# 基于患者E-Health诊治数据的安全存储与管理

## 摘要

着眼于传统医疗信息系统中存在的医疗信息的安全存储和授权共享的问题，结合以太坊平台，设计了一个基于电子病历安全存储与授权共享系统。该设计通过以太坊平台来存储患者的电子病历信息和访问权限。患者可以安全有效地管理自己的诊治数据，还可以授权给医疗机构等第三方进行读取。该设计有效的实现了患者对于自身医疗数据的访问控制权限和对数据的安全存储。

关键词：安全存储、授权共享、以太坊、电子病历、访问控制权限

## Abstract

Focusing on the problem of secure storage and authorization sharing of medical information in traditional medical information systems, combined with the Ethereum platform, a secure storage and authorization sharing system based on electronic medical records was designed. The design stores the patient's electronic medical record information and access rights through the Ethereum platform. Patients can manage their diagnosis and treatment data safely and effectively, and can also be authorized to read by third parties such as medical institutions. The design effectively realizes the patient's access control authority for his own medical data and secure storage of data.

## 绪论

### 研究背景、目的与意义

医疗健康问题与每一个人都息息相关，患者的医疗记录（包括病历、化验单、CT图等）都是患者个人健康状况的有效证明，对于患者了解自身健康状态和调养十分重要[1]。然而我国的医疗发展起步晚，医疗设施落后，医疗信息化和数字化进程缓慢，对于病人的医疗隐私数据也缺乏安全有效的管理。医疗信息数字化是未来的发展趋势，利用现代先进的计算机技术，可以减少传统医疗系统中重复的人力操作，使得诊治过程规范化。其中医疗数据（处方、病历记录、身份信息、文档信息）的安全存储和管理更是重中之重，在传统的解决方案中，需要提高电子医疗数据的权限、隐私、安全等，以维护数据的完整性、正确性、可靠性、安全性和隐私性。

另一方面，当前时代是大数据时代，数据的作用被极大的释放，医疗数据更是数据分析的宝贵来源。在传统的医疗信息系统中，由于信息缺乏安全存储，信息容易被不法分子利用，进行贩卖。同时，由于各种医疗机构之间的不信任，医疗数据的共享更是困难，逐渐的形成了"数据孤岛"。数据的不流通导致了协同诊治平台的推进困难，患者往返于各种医院和医疗机构，无法综合各个机构的诊治结果形成完善的就诊报告。

综上所述，设计一套可以独立于第三方医疗机构，对患者个人的医疗数据进行安全存储，同时可以由患者授权给第三方读取，方便不同的医院和医疗机构进行协同诊治的系统十分必要。

### 国内外研究现状

21世纪是信息化时代，个人数据被大量收集和利用，医疗数据也不例外。对医疗数据的挖掘和分析，有利于医疗机构进行对于疾病的研究和防控，但同时也会带来隐私问题。隐私保护是针对公开发布的数据采取的一系列措施，防止第三方通过数据挖掘等不法手段获取病人的敏感信息。在数据发布领域，学者提出了多种隐私保护模型，主要包括：t-closeness、I-多样性、k-匿名模型等。这些模型的原理都是通过隐藏公开的隐私数据和具体个人之间的关系，但是保证发布数据的信息公开可用[2]。

### 论文内容与组织

本文主要描述电子病历安全存储和授权共享系统的实现。本文的主要内容共分为六章。

第一章，绪论，分析我国医疗系统数字化的进程，讨论了医疗数据安全存储和授权共享的重要性，对比了国内外关于隐私数据存储算法的发展和趋势。

第二章，相关理论介绍，介绍本文中研究设计的系统中所涉及的关键理论，主要包括：区块链技术、比特币技术、以太坊平台等。

第三章，提出基于以太坊平台的电子病历存储和授权共享方案，详细讨论了系统的整体框架和所用技术，经过编程验证得出实验结果。

第四章，主要讨论了系统设计开发过程中遇到的困难和解决办法。

第五章，总结和展望。

## 本论文相关理论

### 区块链技术

#### 定义

2008年，中本聪在《Bitcoin：A peer-to-peer Electronic Cash System》一文中首次提出了“区块链”的概念。但是该文着重描写比特币系统，对于区块链技术并没有给出具体的定义。在中本聪的论文中，区块和链被描述为一种用于记录比特币交易过程中交易历史的数据结构。维基百科对于区块链的定义是：区块链（英语：blockchain或block chain）是一串连接起来的加密过的交易记录，每一个记录被称为一个区块。每一个区块包含了前一个区块的hash值、对应时间的时间戳以及交易记录，上述特点使得区块内容具不可篡改性。本文认为区块链既是一种数据结构，又可以认为是分布式数据库，从广义上看它也代指一系列分布式记账技术，包括智能合约、共识机制等。

#### 区块链体系架构

区块链平台自底向上可以分为数据层、网络层、共识层、合约层、应用层五个层次[5]。数据层采用默克尔树、交易结构、签名结构等数据结构与数据库对区块、交易进行存储；网络层采用P2P协议完成各个节点间的数据传输；共识层通过对应的算法和激励机制，解决了拜占庭容错和分布式一致性问题；合约层通过编写智能合约，让开发者再合约上可以进行交易；应用层提供公开的API接口，允许用户创建和运行智能合约。

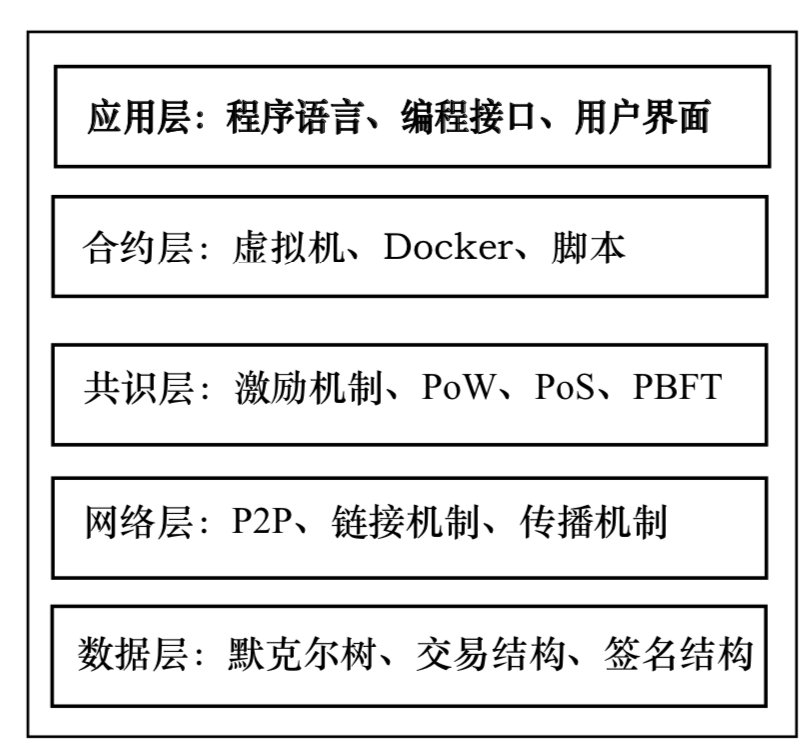


图 区块链技术架构

#### 区块链的应用场景

根据区块链的设计技术，区块链系统具有分布式高冗余存储、不可篡改、去中心化、自动执行的智能合约等特点。区块链不仅可以应用在金融领域，在医疗、社会系统中也有着巨大的应用潜力。区块链目前的应用场景主要有以下几方面：数字货币、数据存储、数据鉴证、金融交易、资产管理和投票选举六大场景[7]。

#### 工作原理

传统的金融交易系统，都需要可信赖的第三方金融机构作为担保，实践证明这种机制一直运转良好，但是这类系统仍然天生受制于“基于信任的模式”的弱点。传统的交易系统并不能实现完全不可逆的交易，因为现实中存在的第三方机构不能保证不会发生分歧。另一方面，第三方金融机构的存在，一定程度上增加了交易的复杂性。所以，人们需要一种电子支付系统，基于密码学原理而不基于信用。达成共识的买方和卖方可以直接进行交易，交易过程不需要第三方金融机构的参与，并且产生的交易是不可逆的。

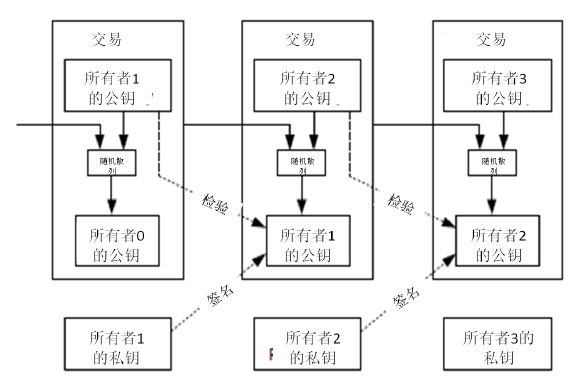


图1 基于数字签名的电子货币

区块链或者说比特币的出现解决了上述难题，中本聪的论文中提出了一种完全通过peer-to-peer技术实现的电子现金交易系统。一般的电子交易系统无法解决"双重支付"的难题。所谓"双重支付"指的是同一枚电子货币被支付两次，为了解决"双重支付"的问题我们又要引入第三方机构来作为公证，这无疑不符合去中心化的特点。因此中本聪提出了一种解决方案，该系统通过随机散列对全部交易记录加上时间戳，将他们加入到一个不断延伸的基于hash值的工作量证明的链条作为交易记录，除非重新完成所有的工作量证明，否则这条链条被认为是不可更改的。只要互联网中大多数的计算能力没有合起来对系统发起攻击，那么可以认为这条链条是安全的[3]。

#### 共识机制

共识机制常见于区块链领域中，即多个节点如何达成共识的机制。由于区块链去中心化的特点，区块链中的各个节点是分布式的。要维护系统的运作顺序和公平性，必须设计一套算法，来统一调度和提供激励机制和惩罚机制。这样的制度，必须依赖某种方式来证明，由哪个节点获得记账权，并且可以获取记录这一个区块的奖励；或者惩罚攻击诚实节点的攻击者。这就是共识机制。常见的共识机制有以下几种：

1. **工作量证明**（Proof-of-Work，PoW），典型案例：比特币网络
2. **权益证明**（Proof-of-Stake，PoS，又译持有量证明），典型案例：以太坊
3. **股份授权证明**（Delegated-Proof-of-Stake，DPoS），典型案例：EOS
4. **容量证明**（Proof-of-space，PoSpace，又称 Proof-of-Capacity，PoC）

本文主要讨论工作量证明：

工作量证明（Proof-of-work）是比特币和以太坊等货币中最常用的算法，其核心依赖于密码学。在分析工作量证明算法前先介绍散列函数（哈希函数）的概念。散列函数是一种函数映射关系，可以将任何长度的数据映射到同一大小的数据。以MD5算法举例如下：

MD5(“Hello World”) = B10A8DB164E0754105B7A99BE72E3FE5

MD5(“Hello World!”) = RNG76287532E96365E841E92BFC50EDG

由上述结果可见，原文细小的改变都会导致随机散列值巨大的变化。

工作量证明算法中，区块链通过共识算法来推选一位Leader来记录下一个区块的内容。Leader还要负责把该区块广播到外界网络，方便其他节点检验该区块的有效性。对该问题简单叙述如下：给定数据X，找到一个数n，使得X+n得到的哈希值是一个小于Y的数。举例如下：

Y = 10，X = “Hello World!”

hash(X) = hash(“Hello World”) = 0x0f = 15 ＞ 10

hash(x+1) = hash(“Hello World1”) = 0xff = 255 ＞ 10

hash(x+2) = hash(“Hello World2”) = 0x09 = 9 ＜ 10

因此，找到n为2满足要求。由于MD5函数式加密安全的，所以我们通过暴力求解来尝试所有的组合。我们把最快计算出上述结果的成员称为矿工，每一个区块被开发的时候，矿工会获得一部分奖励。



图2 区块链数据结构

另一方面，工作量证明解决了确定最长的链条作为可信任的区块链的问题，因为最长的链条被认为是安全的。如果有攻击者想要对现在的区块进行篡改，那么攻击者需要完成当前区块之前的所有工作量，并最终超越现有的区块的工作量。经过证明，攻击者想要追赶上现有的区块，其成功概率呈指数化递减。但是，这不意味着区块链系统是绝对安全的。根据摩尔定律，硬件的运算速度在高速增加，而且诚实节点参与的网络状况也会有所变化。为了解决这种不安全的因素，工作量证明将难度与区块的生成速度的平均值相关联，如果区块生成的速度过快，那么难度将会提高。

#### 区块链的类型

区块链技术根据实际应用场景和需求具有公有链、联盟链和私有链三种应用模式[4]。他们的特点如下：

1. 公有链。公有链允许任何人自由进出，共识机制采用pow/pos/Dpos，需要所有人参与记账，需要激励机制，完全去中心化，每秒可以承载最多100笔交易，典型应用场景是加密货币，代表项目为：比特币、以太坊和EOS。
2. 联盟链。联盟链只允许联盟成员参与，共识机制采用分布式一致性算法，记账人由联盟成员协商确定，激励机制可选，弱中心化，每秒可承载最多10万笔交易，典型应用场景包括供应链金融、银行、物流、电商等，代表项目有R3、Hyperledger。
3. 私有链。私有链只允许链的所有者，共识机制为solo/pbft等，记账人为链的所有者，不需要激励机制，强中心化，承载能力视配置而定，典型场景包括大型组织、机构。



### 智能合约

智能合约(smart contract)是一种特殊协议，在区块链内制定合约时使用，当中内含了函数(Function)，也能与其他合约进行互动、做决策、存储资料以及传送以太币等功能。智能合约主要提供验证及执行合约内所订立的条件。

智能合约容许在没有第三方参与的情况下进行可以被信赖的安全的交易，这些交易具有可追踪性和不可篡改行。

智能合约的概念于1994年由从事智能合约和数字货币研究的尼克萨博(Nick Szabo)博士提出，是与互联网同一时期的产物。当时因为智能合约没有可信的执行环境，智能合约只是一种理论，并没有被技术落地。比特币诞生后，技术人员发现比特币的底层技术支撑区块链是智能合约有效的执行环境，以太坊创始人首先发现了这一点，发布了白皮书《以太坊：下一代智能合约和去中心化应用平台》，并一直致力于将以太坊打造成最佳的智能合约平台，所以比特币依赖于区块链，而以太坊复活了智能合约。

比特币是第一个实现了智能合约的产品，比特币的五种标准交易脚本提供了多方共同签名支付的脚本，又称多重签名支付，多重签名支付可以看做是智能合约的一种表现形式。



图 区块链上的智能合约模型结构

### 以太坊

以太坊（英语：Ethereum）最初由 Vitalik Buterin 在2013年提出，是一个开源的具有智能合约功能的公共区块链平台。以太坊通过其专用加密货币以太币(英语：Ether)提供去中心化的虚拟机来处理点对点合约。以太坊创建的目的是创建一个可自我执行、抗审查和可以自我维护的去中心化的世界级计算机。它延伸了比特币的区块链概念：在分布式的计算机上存储、交易和验证数据。以太坊（Ethereum）在比特币的概念上作出了更加巨大的创新，使在被互联网连接的多个计算机上运行代码成为现实。可以简单的理解：比特币的基础区块链上面存储的是交易记录，而以太坊的基础区块链上存储的是一段可运行的程序代码。

#### 以太坊账户

以太坊的账户主要包括以下四个部分：

1. 一个随机值用作计数器，记录交易的次数，保证只能交易一次
2. 账户当前的以太币余额
3. 账户的合约代码(若无则为空)
4. 账户的存储(默认为空)

两个账户之间的交易信息和信息的状态转换构成了以太坊系统的"状态"。以太币(Ether)是以太坊里面的加密用度，主要用来支出交易用度。以太坊有两种类型账户：合约账户CA（Contracts Accounts）和外部账户EOA（Externally Owned Accounts）。

合约账户是智能合约代码用的账户，外部账户是人用的账户；所以合约账户可以存储并执行智能合约代码，它的"智能性"被外部账户激活。合约账户不存储和存储私钥，合约账户还可以调用其他合约。

外部账户是由使用以太坊的人直接使用的账户，可以存储以太币，可以发送交易到合约账户，触发既定的代码逻辑。外部账户由公钥标识，由对应的私钥控制。

当合约账户被调用时，存储在其中的智能合约可以在矿工处的虚拟机中自动执行，并消耗Gas，如果Gas不足则会触发"Out of Gas"异常，被终止执行。无论是合约账户还是外部账户，在以太坊内部都被看做状态对象。其中合约账户存储了余额和代码，外部账户存储了以太币的余额。

#### 消息和交易

以太坊的消息和比特币的交易相似，但是存在三点不同：

1. 比特币的交易只能由外界来创建，以太坊还可以通过合约创建
2. 以太坊消息可以选择包含数据
3. 以太坊消息包含了函数的概念

Ethereum中的“交易”指存储来自外部帐户的消息的签名包。交易包含消息的接收方、用于确认发送方的签名、以太网帐户余额、要发送的数据和两个名为Startgas和Gasprice的值。为了防止代码的指数级爆炸和无限循环，每个交易都需要限制执行代码所涉及的计算步骤，包括初始消息和在执行过程中引发的所有消息。Startgas是一个限制，而天然气价格是矿商每一步计算的成本。如果在交易执行期间“Gas用尽”，所有状态更改将返回到原始状态，但是所支付的交易成本是不可恢复的。如果交易执行中止时仍然存在Gas，则Gas将返回给发送方。所述创建契约具有单独的交易类型和相应的消息类型;契约的地址是根据帐户随机数和交易数据的哈希值计算的。

#### 以太坊状态转换函数

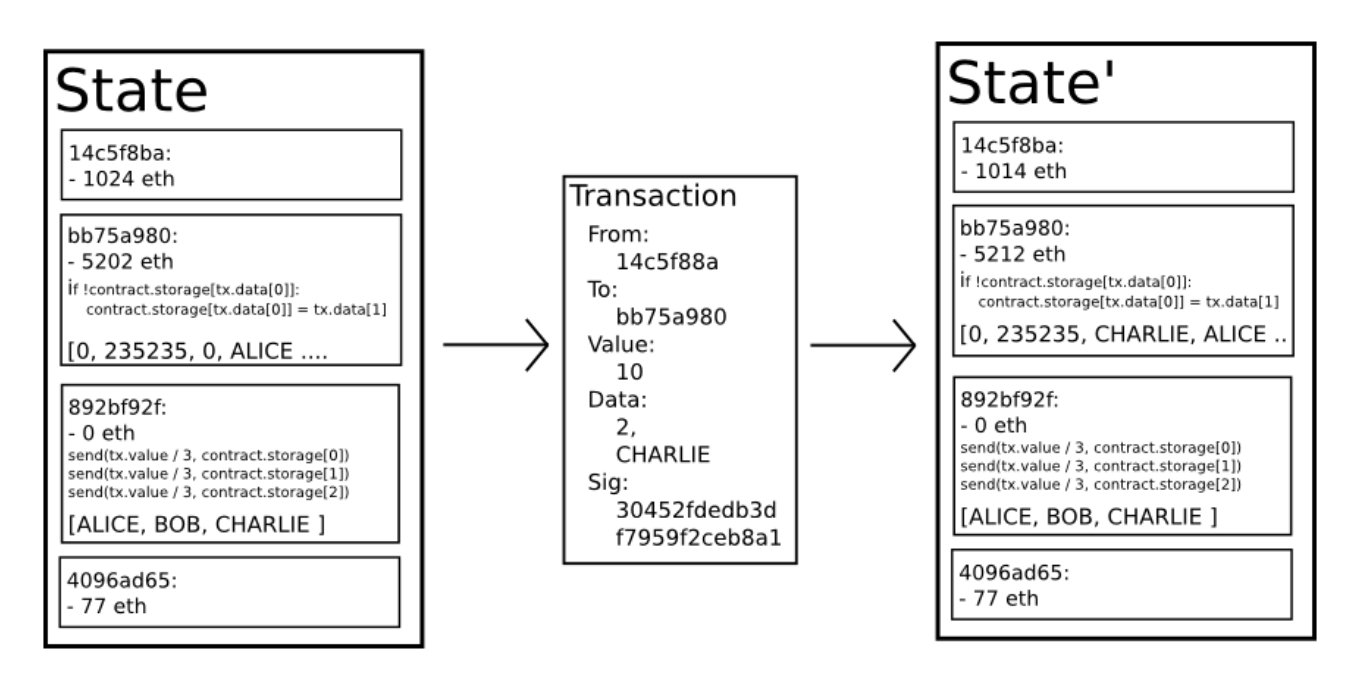


图 以太坊状态转换函数

以太坊的状态转换函数：APPLY(S,TX) -> S'，可以定义如下：

1. 检验输入的格式是否正确。
2. 计算交易费用:cost = STARTGAS \* GASPRICE，根据数字签名找到发送者的地址，并且发送者的账户中减去交易费用和增加发送者的计数器的取值。如果账户余额不足，返回错误。
3. 设定初值cost\_gas = STARTGAS，并根据交易中的字节数减去一定量的Gas。
4. 从发送者的账户转移价值到接收者账户。如果接收账户还不存在，创建此账户。如果接收账户是一个合约，运行合约的代码，直到代码运行结束或者Gas用完。
5. 如果因为发送者账户余额不足或者Gas用完导致转移失败，则恢复账户原来的状态，但是还需要支付交易费用，交易费用转移至矿工账户。
6. 反之，将剩余的Gas归还给发送者，消耗的Gas支付给矿工。

## 基于以太坊的电子病历安全存储和授权共享方案

### 研究目标和研究内容

本系统是对原慢病移动监护协同诊治平台的子系统进行升级。“慢病移动监护协同诊治平台” (简称：“慢病家居监护平台”)，该开发平台通过APP端的手机移动模式，提供直观、方便、快捷的家居移动医疗监护数据的实时采集、上传云服务器，原有平台的功能组成图如下所示：



图 3 平台功能组成图

本系统主要对原医疗协同平台系统中的电子病历管理和权限管理进行升级。



图 4 平台模块调用关系图

原有平台的用户数据、医疗信息数据都存储在医院的关系型数据库中，由医院统一进行管理。现代医疗的一大难题就是数据的获取，高昂的数据成本和患者对于隐私的谨慎使得获得医疗数据十分困难。在传统的医疗信息系统中，患者的医疗信息都由医院统一管理，患者的访问权限十分有限。

针对上述问题，本文提出了一种基于区块链的电子病历管理方案，该方案采用以太坊智能合约为跨医疗机构的医疗数据创建去中心化的病历管理系统。该系统可以为用户提供一个全新的、去中心化的病历管理方案，使用区块链来保存管理电子病历。所有存储在这个系统的记录有以下特点：

1. 患者可以管理自己的病历，
2. 患者可以将自己病历的权限赋予第三方机构，医生获得权限后可以查看和编辑患者的病历
3. 利用独特的区块链属性，以及内含的加密模块，能够在处理敏感信息时为用户提供强大的保密技术

### 拟解决的关键科学问题

去中心化的数据存储。如何保证数据能够安全的存储，不被第三方破解拿到数据，可以由患者自己管理数据，而且不依赖于某个特定的医疗机构。

数据共享。传统医疗机构间共享数据困难，困难点在于医疗数据的校验、保存和同步[6]。由于访问权限的限制，患者、医生和研究人员在读取和共享医疗数据十分困难，需要大量的权限检查和数据校验的过程。

## 基于以太坊的电子病历安全存储和授权共享的具体实现

### 整体框架设计

### 智能合约设计

### 运行环境

### 开发技术

本系统基于Truffle框架搭建以太坊开发平台，通过Ganache搭建本地私有区块链进行验证，其中智能合约的开发语言为Solidity开发语言。

#### Solidity

Solidity是一种面向对象的高级语言，用于实现智能合约。 智能合约是管理以太坊内账户行为的程序。Solidity受C++，Python和JavaScript的影响，旨在针对以太坊虚拟机（EVM）。Solidity是静态类型的，支持继承，库和复杂的用户定义类型以及其他功能。通过Solidity，可以创建智能合约，例如投票，众筹，盲目拍卖和多重签名钱包。

#### Truffle

Truffle是一个基于以太坊虚拟机（EVM）的区块链开发环境，单元测试框架和资源管理中心，意义在于使开发人员的可以更加轻松的搭建以太坊开发环境。 通过Truffle，开发者可以：

* 内置智能合约编译，链接，部署和二进制管理。
* 快速开发的自动合同测试。
* 可编写脚本的可扩展部署和迁移框架。
* 用于部署到任意数量的公共和专用网络的网络管理。
* 使用ERC190标准，使用EthPM和NPM进行包装管理。
* 交互式控制台，用于直接合同通
* 可配置的构建管道，支持紧密集成。
* 在Truffle环境中执行脚本的外部脚本运行器。

#### Truffle-Contract

Truffle-Contract是对Web3.js的封装，是对以太坊智能合约的抽象，适用于Node和浏览器。

Truffle-Contract具有的特性：

* 通过同步交易来获得更好的流量控制(在开发者确认已经开采之前，交易无法完成)
* 承诺。减少回调地狱的产生，适用于ES6的async和await
* 对交易提供默认值，例如address或者gas
* 返回日志、交易的回调、同步产生的交易的哈希值

#### Ganache

Ganache是以太坊开发的个人区块链，可用于部署合同，开发应用程序和运行测试。它既可用作桌面应用程序，也可用作命令行工具（以前称为TestRPC）。Ganache适用于Windows，Mac和Linux。

#### Web3.js

web3.js是一个库集合，允许开发者使用HTTP，WebSocket或IPC连接与本地或远程以太坊节点进行交互。

## 问题与解决办法

### 遇到的问题

首先，区块链技术是比较新颖的一门技术，在对其开发前我们需要先了解其基本原理、工作机制。刚开始学习区块链时遇到的比较大的困难对共识机制的了解，共识机制指的是在分布式的环境下多个节点如何达成共识，在比特币中主要应用了工作量证明(Proof-of-Work)的方法。除此之外，对于时间戳服务器、双重支付的问题的理解也十分关键。

设计采用以太坊作为开发平台，以太坊是一个生态圈，含有许多的开发工具和开发框架。其中最重要的是Solidity语言，Solidity语言用来开发以太坊智能合约的语言，掌握其语法与特性十分重要。其次遇到的困难是学习Truffle框架，Truffle框架是一款以太坊的开发框架，帮助开发者优化了智能合约的编译、发布、交易等流程。Truffle框架还包含了一系列的子框架，包括：Truffle-Contract等。除此之外，对于Ganache等以太坊客户端的学习也十分困难，开发者要访问以太坊区块链必须通过相关客户端提供的API，有官方提供的Go语言客户端还有方便本地使用的Ganache客户端。其次，比较困难的是对于Web3.js库的学习，Web3.js库是一个封装了访问区块链相关API的JavaScrtpt库。简单来说，Web3.js是前端界面与区块链智能合约的桥梁，对于项目的搭建十分重要。

### 解决办法

接触一个新的领域，或者接触一门新的技术，都需要花大量的时间去熟悉才能掌握。特别是对于Solidity语言、Truffle框架、Web3.js库，国内非常少中文资源，最好的学习资料是官方文档。通过查阅官方文档，学习官方文档里的Demo，可以快速的入门。除了官方文档，Github上有非常优秀的开源项目，通过学习前人的代码，可以掌握高质量的开发技巧。

### 开发技巧

本系统的开发主要包括两方面：前端界面的开发和后台以太坊区块链的开发。合理的利用Truffle框架可以帮助开发者一键搭建项目，包括智能合约、前端项目、编译模块和服务器脚本。前端开发要注意与以太坊的接口对接，常见的可以通过Web3.js或者Truffle-contract库进行读取智能合约的示例，从而获取智能合约的数据和调用智能合约的函数。以太坊的开发则要合理利用Ganache工具，在本地搭建私有的区块链作为测试，可以迅速的返回结果，不需要等待出块的时间，方便开发者进行测试和开发。

## 总结与展望

### 总结

本文主要介绍了一个基于以太坊的电子病历安全存储和授权共享系统。对于区块链技术的基础知识和应用给予了简单的解释，还介绍了智能合约和以太坊的相关背景知识。

### 展望

## 致谢

在完成这篇毕业论文的过程中，经历很多很困难的时期，从最初的选题，到项目整体框架的设计，到最后的实现过程，内心感受的变化非常的巨大，毕业设计的完成要比想象中的要难得多。要特别感谢朱诗生老师，从立题开始就为我的设计确立了一个明确的方向，在探索过程中即使遇到困难止步不前也一直鼓励我继续努力，在束手无策的时候又能为我想出解决方案，所以在我的毕业设计中，朱老师功不可没。同时，还要感谢在整个设计过程中提供帮助的黄仁俊师兄，帮我收集相关的资料，督促我的工作。还有某些同学，为我的毕业设计提供了不少宝贵的意见。

最后还要衷心感谢我的家人和朋友，一直以来对我的支持和鼓励。

## 参考文献

[1] 梅颖. 安全存储医疗记录的区块链方法研究[J]. 江西师范大学学报(自然科学版)，2017，41(5):484-490.

[2] 曹敏姿，张琳琳，毕雪华，赵楷. 个性化(α ，l) - 多 样 性 k - 匿 名 隐 私 保 护 模 型[J]. 计算机科学，2018，45(11).

[3] Nakamoto S. Bitcoin: a peer-to-peer electronic cash system [Online], available: <http://bitcoin.org/bitcoin.pdf>, June 12, 2016

[4] 赵延红,原宝华,梁军.区块链技术在医疗领域中的应用探讨[J].中国医学教育技术,2018,32(01):1-7.

[5] 张亮,刘百祥,张如意,江斌鑫,刘一江.区块链技术综述[J/OL].计算机工程:1-15[2019-04-17].http://kns.cnki.net/kcms/detail/31.1289.TP.20190316.1352.002.html.

[6] 薛腾飞,傅群超,王枞,王新宴.基于区块链的医疗数据共享模型研究[J].自动化学报,2017,43(09):1555-1562.

[7] 袁勇,王飞跃.区块链技术发展现状与展望[J].自动化学报,2016,42(04):481-494.