability, data integrity and interoperability. Therefore, this thesis investigates the distributed ledger technologies in the context of healthcare related data exchanges fulfilling the following requirements such as

a) interoperability between systems,

b) data integrity,

c) security,

d) reliability of the database,

e) openness and thus access to critical information,

f) accountability,

g) auditability,

h) privacy and confidentiality, and finally

i) control over the data by the owner.

For this purpose, this thesis is based upon the healthcare system common in Germany for the scientific analysis and conception of a distributed ledger technology.

2.1 Definitions

Distributed Ledger Technology (DLT) is based on a ledger that represents a shared data structure among different parties. This distributed ledger is either public or private and functions as a shared view of the current reality in form of a replicated, shared, and synchronized data structure among participating parties. Neither a central administrator nor a centralized data storage is needed using a DLT in a permissionless setting, where the ledger is accessible to anyone, who wants to participate. By changing this into a private or permissioned setting, the adaptation of the protocol enables a central administrator the power to modify the ledger and grant or revoke certain rights to users [21].

Blockchain belongs to the family of DLTs and uses a distributed ledger of transactions stored by each participating node in a peer to peer network (P2P). The ledger represents a linked list, where each block includes the reference to its previous block and thus equals a chain of blocks. The consensus method is usually competitve

based on proof-of-work but it may differ, depending on the implementation of its underlying protocol. The main representatives of blockchain are Bitcoin and Ethereum [21].

Directed acyclic graph (DAG) uses a finite directed graph with no directed cycles as a ledger. This graph consists of vertices and edges, which represent transactions (vertices) and proof-of-work (edges). Here, the consensus method is usually cooperative and its main features consist in low energy consumption next to its absence of fees. The main representatives are IOTA, Byteball and Nano [21].

Smart Contracts are computer programs stored in the ledger and each participating node storing a replica of this ledger can run these programs. These contracts allow operations to be performed on a ledger and are capable to interact with other contracts. Smart contracts may be written in any language, but Solidity and LLL are specifically designed for this purpose [3, 15, 22] and therefore recommended for the creation of smart contracts.

Decentralized Application (DApp) consists of a user interface called the frontend and a backend. The backend is the collection of smart contracts as well as the data needed for the DApp. To sum up, DApps are applications that run on top of the Ethereum network and allow users to access and use the smart contracts [3].

Proof-of-Work (PoW) scheme is based on the solution of a hash puzzle in a DLT. All nodes attempt to find the solution and the winning node gets the reward for the newly appended block to the ledger. The hash puzzle is solved when the hash function results in a value below a certain threshold. The input of the one-way hash function consists of the data in the next block to append as well as the reference value to its previous block. In order to solve the hash puzzle the right nonce (number only used once) needs to be found, that together with the given input results in a hashed value below the threshold and thus the solution of the hash puzzle. By being the first to publish this solution to all other nodes, the new block will be appended to the current ledger. Depending on the protocol, the solution is checked by each participating node before appending any new block [4].

Proof-of-X (PoX) is an umbrella term for schemes that replace PoW with a suitable alternative. These optimizations includes energy-efficiency, usefulness, compliance with the regulation, etc [4].

Proof of Stake (PoS) scheme provides a different approach to establish consensus. Nodes vote on new blocks weighted by their hash power, amount of currency or another suitable metric. Furthermore, PoS includes the random election of a leader among the stakeholders, that is allowed to announce a new block appended to the ledger. Unfortunately, the last approach is susceptible to Denial-of-Service attacks, which is usually prevented by using a confidential communication method among the stakeholders to elect a leader [4].

Proof-of-capacity is similar to PoS, such that it uses the capacity to allocate a non-trivial amount of disk-space as the discriminating metric [4].

Proof-of-Elapsed-Time uses a trusted execution environment (TEE) and is similar to PoS. Nodes request a wait time from their TEE and the chip with the shortest wait time is elected as the leader [4].

Decentralized Autonomous Organization (DAO) is an organization that acts autonomously and makes decisions electronically by a software application or through the vote of its members. In essence it is a system of hard coded rules that define which actions an organization will take. The financial transaction record and program rules are maintained on a distributed ledger [3, 40].

Certificate Transparency (CT) is an approach to increase transparency in the process of issuing certificates such that certificate authorities can be held responsible for their actions. Furthermore, bad certificates can be caught and revoked early on. This approach introduces three entities: a *log server* keeping track of issued certificates, an *auditor* responsible for keeping track of the presence or absence of certificates in the log server, and finally a *monitor* responsible for checking the quality of the certificates in the log server [8].

General Data Protection Regulation (GDPR) specifies the rules on data protection and privacy in the European Union (EU). This regulation applies to all individuals within the EU and the European Economic Area (EEA). It specifies the export of personal data outside the EU and EEA and regulates the storage, access and processing of this data [10].

Trusted Execution Environment (TEE) represents a secure enclave on a hardware with the guarantees, that code and data loaded inside a processor are protected with respect to confidentiality and integrity. These guarantees still hold, when such an enclave is run on a hostile operating system. Nevertheless, availability can not be guaranteed, as the hardware may shut down unexpectedly [7].

Oblivious Random Access Memory Machine (ORAM) is a cryptographic tool that allows a client to store data on a trustless server without compromising safety requirements. The client performs reads and writes remotely by mapping logical memory addresses to remote physical addresses. Freshness, integrity and confidentiality are ensured by using authenticated encryption and minimal local state. ORAM hides memory access patterns such that a malicious server can not distinguish two client operations having the same length. In ORAM, only the client needs to verify the integrity [7].

Publicly-Verifiable Oblivious RAM Machine (PVORM) is an extension of Oblivious Random Access Memory (ORAM). ORAM can be extended by defining a set of application-specific operations, legal and business requirements that are imposed on all updates. This extension of ORAM is defined as Publicly-Verifiable ORAM. PVROM allows to hide account balances and data exchange details from non-transacting entities. These data exchanges are placed on the ledgers as cipher texts in order to impose strong confidentiality. Consequently, parties can only deduce the identities of the sending and receiving entity, but not of the explicit client behind each entity nor its content. Not even the sending entity can deduce the receiving client behind

the receiving entity and vice versa. In addition, the extension allows to publicly verify, that such transactions have been correctly processed by both entities by an external institution [7].

Transaction-graph confidentiality is based on hiding the graph edges (e.g. network nodes) in order to protect the network against deanonymization attacks, which would otherwise allow an exploitation of the current structure such as the discovery of the identities of the transacting parties. Thus, this term specifies the confidentiality requirements in conjunction with a graph used in distributed ledger technologies [7].

Universal Composability (UC) is a general-purpose model for the analysis of cryptographic protocols. By emulating a protocol with very strong security properties by another protocol without external observers being able to discern these two, the other protocol provides the same security properties as the former one. There also exists a simplified form of this model, which has been postulated by Douglas Wikström in [41]. The authors in [28] were the first to propose a formal UC-based framework for describing and proving the security of distributed protocols that interact with a blockchain. These authors refer their proposal as ”the blockchain model of cryptography”.

Zero-knowledge proof allows a user *u* to another user *v* to proof that user *u* knows a given information without publishing the information by simply providing a secret information that allows user *v* to verify the proof. This procedure is based on simple computational assumptions and no further knowledge nor assumptions are needed [34].

Non-interactive zero-knowledge proof (NIZK) allows a user *u* to another user *v* to proof that they know a given information without publishing the information nor establishing any interactions between them. This is done by simply using a common reference string that allows the user *v* to create a secret information and a verification key and thus to verify the proof from user *u*. This reference string allows both users to share information without interacting with each other. In addition, this reference string provides the key to build the necessary blocks for the proof [34].

Zero-knowledge succinct non-interactive argument of knowledge (zk-SNARK) allows a user to construct a proof for a non-polynomial problem that is publicly verifiable without having to disclose its identity. This proof allows a non-interactive procedure to show that a certain information belongs to the user, who constructed the proof, without even disclosing the information itself. This is done by providing one-way mapping functions and the information as output of one of these functions. Next to the proof there exists a verification key which verifies the proof based on the claimed information. The proof is very short and easy to proof why it is called succinct. Moreover, verifying parties do not need to have any prior knowledge given the proof and the verification key next to the information (e.g. Bitcoins, possessions, etc.) [36].

Decentralized anonymous payment (DAP) schemes allow direct anonymous payments of any amount and are used in ZeroCash. In order to provide a well functioning DAP scheme several requirements need to be fulfilled such as availability of

collision-resistant hashing and pseudorandom functions, statistically-hiding commitments, one-time strongly-unforgeable digital signatures and asymmetric key encryption. The conjunction of these requirements are the cryptographic building blocks of a DAP scheme and parameterized by a security parameter π [36].

2.2 Methodology

For the purpose of this thesis, information needs to be collected, assembled and analyzed for the design and conception of a DLT prototype. Therefore, current policies on data exchange practices for highly confidential data in the healthcare sector between healthcare providers, insurance companies and patients are analyzed. Participating parties are put into the center of the investigation and the priorities as well as requirements for the DLT prototype are dynamically adapted along the design process. The collection of information relies on questionnaires, one-on-one as well as group interviews after the analysis of the current regulations and well established common procedures in Germany. In this regard, data security officers in the primary care setting are consulted to ensure that the information retrieved and assembled is up-to-date and for the interpretation of the current laws and regulations in regard to the prototype.

Along the analysis of the current practices, an overview of the different types of data is generated in order to categorize the data accordingly. These categories provide a basis for a separate analysis in conjunction to applicable laws and regulations, such that security concerns can be applied on a subset of categories, relaxing restrictions on different data categories.

Laws and Regulations are analyzed and consequently implemented into the design and concept process. For this purpose, a data security officer in a primary care facility is interviewed having the relevant knowledge what laws and regulations apply for data exchanges between the respective parties.

Process and Data The current established process in Germany for the exchange of health related between clients, healthcare providers and insurance companies are analyzed. Furthermore, the nature of exchanged data are examined and categorized accordingly in order to allow different applicable restrictions on different categories of data. As in the previous step, a data security officer in a primary care facility is interviewed in order to supplement the findings.

Questionnaires and Interviews are performed in order to collect further information. The questions will be based on the analysis of the current state in the German healthcare system and the findings along the research. The target groups consist of clients requiring health related services, healthcare providers from the primary and secondary care, and finally the private and public health insurance companies. The insights from this approach form the basis for the design and concept of a potential prototype.

Models are chosen that apply DLT as a solution in the current healthcare system. Therefore, the Estonian e-health system, MedRec as well as companies using DLTs to share data among parties are analyzed. The findings flow into the design and concept process. For this purpose, all relevant information is collected and the respective authorities contacted for further information.

Requirement Analysis for each potential user is performed and forms the basis for the design and concept of the prototype. For this purpose, an iterative approach is chosen by integrating the given information and insights into the analysis. By putting the given information and insights into the context of the applicable laws and regulations, the requirement is updated according to the progress of the research till the conception of the prototype.

Design and Concept are based on the findings from the previous steps. The final prototype is based on a DLT allowing the data exchange between clients, healthcare providers and insurance companies. Along the design process, a minimal viable product is created and iteratively adapted to fulfill the requirements generated in the previous step.

Evaluation is performed using a testing network with different nodes representing the respective users and benchmarked against security parameters such as interoperability, data integrity, security, realiability of the database, openness, privacy and confidentiality, control over the data by the owner and finally accountability and auditability. Finally, the prototype is analyzed in regard to the laws and regulations that apply in conjunction with a data security officer in a primary care setting.

User Tests are performed by recruiting a number of volunteers playing a given role using a web application with a sleek user interface. The results are collected using a questionnaire, that pop ups, when the test run is finished, providing a feedback on its usability and other aspects not covered in the previous steps. The user test is tailored specifically on the maturity of the prototype.

Adaptation incorporates the final step and consists on minor adaptions based on the final findings. Finally, the prototype is published under the appropriate licenses.

2.3 Related Systems

There exist several applications using DLT to exchange data between different parties. The listing gives an overview of currently known applications and companies attacking the problem of securely exchanging health related data using DLT.

Nortal *"In 2016, the Estonian E-Health Foundation launched a development project aimed at safeguarding patient health records using blockchain technology in archiving related activity logs. "We are using blockchain as an additional layer of security to help us ensure the integrity of health records. Privacy and integrity of healthcare*

information are a top priority for the government and we are happy to work with innovative technologies like the blockchain to make sure our records are kept safe," said Artur Novek, the foundation's Implementation Manager and Architect." [17]

Guardtime "*The Estonian company, a team of over 150 cryptographers, developers and security architects with decades of experience defending networks from state attacks, is a pioneer in K.S.I. blockchain technology. They operate in the Netherlands, Estonia, US, UK and Singapore and tackle data breach management, secure critical infrastructure, enable the safe operation of e-governments, work with financial markets, telecommunication stakeholders, insurance companies or defense and aerospace enterprises. Regarding healthcare, In March 2017, Estonia's eHealth Authority signed a deal with Guardtime, a blockchain pioneer, to secure the health records of over a million Estonians. These highly sensitive records also contain genetic information of patients in thousands of cases. In January 2018, the company signed a deal with a private sector healthcare provider in the United Arab Emirates to bring the technology to the Arabic country.*" [23]

Iryo "*The Central-European, more closely Slovenia-based enterprise says it is creating a global and participatory healthcare ecosystem. The ambitious start-up was established only two years ago, but it is building an open-sourced OpenEHR platform with tremendous efforts and zero-knowledge data repository. The latter ensures that sensitive medical data will remain secure and utterly impervious to cybersecurity breaches, including state-sponsored attacks. The overall goal of Iryo is to build an appropriate platform for keeping health records unified. Instead of all kinds of medical data from various providers stored in different formats and scattered across different systems, Iryo's solution promises to store data securely and allow patients to share their medical history anywhere in the world. If they are testing their technology in Slovenia first, they will have a reasonably manageable pilot cohort.*" [23]

HSBlox "*brings patient-centric solutions to the healthcare ecosystem, combining machine learning and blockchain (distributed ledger technology) to address the healthcare industry's demand for secure, real-time information sharing and interventions. To support value-based care programs, HSBlox deploys smart contracts to automate multi-party transactions, such as bundled payments and patient referrals. The proven technology enhances the provider, payer and patient experience throughout the care continuum, driving better outcomes for each healthcare stakeholder.*" [25]

MedicalChain "uses blockchain technology to securely store health records and maintain a single version of the truth. The different organizations such as doctors, hospitals, laboratories, pharmacists and health insurers can request permission to access a patient's record to serve their purpose and record transactions on the distributed ledger. Medicalchain provides solutions to today's health record problems. The platform is built to securely store and share electronic health records. By digitizing health records and empowering users we can leverage countless industry synergies." [32]

Corda "The healthcare industry represents a large and complex network of institutions that must manage patient data in a highly regulated and fragmented environment. The challenges include:

- Sharing data to facilitate a more comprehensive view of a patient while retaining a patient's privacy
- Synchronizing of numerous disparate systems that hold patient data with disparate identifiers and data formats
- Intersecting with healthcare insurers and providers who require assurance on data validity and provenance
- Managing inconsistent and complex rules and processes

Corda unites disparate processes, increases data flow, and reduces costs resulting in improved patient experience and outcomes. The opportunities for applying Corda to solve the systemic issues of healthcare include:

- A global network of interoperable entities that can exchange data in a secure, confidential manner while ensuring compliance to healthcare regulation
- Framework for identity that preserves confidentiality while putting patients in control of their information across providers
- Smart contracts for a consistent, rule-based method for accessing patient data
- Connectivity to insurance providers to verifiable, immutable records on which to base claims in an automated process" [11]

2.4 Related Work

Related Work provides an overview on related applications, academic research papers and scientific books that deal with and provide a basis for the understanding of DLTs in the context of secure data exchanges. This section is structured into three parts *Applications*, *Security Aspects* and *Further Aspects*.

2.4.1 Applications

In [2], **Asaph Azaria, Ariel Ekblaw, Thiago Vieira, and Andrew Lippman** present a decentralized record management system using blockchain technology that handles electronic medical records. MedRec manages authentication, confidentiality, accountability and data sharing using a modular design, such that existing electronic health records can be integrated, which facilitates the interoperability between legacy systems. MedRec runs on top of Ethereum and each node can be represented by a healthcare provider or a patient, who controls the data in MedRec. MedRec addresses the problems with fragmented and slow access to medical data, system interoperability, patient agency and finally data quality and quantity for medical research.

The architecture of MedRec consists of three entities that control how data is organized giving the control to the patient. The Registrar Contract (RC) maps participant identification strings to their Ethereum address. The Patient-Provider Relationship Contract (PPR) lists which nodes store and manage health records for the respective

nodes. And finally the Summary Contract (SC) functions as a bread crum trail for participants to localize their respective data.

MedRec provides a suitable starting point for a prototype as its source code is available on github and furthemore MedRec is currently being tested at the Beth Israel Deacon Medical Center in Boston, Massachusetts [19].

Their findings are interleaved along the analysis and design process of a suitable prototype for the German healthcare system without having to solve every problem along the design process.

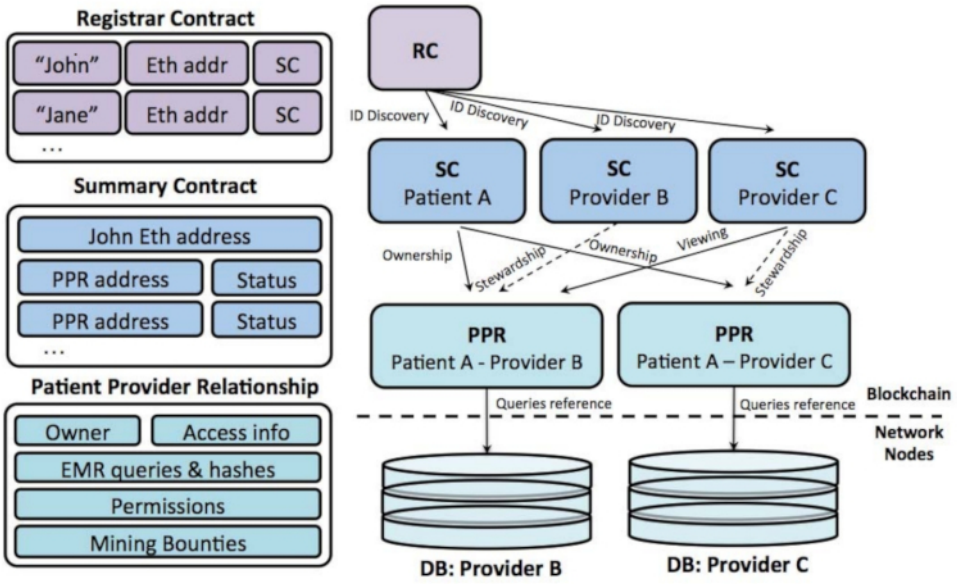

Figure 2.1: The architecture of MedRec runs on a smart contract enabling DLT such as Ethereum that acts as the database. Three smart contracts control how data is organized. The Registrar Contract (RC) lists the participant identification to their Ethereum address. The Patient-Provider Relationship (PPR) Contract lists which nodes store and manage health records for the respective patient. The Summary contract (SC) maps all data managed by a provider to the respective patient and its respective status such as public or private [2].

In [28], **Ahmed Kosba, Andrew Miller, Elaine Shi, Zikai Wen, and Charalampos Papamanthou** show how a decentralized smart contract system can be built using their smart contract compiler named Hawk such that financial transactions are stored as encrypted entries in the ledger. The compiled smart contract does not store financial transactions in the clear and consequently ensures transactional privacy in a public ledger unless the contractual parties do disclose such information themselves. This transactional privacy is based on sending encrypted information to the blockchain and relies on zero knowledge proofs to ensure the correctness of contract execution and money conservation. Current distributed ledger technologies lack

such a transactional privacy feature. Unfortunately, this solution does not allow for auditability, which is mandatory for insurance companies, healthcare providers and governmental institutions. Nevertheless, Hawk allows programmers to write smart contracts without having to deal with the implementations of cryptographic tools. The Hawk compiler generates an executable including the cryptographic protocol between the blockchain and the users. This allows contractual parties to interact with the blockchain without disclosing its identity nor transactional information in the public ledger.

They assume that the generalized consensus protocol is secure and that the blockchain can be trusted for correctness and availability but not for privacy. They use a modified version of ZeroCash to achieve stronger security.

By interchanging financial data in a transaction with sensitive data representing a String, we could theoretically achieve the same conclusion as stated in the paper. Therefore their findings can be applied in the healthcare setting when transacting patient related information.

Thus, this paper gives a solution for the implementation of a DLT in the healthcare environment. These findings play a fundamental role along the design and implementation of a DLT by using Hawk as the framework for the conception of the prototype. Moreover, this paper is one of the fundamental papers on privacy protection in blockchain using smart contracts providing substantial background information for a potential prototype.

In [36], **Eli Ben-Sasson, Alessandro Chiesa, Christina Garman, Matthew Green, Ian Miers, Eran Tromer, and Madars Virza** discuss a solution named ZeroCash, that allows two parties to transfer a value among them without disclosing their identities to each other nor the amount transferred to external observers. ZeroCash uses a set of cryptographic building blocks in order to instantiate a DAP scheme, which forms the infrastructure for ZeroCash. Value among parties are transferred using sk-SNARKs. ZeroCash relies on Bitcoin as the base ledger but their proposal can be applied to any other DLT.

Hawk provides a good solution how to establish confidential transaction between parties and uses a modified version of Zerocash in order to enhance its inherent security and privacy characteristics.

This paper will be mentioned along the design and implementation of an ideal DLT solution for the healthcare setting in order to complement the findings in [28], as Hawk relies on features of Zerocash.

In [9], **Raymond Cheng, Fan Zhang, Jernej Kos, Warren He, Nicholas Hynes, Noah Johnson, Ari Juels, Andrew Miller, and Dawn Song** present Ekiden, a system, that combines DLTs enabling smart contracts with TEE. This system allows smart contracts to run on any system without compromising confidentiality and security. TEE provides the necessary security features such that the application runs as an enclave on any system by protecting its data from any other application and operating system. This system works furthermore on any desired DLT. Ekiden helps

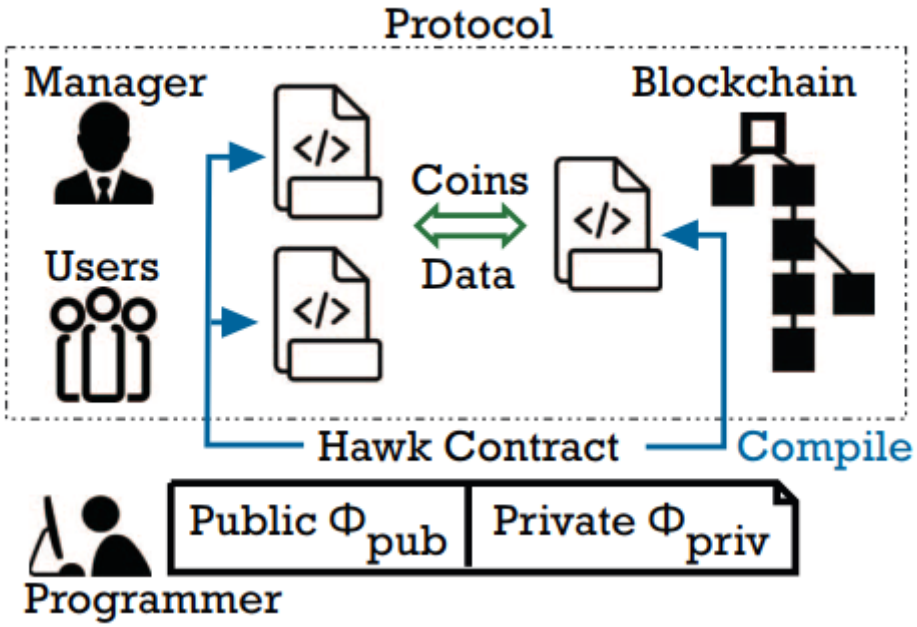

Figure 2.2: The Hawk compiler generates a cryptographic protocol between the blockchain and the users. This protocol represents the program, which consists of a private Φ_{pub} and public Φ_{priv} portion. The public portion Φ_{pub} acts as the application interface. The private portion Φ_{priv} takes the input data of all the parties and performs the computation. This private portion is meant to protect the parties' data and the exchange of value. The manager is a special facilitating party, that has access to the inputs and is trusted not to disclose any private data. This party supervises the correct execution of the smart contract such as in auctions [28].

to keep sensitive data confidential and smart contracts are endowed with strong confidentiality and integrity guarantees by using a hybrid architecture combining trusted hardware and the DLT. There are three types of entities when using this system: *clients*, *compute nodes* and *consensus nodes*.

Healthcare related data is mostly regarded as sensitive data which usually needs to be kept confidential, if we want to enable the execution of smart contracts on such data. Even if the system/node is compromised, Ekiden protects the data from a malicious system, that runs the smart contract. They used several models to evaluate Ekiden as well as a machine learning model detecting heart diseases in patients. They showed, that the ”machine learning contracts allow clients with sensitive data to train a shared model in a secure setting” [9] without exposing any plain text training data. Nevertheless, nodes may shutdown and therefore the availability of such smart contracts on any node cannot be guaranteed with TEE and consequently not in Ekiden.

This paper gives next to Hawk, a solution for the implementation of a prototype. Thus, their findings play a crucial role along the design and prototyping of an applicable DLT for the healthcare setting. Moreover, this paper is one of the fundamental papers on privacy protection in DLTs providing important background information for the thesis, especially as Ekiden represents an improvement compared to Hawk. Finally, Ekiden provides a practical example on healthcare related data, as they showed how to use sensitive data to train a machine learning model for diagnosis heart diseases and consequently showing how smart contracts can be used for such applications.

In [7], **Ethan Cecchetti, Fan Zhang, Yan Ji, Ahmed Kosba, Ari Juels, and Elaine Shi** present Solidus, a protocol, that keeps transactions confidential on public DLTs using the concept of Publicly-Verifiable Oblivious RAM Machine (PVROM). In the setting of the financial banking sector with several banks representing a modest number of clients, a client can request a secure transaction, that remains confidential even if stored on a public ledger. The bank will perform the transaction for the client without disclosing the identity of neither the sending nor receiving client represented by the other bank. The transaction is stored on a permissioned or private ledger among participating banks. The bank internally maps the identity of its client with the identification used on the public ledger, without disclosing the true identity of its client to external observers nor other banks. Solidus therefore allows auditability that is mandatory for financial institutions without impacting the confidentiality towards their clients.

In a public ledger, a secure data exchange between participants need to remain confidential in order to comply with the current data protection laws for highly sensitive data. By interchanging banks with healthcare providers and insurances, we can establish a similar scenario without having to heavily modify the protocol. If such a given infrastructure exists, insurance companies can be transformed into decentralized autonomous organization allowing for a more cost efficient healthcare system in the near future.

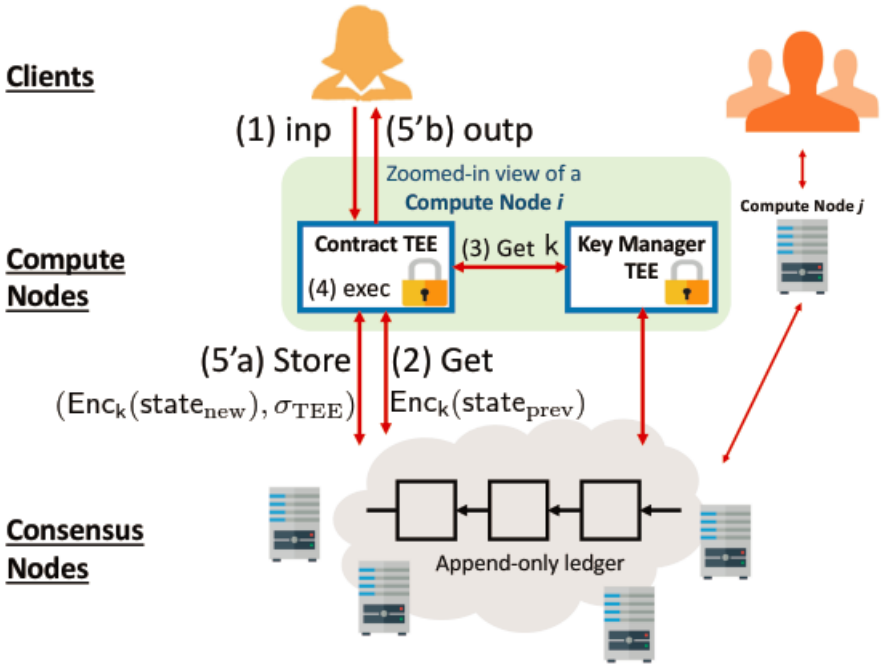

Figure 2.3: The architecture of Ekiden consists of clients, compute and consensus nodes. Clients send inputs (1) to smart contracts, which are executed within a TEE (4) preserving the confidentiality of the input on a compute node. The contract within the TEE retrieves the respective data from the ledger (2) in order to process the clients' input. The key manager running within a separate TEE stores and provides the key associated with the respective smart contract (3). The compute node executes the smart contract with the clients' input. The smart contract stores the output encrypted with its respective keys on the ledger (5'a) and returns the output for the clients encrypted with their respective public keys (5'b). The consensus node verifies that the attestation σ_{TEE} is correct before appending the associated output on the ledger, that stores the encrypted contract state [9].

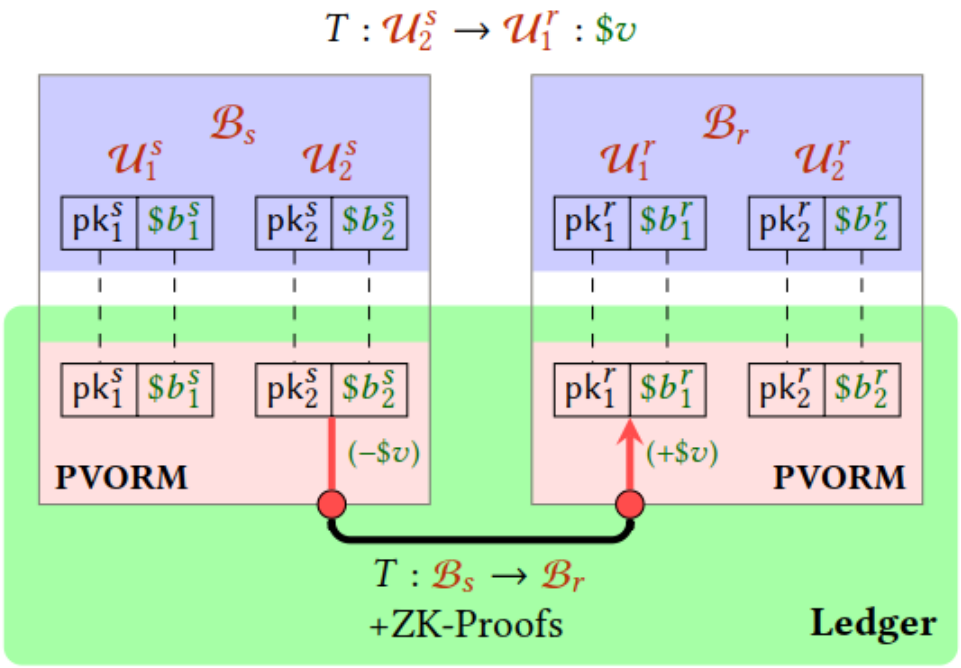

Figure 2.4: An example transaction T in Solidus where B_a represents a bank among a set of $a \in \mathbb{N}$ different banks and U_i^a represents a user of a bank B_a which oversees $i \in \mathbb{N}$ different users. In this example a user U_2^s at a bank B_s sends the amount of \$v $\in \mathbb{R}_{\geq 0}$ Dollars to user U_1^r at the bank B_r. Both banks represent two users, that are stored in a logical (plaintext) memory of each bank's PVORM represented as blue boxes and the lower pink boxes are the associated public (encrypted) memories. External observers only see that a user at a bank B_s sent money to a user at a bank B_r but nothing about the sending nor receiving user. After a successful transaction, both banks update their respective PVORMs correctly [7].

This paper forms the basis for the conception of a potential prototype next to Hawk and Ekiden. Furthermore, their findings addresses fundamental aspects on privacy protection in a public DLT. Solidus represents an improvement compared to Hawk as it allows for auditability, making Solidus a good candidate as main framework for the prototype in the healthcare setting. Finally, Solidus shows that secure data exchanges between parties are possible when using a public DLT.

2.4.2 Security Aspects

In [30], **Li Xiaoqi, Jiang Peng, Chen Ting, Luo Xiapu and Wen Quaoyan** systematically examine the security risks to popular DLT systems. The divide the common risks into nine categories such as

- 51% vulnerability,
- private key security,
- criminal activity,
- double spending,
- transation privacy leakage,
- criminal smart contracts,
- vulnerabilities in smart contract,
- under-optimized smart contracts and
- under-priced operations.

For each of these security risks, they explain the mechanism and provide a good starting point for further research.

In addition, they provide, explain and analyze attack cases in order to understand further security flaws in current DLT systems such as

- selfish mining attacks,
- DAO attacks,
- Border Gateway Protocol (BGP) hijacking attacks,
- eclipse attacks and
- balance attacks.

Finally, their work list and explain security enhancements to blockchain systems. These enhancements include solutions such as SmartPool, Quantitative Framework, OYENTE, Hawk and Town Crier, which allow developers to build a more robust DLT implementation.

Their work give a systematic overview of current security risks inherent in popular DLT solutions such as Ethereum. Their findings are interleaved in the conception of a potential prototype and act as a framework in the analysis of the prototype.

In [44], **Guy Zyskind, Oz Nathan, and Alex 'Sandy' Pentland** demonstrate the use of blockchain protocols for permission management. A third party service stores encrypted data on a distributed hashtable (DHT) that points to the data held by the user on its local device. Their solution, a decentralized personal data management system, allows users to own and control their data. They apply a protocol, that turns a blockchain into an automated access-control manager. Their system ensures data ownership, transparency and auditability as well as fine-grained access control.

This paper explains the underlying principles for MedRec and their findings are included in the design and conception of the prototype. MedRec is built on these ideas and features distributed record retrieval, smart contract permissioning schemes, data sharing, and the economics of information supply and demand via blockchain mining [2].

In [43], **Aviv Zohar** shows the weaknesses inherent to first generation DLTs such as Bitcoin and discusses different protocols and how to improve them. The protocols such as GHOST (Greedy Heaviest Observer Sub-Tree), SPECTRE (Serialization of Proof-of-Work Events: Confirming Transactions via Recursive Elections [37]) and off-chain transaction channels improve the scalability by either modifying the data structure, adapting dynamically the time interval for the creation of a new block based on the delay of the network propagation of new blocks, and finally by utilizing off-chain transaction channels, respectively.

From a security perspective, this paper gives a short overview over the inherent weaknesses but does not provide any insightful solution on how to improve them.

This paper acts a supplemental source to explain inherent problems with first generation DLTs and how they can be addressed.

In [26], **Ari Juels, Ahmed Kosba, and Elaine Shi** show how to facilitate leakage of confidential information, theft of cryptographic keys as well as a physical-world crimes (e.g. murder, poison, etc.) using smart contracts. Through their examples, they illuminate the extend of problems with current DLTs enabling smart contracts.

The healthcare economy relies on confidential information, which needs to be protected against malicious attacks and abuses. Their findings show, that safeguards are needed in the design of a DLT solution. Aside of these threats, their example on how to leak confidential information between two parties provides a good solution for clients and healthcare providers to share information, which could be stigmatizing and provide an economic disadvantage for the client (e.g. HIV, cancer, etc.).

In [20], **Ittay Eyal and Emin Gün Sirer** demonstrate that the distributed protocol of Bitcoin is not incentive-compatible such that an attack on the public ledger may successfully be centralized by a minority pool of colluding miners. These colluding miners need to run a modified bitcoin protocol that would favor their own mined blocks and ledger. By applying a stochastic strategy these miners can put themselves into the most favorable position to get the rewards for mined blocks. They will keep mined blocks secret as long as they are ahead of the public ledger and will ignore any block mined outside their pool. By publishing their private branch they are able to invalidate the public branch as new blocks are only appended on the longest chain and hence claiming the rewards for all privately mined blocks in the ledger. In case their branch is equal in length to the public branch, the colluding miners simply publish their branch hoping that the majority will apply their previously private instead of the public branch when appending new blocks and thus reaping the rewards for the selfish pool. This strategy allows a minority of miners to reap a disproportionate amount of rewards and moreover giving incentives to honest miners to join their pool further centralizing the mining power and consequently destroying the decentralized nature of Bitcoin. When these colluding miners reach the majority in the network no new members will be accepted making Bitcoin a centralized cryptocurrency.

This paper shows a weakness inherent to first generation distributed ledger technologies like Bitcoin, what can happen in any environment applying a distributed ledger technology. Therefore, a solution for the healthcare setting needs to be set in place avoiding any injustices between insurers, healthcare providers and clients.

To sum up, their findings help to define the weaknesses of the first generation distributed ledger technologies and explain why there are further perspectives to consider, when designing DLT implementation for the healthcare sector.

In [24], **Stuart Haber and W. Scott Stornetta** provide a practical solution on how to time-stamp digital documents , that are so easy to tamper with, in a distributed network of users with a unique identification number. Their solution makes it difficult to back-date or forward-date such documents by still maintaining complete privacy of the documents without requiring any record-keeping by the time-stamping services. A hash function takes the digital document as input and outputs a hash value. This hash value is then signed by $k \in \mathbb{N}$ external clients and together added to a ledger as a shared data structure.

In the healthcare setting time-stamping document is a useful tool and allows medical professionals, insurance companies as well as clients to put this information into context.

This paper explains the essence of the current distributed ledger technologies and provides a practical solution on how to time-stamp any document as well as a batch of documents. Their findings show how to fulfill important constraints in the healthcare setting like auditability, data integrity and authenticity.

In [4], **Shehar Bano,Alberto Sonnino,Mustafa Al-Bassam, Sarah Azouvi, Patrick McCorry, Sarah Meiklejohn, and George Danezis** give a systematic overview of

consensus protocols used in DLTs and show alternatives to the Proof-of-Work (PoW) schemes. These alternatives are summed up under the the term Proof-of-X (PoX) and are presented next to hybrid consensus schemes. These hybrid consensus schemes rely on single or multiple committees with different advantages and disadvantages.

The healthcare setting is already working on a given infrastructure with given laws and regulations, why PoW might not be the best choice availabe when implementing a DLT solution, especially when trust in the infrastructure is enforced by law. Therefore, other consensus schemes may be preferred, that can be implemented on top of the given infrastructure.

Their work will be included along the design and implementation of a DLT prototype. Aside, they provide a useful overview over the different consensus schemes available for DLTs.

2.4.3 Further Aspects

In [6], **Ken Binmore** provides insights from game theory, which are implemented in different DLT implementations. This book gives a reasonable introduction and overview for this thesis. The healthcare system relies on many political and economic incentives which may be enforced by applying game theoretical frameworks to favor certain outcomes and to facilitate desired actions.

Game theory helps to design a system that favors certain outcomes when a goal oriented approach is desired. Therefore, game theory gives insightful and useful solutions for the healthcare setting, where many different parties compete with each other for different resources and rewards.

Insights from game theory are included along the analysis and design process of a suitable prototype.

In [5], **Moritz Becker** investigates the driving factors of health information privacy concerns and the underlying rational of users' privacy perception with respect to commercially available health wearables without the need of healthcare professionals. By using an iterative thematic analysis approach health wearables users were assigned to seven semi-structured focus groups with six users of health wearables each, where the participants discussed health wearable and privacy related topics given by a thematic map. This investigative method is based on the Health Information Privacy Concerns Model composed of six dimensions to explicitly address privacy concerns with health information technologies[27]. This list provides an overview over these six dimensions with respect to the results in this paper.

- *Collection* describes the subjective concern with respect to the accumulation and storage of personal health related information.
- *Unauthorized Secondary Use* addresses the concerns over the use of users' data other than agreed upon.

- *Improper Access* describes users' concerns over the perceived threat of unauthorized access by third parties.
- *Errors* describes the data inaccuracies and its consequences for the respective health wearable users.
- *Control* covers the user's perceived control over the personal health information.
- *Awareness* addresses the concerns regarding the lack of awareness how personal health information is used and moreover protected.

These findings provide an orientation and focus on the most relevant aspects for the patients in regard to their data. This paper influences the outcome of the requirement analysis forming the basis for the design and concept of an appropriate prototype.

In [21], **Pietro Ferraro, Christopher King, and Robert Shorten** explain how a distributed ledger technology enhanced by embedded systems can be designed and applied such that compliance levels of users can be controlled. They discuss the differences between blockchain and directed acyclic graph (DAG) technologies and conclude, that DAG are more suitable for controlling the compliance of users that share access to a common resources. DAG allow for high-frequency micro-trading as well as fee-less transactions compared to the blockchain. Furthermore, DAG provides the necessary privacy protection for its users.

The example is based on a street junction and how their solution can favor good behavior from cyclists and drivers. Their results can be applied to other scenarios, where a system is needed to control the human behavior towards shared resources.

They provide an approach on how to design socially conforming system based on DLT. In the future, sensors will become cheaper and their presence ubiquitous This will allow families to perform the necessary care at home with a remote care management system as backbone. Furthermore, it is estimated, that AI applications may soon play a stronger role in the monitoring of vital parameters and alleviating work processes for the healthcare professionals making their findings interesting to consider.

In addition, the healthcare system suffers from a shortage of many different kinds of resources such as medical devices, medical professionals, organs, equipment and so forth. Consequently, the implementation of a system that allows to orchestrate its users towards a fairer use of resources can help to improve some of the current problems in this sector. As an example, such as system could reward users taking care of their family members. With this rewards, family members can request health services in return.

The insights from this research will be included in the design process of a suitable DLT prototype such that processes can be controlled to improve the quality of care as well as clinical outcomes.

In [8], **Melissa Chase and Sarah Meiklejohn** explain how to implement a distributed ledger technology without needing to store the whole ledger as a user and thus to eliminate the hash-based mining. Their proposal relies on the assumption that one

is willing to adopt a distributed rather than a fully decentralized solution and thus to trust any set of named properties. Under this assumption a certificate transparency (CT) solution can be applied on top of a DLT, where certificates are issued and appended as *dynamic list commitments*, that represents a merkle tree with committed events stored as certificates. This transparency overlay consists of *log servers* (stores events as certificates in a publicly available merkle tree), *auditors* (checks that specific events are in the log) and *monitors* (flags any problematic entries in the log). This allows users to request transfers without having to store the complete ledger. Users directly communicate with the log servers and act concurrently as an auditor and a monitor. Consequently, users can autonomously check for correctness and finally ensure accountability in a distributed system.

Their proposal allows the implementation of an energy efficient DLT without the need of hash-based mining by using the underlying infrastructure that runs the healthcare setting in each country. Hence, this approach reduces the costs for maintaining a DLT and allows clients, insurance companies as well as healthcare providers to check for correctness without having to rely on a central authority for bad certificates.

These findings are interleaved along the design and concept of a DLT prototype.

3 Planned Work

The third section *Planned Work* is organized into four parts. First, *Analysis* explains the steps to acquire the necessary information to generate an overview of the current state of data exchange procedures in the German healthcare system. With this information, the requirements are stated for the concept of a DLT. Second, *Design Concept* uses the findings from the first part and explains the design and concept of a DLT as part of the solution. Third, *Evaluation* shows how the prototype is tested based on benchmarks given by the requirement analysis such as security features. This evaluation takes the node, content, ledger and network into consideration in order to give an overview over the strengths and weaknesses of the prototype. And finally, *Implementation* introduces the architecture and technologies used for the DLT prototype.

3.1 Analysis

This part is split into four parts *Law and Regulations*, *Process and Data*, *Requirement Analysis* and finally *Models*. *Law and Regulations* introduces the current applicable laws and regulations for the exchange of healthcare related data. *Process and Data* gives an overview of the different types of exchanged data. The given information is generated using interviews and questionnaires with the different parties in the German healthcare system. *Requirement Analysis* shows the procedures and the questions for the interviews in order to gather and complement existing information. This step summarizes the findings so far and provides the blueprint for the design and implementation of a DLT. Finally, *Models* present DLT based applications, that are used to compare and further modify the requirements created in the previous step.

3.1.1 Laws and Regulations

A data security officer in a primary care setting is contacted and paid accordingly to provide the listing of the laws and regulations that apply in the given setting. This listing forms the boundary for the DLT prototype. In general, the GDPR applies but further aspects need to be investigated.

3.1.2 Process and Data

The current processes and procedures in the German healthcare system are analyzed by contacting primary and secondary healthcare providers, private and public health insurances and finally clients using health related services. In total two insurance companies (a public and private each), three healthcare providers (a general

practitioner, an ophtalmologist and a dentist) and two clients are interviewed. These findings gives an overview over the current processes of data exchanges in the health-care industry. Moreover, these interviews help to understand what type of data are exchanged, which allows to categorize data according to sensitive and non-sensitive data.

Questions to primary and secondary health care providers

1 What data do you collect from a new client?

2 What data do you collect from a regular client?

3 What data do you collect during a consultation?

4 What data do you collect after a consultation?

5 Where do you store which data?

6 What channels and media do you use to collect and transfer data?

7 When and what data do you request and send to other healthcare providers?

8 When and what data do you request and send to the insurance company?

9 When and what data do you request from and send to the client?

10 When and how often is data requested or modified on a client's request?

11 For how long is data stored?

12 What happens to the data, when a client passes away?

13 Has there ever been a data breach?

14 If yes, what data has been stolen or leaked?

15 Did GDPR have an impact on how you perform your work in this regard?

16 If yes, what changes came into effect?

17 Do you employ or use the service of a data protection officer?

18 How often do you consult a data protection officer?

19 Are there any other points that are important to you, that I forgot to ask in regard to data collection and transmission?

Questions to private and public health insurance companies

1 What data do you collect from a new client?

2 What data do you collect from a regular client?

3 When and what data do you collect when a client uses health related services?

4 What data do you request and send to other insurance companies?

5 What data do you request and sent to other institutions?

6 When and how often is data requested or modified on a client's request?

7 When and how often is data requested or sent to a healthcare provider?

8 For how long is data stored?

9 What happens to the data, when a client switches the insurance policy in house?

10 What happens to the data, when a client switches to another insurance company?

11 What happens to the data, when a client passes away?

12 Has there ever been a data breach?

13 If yes, what kind of data has been stolen or leaked?

14 Did GDPR have an impact on how you perform your work in this regard?

15 If yes, what changes came into effect?

16 Do you employ or use the service of a data protection officer?

17 How often do you consult a data protection officer?

18 Are there any other points that are important to you, that I forgot to ask in regard to data collection and transmission?

Questions to clients using health related services

1 What data do you collect when you are looking for health related services?

2 What type of information do you collect before visiting a healthcare provider?

3 What data do you provide when visiting a healthcare provider?

4 What data do you provide to your insurance company?

5 How often do you request your data from healthcare providers or insurance companies?

6 What data do you request from the healthcare providers?

7 What data do you request from the insurance companies?

8 How often do you request any corrections on your data?

9 How often do you change your health care provider?

10 Do you have a complete medical record at your disposal?

11 Do you store health related information on your mobile phone such as heart frequency, blood pressure, weight, blodd oxygen saturation or other vital parameters?

12 Do you use your mobile phone when using health related information or services?

13 If yes, do you send sensitive personal data to other healthcare providers, insurance companies or friends with your mobile phone?

14 Do you use other devices to manage your health related information?

15 Has there ever been a data breach or a cyber attack on one of your devices?

16 If yes, what kind of data has been stolen or lost?

17 Since beginning of 2018, did the way how you interact with health related services change from your perspective?

18 If yes, what did change from your perspective?

19 Are there any other points that are important to you, that I forgot to ask in regard to data collection and transmission?

The answers of these three entities are analyzed and the data categorized into sensitive and non-sensitive components, if applicable. In addition, current data exchange practices are illuminated and scrutinized. Based on these findings, a map of current procedures is created forming the orientation plan for the requirement analysis.

3.1.3 Requirement Analysis

Combining the findings from *Law and Regulations* and *Process and Data* an analysis is performed over the given procedures, data and applicable regulations. These findings provide the source for the next questions asked to the same primary and secondary healthcare providers, insurance companies and clients. For these three entities, the questions are:

1 What tasks have to be solved when exchanging health services related data with other parties?

2 Which problems and frustrations do you face along such data exchanges?

3 What health services related data do you consider to be the most sensitive?

4 And what of this data do you consider to be the most valuable for you personally?

5 In which cases would a new solution support you?

6 What security features should this new solution include: *privacy, confidentiality, anonymity, data integrity, control*?

7 What further features should this new solution include: *openness, interoperability, reliability, accountability and auditability*?

8 Should there be any exceptions to transfer certain data without your consent such as in emergencies, consultations among medical experts or insurance companies, and even family members or friends?

3.1.4 Models

There already exist applications using DLT that transfer health related data in real world settings. These applications include MedRec, that is currently being tested at the Beth Israel Deaconess Medical Center, USA [19]. In addition, the Estonian government applied blockchain technology on a national scale for the healthcare sector in 2016, which serves as a role model for this thesis. Aside, from these two examples, there exist many companies and applications, that could potentially fulfill the set of deduced requirements. In such a case, this model serves as the DLT prototype and is implemented according to the established requirements.

3.2 Design Concept

This part uses the findings from the previous part to design a DLT prototype. For the design process the methods and suggestions by John F Dooley [13] are used as guideline. The resulting design provides the blueprint for the DLT, that is iteratively assembled starting with the requirements given by the laws and regulations. The goal is to include the legal restrictions as well as forty percent of the requirements given by the prior research and consecutive interviews. Therefore, the design concept is progressively upgraded and implemented in order to build a prototype by including one requirement at a time. This is accomplished by using the divide and conquer approach. Along the implementation for each requirement, this approach is repeated until no major improvements arise. The prototype is built using Truffle, but in case there exists an appropriate reusable model, the model is used and adapted to fulfill the requirements postulated so far. Fortunately, MedRec includes several security features and is furthermore available on github making it the perfect candidate for the prototype.

3.3 Evaluation

The prototype is evaluated along its assembly based on security parameters such as privacy, confidentiality, data integrity, data control and reliability of the database. On top of these benchmarks openness, interoperability, accountability and auditability are evaluated as well. As a guideline, the evaluation process is based on the recommendations by John F Dooley [13] as well as Oliver Jung, master student at the Free University Berlin. Finally, at the end of the design and concept process, user tests are run using participants representing one of the three roles: client, healthcare provider or insurance company. In case user tests may not be performed, a set of nodes is created using the Ethereum testnet on a local network with each node having the protocol to run this prototype as a client, healthcare provider or insurance company.

After this steps, the prototype is adapted to at least fulfill the requirements of current laws and regulations. In addition, necessary modifications are performed in order to submit the prototype as final result on github under the appropriate software licenses.

3.4 Implementation

The part *Implementation* shows how the final product is implemented based on the iteratively built prototype. *Architecture* explains the design and structure of the software application that allows for secure health services related data exchanges. Finally, *Technologies* dives into the different tools that are built in and on top of the final product.

3.4.1 Architecture

The planned software architecture relies on the Ethereum network and consists of a client and server sided application which is called dapp. This dapp allows users to access the smart contracts or data on the Ethereum ledger such that they can perform the data exchanges and requests using DLT as its backbone. In case of a reusable model such as MedRec, the architecture of the model is analyzed and further implementations with respect to the requirements discussed. In general, the architecture of the software relies on a DLT as backbone, that acts as the server. The Ethereum network is the DLT of choice, because smart contracts are enabled with this technology, which allows the implementation of server sided logic. The user interface consists of a frontend application written in HTML5, CSS and Javascript, which allows to access the data on the ledger and perform the required operations on such data.

3.4.2 Technologies

For the final product different technologies are implemented, which are presented in this part. The Ethereum network is the fundamental DLT on which the dapp is built upon. This DLT provides further tools that enhances the functionality such as Whisper, Swarm and Interplanetary File System.

Ethereum is the DLT of choice and provides the infrastructure to build and run the final product as a dapp. The Ethereum is a distributed network of nodes, that allows to run smart contracts with a turing complete programming language. Thus, the full set of server sided logic solutions are available for this technology. In addition, Ethereum provides a testing environment for applications such that no coins need to be purchased in order to run and test the final product [22].

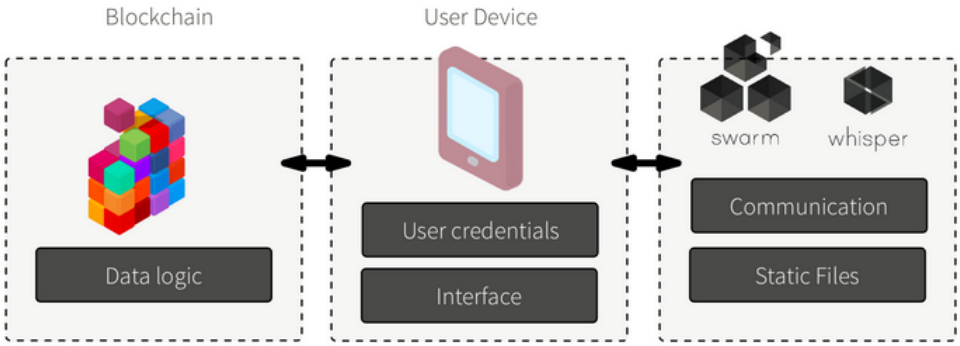

Figure 3.1: The architecture of a dapp is similar to the classical Client-Server Model with the exception, that the server is now a distributed database represented by the blockchain. Inherent technologies of the Ethereum network such as Swarm and Whisper can be used on top of the application. The dapp runs on a user device which can access data and smart contracts on the blockchain with the appropriate credentials [35].

DApp uses HTML, CSS, JavaScript and Solidity for its frontend and backend, but other programming languages such as Python, LLL and C++ can be used as well. For this thesis, the dapp is created with the help of Truffle, that includes all the necessary tools for the development of dapps in the Ethereum network. In addition, Truffle provides different boxes for different end users or use cases [39]. This dapp forms the interface to the DLT for clients, healthcare providers and health insurances from the mobile or desktop environment.

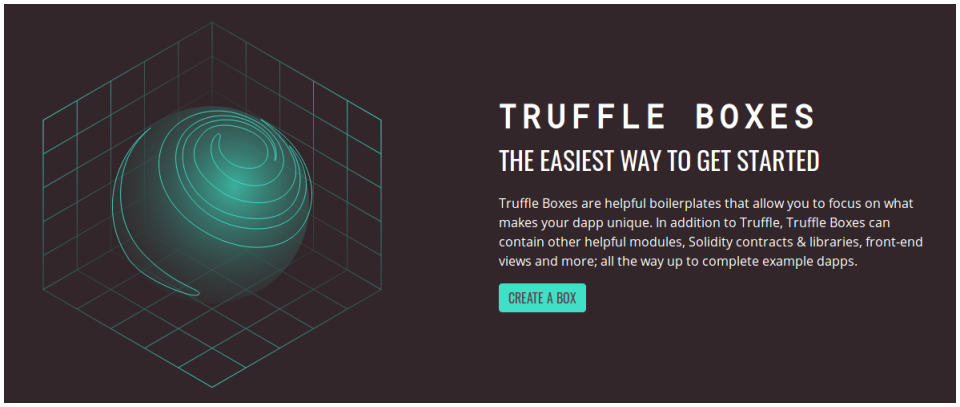

Figure 3.2: The Truffle Suite provides Boxes that include all the necessary tools to start developing a dapp. These boxes include all the necessary libraries that make writing dapp front-ends easier and furthermore provide the necessary tools to develop for specific end user platforms like Android [38].

Whisper is a messaging system that facilitates secure and decentralized communications on top of the Ethereum network. This system can be used to contact insurance

companies, health care providers and other clients with each other and request data exchanges [3].

Interplanetary File System (IPFS allows for a distributed storage system by resolving human readable addresses into hash values by double hashing the data. When data is moved somewhere else, the address still resolves to the actual data on a new site [3]. IPFS serves as a solution on how to externalize encrypted personal health related data.

Swarm is a distributed storage platform and content distribution service and acts as a native base layer service of the Ethereum web3 stack. This platform provides the storage solution for certain data, that needs to be specified [3].

4 Schedule

In section *Schedule* an approximate schedule of the planned work is provided. *Task Schedule* specifies the time frames for the research, analysis, design concept, implementation, evaluation and writing along this thesis. In addition, the section *Milestones* clarifies the respective tasks and their completion weeks.

4.1 Task schedule

Task	May 2019				June 2019	
	W-01	W-02	W-03	W-04	W-05	W-06
Research	yes	yes	yes	yes	no	no
Analysis	no	yes	yes	yes	yes	yes
Design Concept	no	no	yes	yes	yes	yes
Implementation	no	no	no	yes	yes	yes
Evaluation	no	no	no	no	yes	yes
Writing	yes	yes	yes	yes	yes	yes

Task	June 2019		July 2019			
	W-07	W-08	W-09	W-10	W-11	W-12
Research	no	no	no	no	no	no
Analysis	no	no	no	no	no	no
Design Concept	yes	yes	no	no	no	no
Implementation	yes	yes	yes	yes	yes	no
Evaluation	yes	no	yes	yes	yes	no
Writing	yes	yes	yes	yes	yes	yes

Table 4.1: Schedule of Bachelor Thesis

4.2 Milestones

- Milestone 1: W-04 - Research completed
  The sections Introduction and in *Background* the parts *Related Systems* and *Related Work* are completed and written.
- Milestone 2: W-06 - Analysis completed
  In section *Planned Work* part *Analysis* is completed, analyzed and written.

- Milestone 3: W-08 - Design Concept completed
  In section *Planned Work* part *Design Concept* is completed and written. In addition, a minimal viable prototype is built and ready for testing and evaluation purposes.
- Milestone 4: W-11 - Implementation completed
  In section *Planned Work* part *Implementation* a final product is built based on the prototype. The software architecture and implemented technologies are written
- Milestone 5: W-11 - Evaluation completed
  The evaluation of the final product is completed and written based on the suggestion by John F Dooley [13] and recommendations by Oliver Jung. Based on the findings, the final product is modified and reevaluated.
- Milestone 5: W-12 - Writing completed
  The thesis is completed and written.

5 References

[1] **Nicola Atzei, Massimo Bartoletti, and Tiziana Cimoli**. “A survey of attacks on Ethereum smart contracts.” In: *IACR Cryptology ePrint Archive* 2016 (2016), p. 1007.

[2] **Asaph Azaria, Ariel Ekblaw, Thiago Vieira, and Andrew Lippman**. “Medrec: Using blockchain for medical data access and permission management”. In: *2016 2nd International Conference on Open and Big Data (OBD)*. IEEE. 2016, pp. 25–30.

[3] **Joseph J Bambara, Paul R Allen, Kedar Iyer, Rene Madsen, Solomon Lederer, and Michael Wuehler**. *Blockchain: A practical guide to developing business, law, and technology solutions*. McGraw Hill Professional, 2018.

[4] **Shehar Bano, Alberto Sonnino, Mustafa Al-Bassam, Sarah Azouvi, Patrick McCorry, Sarah Meiklejohn, and George Danezis**. “Consensus in the age of blockchains”. In: *arXiv preprint arXiv:1711.03936* (2017).

[5] **Moritz Becker**. “Understanding users’ health information privacy concerns for health wearables”. In: *Proceedings of the 51st Hawaii International Conference on System Science*. University of Hawaii. 2018, pp. 3261–3270.

[6] **Ken Binmore**. *Game theory: a very short introduction*. Oxford University Press, 2007.

[7] **Ethan Cecchetti, Fan Zhang, Yan Ji, Ahmed Kosba, Ari Juels, and Elaine Shi**. “Solidus: Confidential distributed ledger transactions via PVORM”. In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM. 2017, pp. 701–717.

[8] **Melissa Chase and Sarah Meiklejohn**. “Transparency overlays and applications”. In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM. 2016, pp. 168–179.

[9] **Raymond Cheng, Fan Zhang, Jernej Kos, Warren He, Nicholas Hynes, Noah Johnson, Ari Juels, Andrew Miller, and Dawn Song**. “Ekiden: A platform for confidentiality-preserving, trustworthy, and performant smart contract execution”. In: *arXiv preprint arXiv:1804.05141* (2018).

[10] **European Commission**. *2018 reform of EU data protection rules | European Commission*. https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en [Online; last accessed 2019-03-04 10:04]. 2018.

[11] **Corda**. *Corda | Industry*. https://www.corda.net/discover/industry.html [Online; last accessed 2019-03-04 10:01]. 2018.

[12] **Dr. Fred Davis**. *Physicians Under Pressure To Sell Practices: How to Remain Independent*. https://physiciansnews.com/2015/01/27/physicians-under-pressure-to-remain-independent/ [Online; last accessed 2019-03-04 09:22]. 2018.

[13] **John F Dooley**. *Software Development, Design and Coding: With Patterns, Debugging, Unit Testing, and Refactoring*. Apress, 2017.

[14] **Daniel Drescher**. *Blockchain basics*. Springer, 2017.

[15] **Ben Edgington**. *LLL Introduction - LLL Compiler Documentation 0.1 documentation*. https://lll-docs.readthedocs.io/en/latest/lll_introduction.html [Online; last accessed 2019-03-04 09:55]. 2017.

[16] *e-Health Records - e-Estonia*. https://e-estonia.com/solutions/healthcare/e-health-record/ [Online: last accessed 2019-03-05 15:37]. 2019.

[17] **Taavi Einaste**. *Blockchain and healthcare: the Estonian experience*. https://nortal.com/blog/blockchain-healthcare-estonia/ [Online; last accessed 2019-03-04 09:21]. 2018.

[18] **Ariel Caitlyn Ekblaw**. "Medrec: blockchain for medical data access, permission management and trend analysis". PhD thesis. Massachusetts Institute of Technology, 2017.

[19] **Ariel Ekblaw, Asaph Azaria, John D Halamka, and Andrew Lippman**. "A Case Study for Blockchain in Healthcare:"MedRec" prototype for electronic health records and medical research data". In: *Proceedings of IEEE open & big data conference*. Vol. 13. IEEE. 2016, p. 13.

[20] **Ittay Eyal and Emin Gün Sirer**. "Majority is not enough: Bitcoin mining is vulnerable". In: *Communications of the ACM*. Vol. 61. 7. ACM, 2018, pp. 95–102.

[21] **Pietro Ferraro, Christopher King, and Robert Shorten**. "Distributed Ledger Technology, Cyber-Physical Systems, and Social Compliance". In: *arXiv preprint arXiv:1807.00649* (2018).

[22] **Ethereum Foundation**. *Ethereum Project*. https://www.ethereum.org/ [Online; last accessed 2019-03-04 09:50]. 2018.

[23] **The Medical Futurist**. *Top 12 Companies Bringing Blockchain To Healthcare*. https://medicalfuturist.com/top-12-companies-bringing-blockchain-to-healthcare [Online; last accessed 2019-03-04 09:59]. 2018.

[24] **Stuart Haber and W. Scott Stornetta**. "How to time-stamp a digital document". In: *Conference on the Theory and Application of Cryptography*. Springer. 1990, pp. 437–455.

[25] **HSBlox**. *Blockchain Healthcare and Distributed Ledger Tech Solutions*. https://hsblox.com/solutions/ [Online; last accessed 2019-03-04 09:57]. 2018.

[26] **Ari Juels, Ahmed Kosba, and Elaine Shi**. "The ring of Gyges: Investigating the future of criminal smart contracts". In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM. 2016, pp. 283–295.

[27] **Grace Kenny and Regina Connolly**. "Drivers of Health Information Privacy Concern: A Comparison Study". In: *Proceedings of the 22nd Americas Conference on Information Systems*. AMC. 2016, pp. 1–10.

[28] **Ahmed Kosba, Andrew Miller, Elaine Shi, Zikai Wen, and Charalampos Papamanthou**. “Hawk: The blockchain model of cryptography and privacy-preserving smart contracts”. In: *2016 IEEE symposium on security and privacy (SP)*. IEEE. 2016, pp. 839–858.

[29] **Kai-Fu Lee**. *AI Superpowers: China, Silicon Valley, and the New World Order*. Houghton Mifflin Harcourt, 2018.

[30] **Xiaoqi Li, Peng Jiang, Ting Chen, Xiapu Luo, and Qiaoyan Wen**. “A survey on the security of blockchain systems”. In: *Future Generation Computer Systems* (2017).

[31] **Loi Luu, Duc-Hiep Chu, Hrishi Olickel, Prateek Saxena, and Aquinas Hobor**. “Making smart contracts smarter”. In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM. 2016, pp. 254–269.

[32] **Medicalchain**. *Medicalchain - Blockchain for electronic health records*. https://medicalchain.com/en/ [Online; last accessed 2019-03-04 10:05]. 2019.

[33] **Catherine Mulligan, Jennifer Zhu Scott, Sheila Warren, and JP Rangaswami**. “Blockchain beyond the hype; a practical framework for business leaders”. In: *White Paper of the World Economic Forum*. WEF. 2018.

[34] **Koshik Raj**. *Foundations of Blockchain*. Packt Publishing, 2019.

[35] **Alex Van de Sande**. *How to build server less applications for Mist*. https://blog.ethereum.org/2016/07/12/build-server-less-applications-mist/ [Online; last accessed 2019-03-06 20:47]. 2016.

[36] **Eli Ben Sasson, Alessandro Chiesa, Christina Garman, Matthew Green, Ian Miers, Eran Tromer, and Madars Virza**. “Zerocash: Decentralized anonymous payments from bitcoin”. In: *2014 IEEE Symposium on Security and Privacy*. IEEE. 2014, pp. 459–474.

[37] **Yonathan Sompolinsky, Yoad Lewenberg, and Aviv Zohar**. “SPECTRE: Serialization of proof-of-work events: confirming transactions via recursive elections. 2016”. In: *Cryptology ePrint Archive* (2016).

[38] **Truffle**. *Truffle Suite | Boxes*. https://truffleframework.com/boxes [Online; last accessed 2019-03-04 10:10]. 2019.

[39] **Truffle**. *Truffle Suite | Sweet Tools for Smart Contracts*. https://truffleframework.com/ [Online; last accessed 2019-03-04 10:09]. 2019.

[40] **Universa**. *Decentralized autonomous organization—What is a DAO company?* https://medium.com/universablockchain/decentralized-autonomous-organization-what-is-a-dao-company-eb99e472f23e [Online; last accessed 2019-03-04 10:07]. 2017.

[41] **Douglas Wikström**. “Simplified Universal Composability Framework”. In: *Proceedings, Part I, of the 13th International Conference on Theory of Cryptography*. Vol. 9562. Springer. 2016, pp. 566–595.

[42] **Xiao Yue, Huiju Wang, Dawei Jin, Mingqiang Li, and Wei Jiang**. “Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control”. In: *Journal of medical systems* 40.10 (2016), p. 218.

[43] **Aviv Zohar**. “Securing and scaling cryptocurrencies.” In: *IJCAI*. IJCAI, Inc. 2017, pp. 5161–5165.

[44] **Guy Zyskind, Oz Nathan, and Alex ’Sandy’ Pentland**. “Decentralizing privacy: Using blockchain to protect personal data”. In: *2015 IEEE Security and Privacy Workshops*. IEEE. 2015, pp. 180–184.