MASTER THESIS XIDIAN UNIVERSITY

西安电子科技大学

硕士学位论文

基于区块链的医疗数据共享隐私保护问题的研究与实现

作者姓名 王 旭

指导教师姓名、职称 覃桂敏 副教授

申请学位类别 工学硕士

学校代码 10701 学 号 1603121626
分 类 号 TP30 密 级 公开

西安电子科技大学

硕士学位论文

基于区块链的医疗数据共享隐私保护问题的研究与实现

作者姓名：王 旭

一级学科：软件工程

二级学科：软件工程

学位类别：工学硕士

指导教师姓名、职称：覃桂敏 副教授

学 院：计算机科学与技术学院

提交日期：2019 年 6 月

Research and Implementation of Privacy Protection for Medical Data Sharing Based on Blockchain

A thesis submitted to

XIDIAN UNIVERSITY

in partial fulfillment of the requirements

for the degree of Master

in Software Engineering

By

Wang Xu

Supervisor: Qin Guimin Title: Associate Professor

June 2019

西安电子科技大学
学位论文独创性（或创新性）声明

秉承学校严谨的学风和优良的科学道德，本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得西安电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文若有不实之处，本人承担一切法律责任。

本人签名：王旭　　日　期：2019.6.17

西安电子科技大学
关于论文使用授权的说明

本人完全了解西安电子科技大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权属于西安电子科技大学。学校有权保留送交论文的复印件，允许查阅、借阅论文；学校可以公布论文的全部或部分内容，允许采用影印、缩印或其它复制手段保存论文。同时本人保证，结合学位论文研究成果完成的论文、发明专利等成果，署名单位为西安电子科技大学。

保密的学位论文在___年解密后适用本授权书。

本人签名：王旭　　导师签名：[illegible]

日　期：2019.6.17　　日　期：2019.6.17

摘要

数字医疗技术的进步带动了医疗数据共享，但是也造成了医疗机构中出现大量异构信息系统等问题，使得医疗数据不能高效互通。另外，医疗数据在各种中间机构之间不断交易共享，患者无法参与自身医疗数据共享过程且隐私泄露严重。区块链是一种按时间顺序将区块以链的形式组织起来的数据结构，使用分布式节点共识算法和密码学算法提供了抗篡改，可追溯和安全等特性，为医疗数据共享注入了活力。针对医疗数据共享中不能保证数据真实性，同时当对数据有争议时，不具备可追溯性等问题，本文实现了基于区块链的医疗数据共享，并结合医疗数据易复制和易传播的特点，从应用层和交易层分别进行了隐私保护机制的研究与实现。

为达到在应用层隐私保护的目的，文中研究实现了基于默克尔树加密的医疗数据隐私保护。首先分析原始医疗数据，发现了医疗数据本身结构复杂且标准不统一等问题，同时考虑到区块链中单个区块容量有限，对医疗数据进行了结构化和标准化处理。预处理后的数据不仅包含原始医疗数据中存储位置、分类标准等关键信息，而且所需存储空间大大减少。其次，为保证医疗数据能够在提供者和需求者之间高效传输，对标准化医疗数据进行关键字提取，构建医疗数据的关键字索引结构。接着为保证数据提供者对数据的完全控制和数据需求者对数据的安全访问，对标准化医疗数据和关键字的索引结构进行加密，通过区块链网络共享密文，形成医疗数据的默克尔树，保证医疗数据的完整性，并且使得数据需求者快速获取密文的同时验证数据需求者的身份。

在交易层研究超级账本 Fabric 交易过程，针对交易数据在分布式账本中公开透明的问题，文中使用群签名和对称加密算法处理交易提案，对交易过程中背书环节和排序环节进行改进以保护隐私性。首先，研究现有群签名算法，结合超级账本 Fabric 特点，使用了基于 Paillier 的群签名方案。其次，针对该群签名方案中群成员和群管理员交互过程中信息泄露的问题，文中将群成员申请入群过程和密钥颁发过程与超级账本中 PKI 体系结合，则群成员和管理员可以在公开信道交互。接着使用了 AES-128 对称加密算法和改进的群签名方案对超级账本 Fabric 的交易提案进行处理。最后，客户端分别发送加密且群签名的交易提案给背书节点和排序节点，进行不同的处理。在交易提案成功写入账本的前提下，保证了匿名性且起到了交易数据隐私保护的作用。

在医疗数据加密隐私保护研究和交易数据的隐私保护研究的基础上，使用了超级账本 Fabric 对数据共享进行了测试验证。实验表明，数据提供者与数据需求者成功共享数据，且交易数据为密文信息。然后对三种小儿白内障图像数据类型进行共享测试，实验表明，随着共享数据数量增大，耗时性能趋势一致。但是由于三类数据样本量和数据中关键字数量不同，不同类别数据共享耗时差别较大，证明结构化医疗数据的大

小对基于超级账本 Fabric 的医疗数据共享耗时影响更大。然后对数据隐私保护中的加密环节分别使用 RSA 和 AES 进行加密共享，实验表明，随着共享数据量的增大，虽然 RSA 的速率慢于 AES，但是可以验证数据需求者身份，保证对数据的安全访问。最后进行多次签名和验签性能测试，文中使用的具有匿名性的群签名方案性能接近于 Fabric 默认签名 ECDSA。

关 键 词：医疗数据共享， 隐私保护， 加密算法，超级账本 Fabric， 群签名

ABSTRACT

Advances in digital medical technology have led to medical data sharing, but they have also caused many problems. For example, a large number of heterogeneous information systems appear in medical institutions, and medical data cannot be efficiently interoperable. In addition, medical data is continuously traded and shared among various intermediary organizations, and privacy leakage is serious. Patients cannot participate in their own medical data sharing process. Blockchain is a data structure that organizes blocks in the form of chains in chronological order. It provides tamper-resistant, traceability and security features by using distributed node consensus algorithm and cryptographic algorithm, which injects vitality into medical data sharing. Aiming at the problems of data authenticity in medical data sharing and lack of traceability when data is disputed, this thesis realizes the medical data sharing based on block chain, and combines the characteristics of easy replication and transmission of medical data.The research and implementation of privacy protection mechanism are carried out from the application layer and the transaction layer respectively.

In order to achieve the purpose of privacy protection at the application layer, the thesis designs and implements the medical data privacy protection algorithm based on Encrypted Merkel Tree. Firstly, we analyze the original medical data and find that the structure of medical data is complex and the standard is not uniform. Considering the limited capacity of a single block in the block chain, the medical data was structured and standardized. The pre-processed data not only contains storage locations and classification criteria in the original medical data, but also requires a storage space of up to 1KB.Secondly, in order to ensure that medical data can be efficiently transmitted between providers and consumers, keyword extraction is performed on standardized medical data and the keyword index structure of medical data is constructed. Then, in order to ensure the complete control of data by data providers and the safe access to data by data demanders, the index structure of the standardized medical data and keywords is encrypted, and the ciphertext is shared through block chain network to form Merkel tree of medical data. The tree guarantees the integrity of the medical data and enables the consumer to quickly obtain the ciphertext while verifying the identity of the consumer.

In the transaction layer, the transaction data is transparent in the distributed ledger. The thesis's proposes to use the group signature and the symmetric encryption algorithm to process the transaction proposal, and improve the endorsement and sorting links in the transaction process to protect privacy. Firstly, the existing group signature algorithm is studied. Combined with the characteristics of the Hyperledger Fabric, the Paillier signature algorithm is used. Secondly, in order to solve the problem of information leakage during the interaction between group members and group administrators in the group signature scheme, this thesis combines the application process of group members and the key issuance process with PKI system of Hyperledger, then the group members and administrators can interact in open channel. The AES-128 symmetric encryption algorithm and the improved group signature scheme were used to process the transaction proposal of the Hyperledger Fabric. Finally, the client sends an encrypted and group-signed transaction proposal to the endorsement node and the sort node for different processing. Under the premise that the transaction proposal is successfully written into the Hyperledger, the anonymity is guaranteed and the privacy of transaction data is protected.

Combined with the research on privacy protection of medical data and transaction data, the data sharing was tested and verified using the Hyperledger Fabric. Three types of pediatric cataract image data are tested for sharing. The experimental results show that with the increase of the number of shared data, the trend of time-consuming performance is consistent. Due to the different types of data and the number of keywords, the time of different data sharing varies greatly. The size of structured medical data has a greater impact on medical data sharing through the Hyperledger Fabric. Then, RSA and AES are used for encryption in privacy protection. Experiments show that, with the increase of the amount of shared data, the RSA rate is slower than AES encryption sharing, but it can guarantee the data demander access to data Safely. Finally, several signature and verification tests are carried out. The performance of the anonymous group signature scheme proposed in this thesis is close to that of Fabric default signature ECDSA.

Keywords: Medical Data Sharing, Privacy Protection, Encryption Algorithm, Hyperledger Fabric, Group Signature

插图索引

表格索引

符号对照表

符号	符号名称
x	标准化医疗数据集
n	医疗数据集数量
R	01 随机生成器
H	哈希函数
m	数据集中关键字总数
k_i	第i个关键字
Z_{n^2}	小于n^2的整数集合
$Z_{n^2}^*$	Z_{n^2}中与n^2互质的整数的集合
ID	用户真实身份
$ord(u)$	u的阶次
r	用户特性参数
Q_{ID}	用户假名信息
Z_{ID}	与Q_{ID}对应的用户身份信息
a	随机整数
b	随机正整数
λ	两个大质数的最小公倍数
H_1	哈希函数
H_2	哈希函数
s_1	群签名部分信息
s_2	群签名中间参数
s_3	群签名部分信息

缩略语对照表

缩略语	英文全称	中文对照
EHR	Electronic Health Record	电子健康记录
POS	Proof of Stake	股权证明
AES	Advanced Encryption Standard	高级加密标准
HIT	Health Information Technology	医疗信息技术
DICOM	Digital Imaging and Communications in Medicine	医学数字成像和通信
DES	Data Encryption Standard	数据加密标准
ECDSA	Elliptic Curve Digital Signature Algorithm	椭圆曲线数字签名算法
PKI	Public Key Infrastructure	公钥基础设施
E-Cert	Enrollment Cert	身份证书
T-Cert	Transaction Cert	交易证书
TLS-Cert	Transport Layer Security Cert	通讯证书
PKG	Private Key Generator	私钥生成器
XML	Extensible Markup Language	可扩展标记语言
HTTPS	Hyper Text Transfer Protocol over Secure Socket Layer	超文本传输安全协议

目录

第一章　绪论

1.1 研究背景及意义

随着信息数字化的发展，医疗领域迎来数字医疗时代。数字医疗是将计算机科学信息技术与整个医疗过程深度融合，产生的一种新型的、现代化的医疗方式[1]。数字医疗技术的进步和发展颠覆了医疗行业的方方面面。其中，电子健康记录系统（EHR）的发展势头最为迅猛。根据调查研究发现，全球经历了一场卫生领域的信息技术转型，即从主要以纸张为记录基础的信息管理系统转变为几乎人人都有个体数字化信息的模式[2]。如今的美国医疗领域中，几乎所有医院（96%）和越来越多的医生（78%）都在使用已权威认证的 EHR，这种转变是 2009 年颁布经济和临床健康信息技术法案的结果。数字医疗的发展是纸质健康记录数字化的过程，也是卫生医疗机构不屈不挠努力的结果[3-5]。

随着时间推移，数据信息化的普及程度越来越高。如今，每个医院各个部门业务系统每天都会产生大量医疗数据，由于各个医院采用不同医疗数据管理系统，这些业务系统之间没有统一的管理方式，导致不同业务系统输出的数据格式和数据类型千差万别，“数据孤岛”现象普遍[6]。不同医院之间甚至医院内部不同部门之间为进行数据共享，挖掘医疗数据价值，进而更好地服务于医疗过程，都在建设数据共享系统。从数据共享过程来看，获取医疗信息的方式越来越便捷，与此同时，医疗信息数据的泄露成本越来越低，个人医疗信息面临的泄露威胁日益增长。IBM Security 和 Ponemon Institute 联合研究报告显示，2018 年全球数据泄露平均成本与去年相比，高出 6.4%。其中，医疗保健领域的泄露成本最高（泄露成本具体是指单位泄露数据样本价值）。同时报告指出，连续 8 年来医疗保健领域数据泄露成本常位榜首，单位丢失医疗数据记录样本的成本为 408 美元，几乎是跨行业平均值（148 美元）的 3 倍[7-9]。

提升医疗数据的使用价值对医疗数据的开放和共享尤为重要，但是只有在保护患者以及其他参与方隐私的基础上才能真正实现医疗数据共享[10]。在区块链中，首先，信息存储在永远不能删除或者更改信息的分布式账本中，从而提供了完整并且不可辩驳的记录。其次，分布式账本分散在全网各个计算节点上，能够抵御故障和攻击。最后，区块链依赖密码学提供了伪匿名和公钥基础架构来保证链上数据的安全和隐私[11]。区块链中去中心化、全网分布式存储和共识算法等为医疗数据共享中医疗数据确权和共享过程中隐私保护问题的解决提供了有效的措施。

1.2 国内外发展概况

为了促进医疗数据流动，实现医疗数据的自身价值，医疗领域中相关参与方都在努力，以实现更好、更安全的医疗数据共享。相关研究主要集中在医疗数据共享平台建设，制定医疗信息标准，健全医疗数据隐私保护的相关法律法规和应用区块链技术的医疗数据共享研究等方面。

1.2.1 医疗数据共享现状

（1）医疗数据共享平台建设

加拿大 Health Infoway 项目服务于患者和医疗保健服务提供者，提供了兼容的电子健康记录。此项目推进了加拿大的数字医疗进程，为促进医疗数据的互操作做出了巨大的贡献[12]。

我国国家人口与健康科学数据共享平台面向全社会开放提供服务，用户可借助平台完成医疗数据互联、医疗数据使用、医疗数据存储、医疗数据挖掘和共享服务等，此平台为医疗领域中人才培养和科技创新提供了基础化设施建设[13]。

中国科学院大学健康医疗大数据遂宁研究中心构建了互联网加智慧医疗的模式，建立了医疗大数据研究中心，同时整合遂宁市的医疗数据，推出健康遂宁的应用程序，使居民对自己的个人健康有全方位了解[14]。

碳云智能是国内第一个具备很强专业能力的健康大数据平台，可通过自身核心技术和合作伙伴提供数据这两种途径来获取医疗信息。该平台利用数据挖掘和机器学习等先进技术，将人工智能的优势与数据分析和应用相结合[15]。

（2）医疗信息标准

美国成立了医疗卫生信息技术标准委员会，主要用于解决医疗数据共享问题，同时提出了一系列被广泛采用的协调不同标准差异的规范[16]。

英国国家卫生局成立了卫生信息标准管理机构，提出了包含技术标准、健康信息标准、医疗数据标准的卫生信息标准体系[17]。

我国医疗领域的医疗信息标准制定起步比较晚。从 2009 年下半年开始，卫生部开始大力推行健康档案电子化，并且先后发布了《电子病历基本架构与数据标准（征求意见稿）》等多个文件通知和标准[18]。

1.2.2 医疗数据隐私保护现状

现如今处于零信任的时代，数据业务逻辑响应越快越好，但与此同时，受到的安全威胁也越来越多，关于个人数据隐私保护的法律法规建设也在不断地完善发展。

早在上世纪 80 年代，由美国，加拿大和欧洲经济共同体成员国等 20 个国家组成

的经济合作与发展组织提出了《保护隐私和个人数据跨境流动》的指导方针[19]。方针重点在于，使用数据前让用户知道、使用数据目的明确、取得用户同意、保障个人数据安全地使用、使用数据过程中用户可以查明验证等，为数据的跨境流动做出了明确规定。虽然该指导方针具有非常强的指导意义，但是各国内部的隐私保护条例并没有统一实施。

2016 年欧盟通过了保护个人数据规定，此规定包含欧盟公民个人数据的整个处理过程，从收集数据，处理数据，储存数据到最终管理数据。规定从“顾客优先”的角度出发，把个人信息的最终控制权返还给公民个人，以限制不经过用户同意，随意收集和使用用户个人信息的行为[20]。

2018 年，国健委印发《关于印发国家健康医疗大数据标准、安全和服务管理办法的通知（试行版）》，从法律角度确保医疗数据隐私安全[21]。在医疗数据共享方面，国家卫生健康委员会建立了健康医疗大数据开放共享机制和数据共享交换体系，规定在共享过程中发生的数据泄露等安全问题由委托单位和受委托单位共同承担安全责任。

2018 年，任倬辉结合医疗数据敏感度高的特点，在差分隐私保护的前提下，引入增量式聚类算法，同时实现了可进行疾病监测的医疗数据分析系统。最后，通过实现证明差分隐私机制可以保证医疗数据动态分析过程中的隐私保护[22]。

据不完全统计，国内个人隐私数据泄露数高达 55.3 亿条左右，平均每人都有 4 条相关的个人数据泄露。数据隐私保护需要一系列技术来实现，常见医疗数据的隐私保护技术为：数据扰乱技术、数据加密技术、数据匿名和访问控制技术[23]。其中，访问控制技术较为普遍和简单，但是用户访问数据过程复杂。

1.2.3 应用区块链技术的医疗数据共享研究现状

互联网时代为人们的生活带来了巨大的变革，但是仍然存在互联网社会中人与人之间个人信用缺失这一黑暗面。区块链技术提供了信任价值转移网络，是目前能想到的、也比较适合于数据隐私保护的解决方案。

2018 年，Jiang S 等人提出了基于区块链的健康信息交换平台，分析不同医疗数据源之间共享数据的需求，使用了两种松耦合的区块链处理不同的医疗数据。采用离线存储和线上验证两种方式处理数据，从而保护了患者隐私信息[24]。

2018 年，Peng Z 等人针对传统临床数据孤岛的问题，在满足数据互操作性路线图的条件下，设计基于区块链技术的架构，并且展示了基于该架构的应用程序，癌症护理协坐研究案例中的参与者使用该应用程序进行身份验证[25]。

2018 年，张圣垚等人使用区块链技术设计实现了电子病历系统并进行系统测试。该系统具备病例管理、AES 和 RSA 相结合的加密管理、系统权限管理、用户身份认证管理、用户管理等模块，解决了传统电子健康病例系统中共享效率低下，隐私数据

泄露严重等问题[26]。

1.3 论文工作内容

本文首先对目前国内外医疗数据共享现状进行研究和分析，重点研究医疗数据共享中的隐私保护机制。本文使用区块链技术来进行医疗数据共享，其次，为解决医疗数据共享中隐私保护问题，从应用层和交易层两个角度出发，分别研究并实现了隐私保护方案，最后将上述隐私保护方案应用于医疗数据共享过程，基于超级账本 Fabric 实现医疗数据共享并且进行测试验证。

本文主要从以下四个方面来开展研究工作：

（1）研究常见医疗数据共享过程，为解决医疗数据共享过程标准不一、隐私安全等问题，使用区块链技术进行医疗数据共享。

（2）为保护医疗数据隐私，首先分析了现有医疗数据，对医疗数据进行预处理，其次，为了提高数据共享效率和保证数据需求者对数据的安全访问，研究并实现了基于关键字的数据索引结构。最后，从应用层出发研究并实现了医疗数据的加密方案。

（3）研究现有区块链交易隐私保护方案，对比分析群签名算法。对群签名方案结合超级账本 Fabric PKI 进行改进，从交易层出发，给出与超级账本 Fabric PKI 机制结合的基于 Paillier 的群签名方案。研究超级账本 Fabric 交易过程，发现了交易数据在分布式账本中公开透明的问题。针对该问题，使用群签名和对称加密算法处理交易提案来进行交易隐私保护，对交易过程中背书环节和排序环节进行改进以保护隐私性。

（4）结合上述应用层和交易层的隐私保护方案，研究并实现基于超级账本 Fabric 的医疗数据共享并进行测试验证。

1.4 论文组织结构

本文分为七部分。各个部分具体内容如下：

第一章：绪论。首先，讲述了本文研究背景和意义；其次，概要分析了国内外在医疗数据共享、数据隐私保护、区块链技术应用于医疗领域等方面的研究现状；接着，给出本文的主要工作内容，最后介绍了本文的整体组织结构。

第二章：本文涉及相关技术概述。首先介绍了医疗数据标准，其次对相关密码学中加密算法、数字签名技术以及群签名技术进行概述，最后讲述了区块链的特点和分类以及常见隐私保护技术。

第三章：医疗数据隐私与医疗数据安全访问控制研究。首先比较分析了医疗数据，依据医疗数据标准对医疗数据进行预处理。然后，为提高数据共享效率和保证数据需求者对数据的安全访问，研究且实现了基于关键字的数据索引结构，最后从应用层保

护数据隐私的角度出发，研究实现了医疗数据的加密隐私保护方案，并给出理论上的分析。

第四章：区块链中交易数据隐私保护研究。首先比较分析群签名方案和研究超级账本 Fabric 交易过程，然后从交易层进行隐私保护的角度出发，研究并实现结合群签名算法和对称加密算法的方案，并详细阐述使用新方案的超级账本 Fabric 交易过程。

第五章：结合医疗数据数据隐私保护研究和区块链中交易隐私保护研究，研究并实现基于超级账本 Fabric 的医疗数据共享过程。搭建超级账本 Fabric 测试环境，使用不同类型医疗数据进行加密共享测试，同时对结合超级账本 Fabric 的群签名和对称加密算法的交易隐私保护方案进行测试验证。

第六章：总结与展望。总结全文研究工作，得出基于区块链的医疗数据共享隐私保护问题的下一步研究重点。

第二章 相关技术和理论

本章节主要介绍文中涉及的相关技术。介绍了数字医疗和医疗数据标准的概念，对加密算法特别是 RSA 算法和 AES 算法进行介绍，对数字签名算法和群签名算法进行概述，同时介绍了区块链的特点、分类以及区块链中常见的隐私保护技术等。

2.1 医疗数据相关概念

2.1.1 数字医疗

数字医疗作为现代化的医疗方式，包括医疗设备的数字化、医疗设备的网络化、医疗机构管理的信息化和医疗服务的个性化[27]。医疗设备的数字化即使用计算机软件控制的医疗设备，在该医疗设备形成的工作网络中，数据的采集和数据的处理以及数据的存储和传输全部使用了现代计算机科学技术。医疗设备的数字化为数字医疗奠定了基础。医疗设备的网络化具体是指某一医疗机构内部医疗设备之间可以实现文档和影像资料的传输共享，采用这种方式，可以减少相关资源浪费和降低发生医疗错误的概率。医疗机构管理的信息化即医疗机构中相关参与人员通过医疗机构内部的电子健康病例系统开展工作，且相关管理人员和审核人员可以访问系统，查看各部门的具体运行情况。医疗服务的个性化即在数字医疗进程的快速推动下，人工智能技术和穿戴设备等新技术融入了传统医疗业务模式，为患者预测疾病的患病概率和发展情况[28]。

2.1.2 医疗数据标准

随着科学技术和医疗的深度融合，医疗数据体量骤然增大。采用医疗数据标准是确保医疗数据流无缝衔接和安全传输实现数据共享的第一步。

研究对比各种医疗信息共享提案，虽然各种健康医疗信息标准层出不穷，然而信息标准的具体实施因各 HIT 厂商和医疗机构的差异化、个体化，无法准确、客观、及时地评价医疗信息标准实施的正确性。更重要的是，缺乏公开和权威的数据交换标准测试中心，无法具体指导医疗信息标准的实施落地[29]。

医疗数据的标准化处理是医疗数据安全共享的前提，也是实现医疗数据价值的必要条件。医疗领域中常见的相关标准主要有国际健康领域数据标准、国际医学数字影像通讯标准、临床数据交换标准协会、国际疾病分类标准等。其中国际健康领域数据标准指的是标准化卫生信息的用户层交换协议，描述了医疗领域中不同应用之间的电子传输协议。医学数字成像和通信标准是和医学图像相关的，被广泛采用的国际标准。目前推出了最新版本 DICOM3.0 标准，按照 DICOM 标准要求存储的医学文件统称为

DICOM 文件[30]。

2.2 加密算法

区块链技术使用了众多密码学算法来确保隐私性和安全性，加密技术在密码学算法中最为常见。

（1）对称加密算法

按照加解密过程是否使用同一密钥的设计理念，加密算法被分为对称加密算法（private key cryptography）和非对称加密算法（public key cryptography）。前者加解密过程使用了同一套密码，后者加解密过程使用不同的两套密码[31]。人类传统的加密方式，例如莫尔斯电码和 Enigma 电报密码都属于对称加密算法。对称加密算法从实现原理上划分，可以分为分组密码和序列密码。目前，被广泛采用的 AES 加密算法，按照实现原理属于分组密码系列算法[32]。

随着计算机硬件的发展，计算机处理能力越来越强。DES 算法密钥长度为 56 比特，则该算法2^{56}数量级的理论安全强度不能确保足够安全。AES 算法是一个分组密码算法族，根据密钥长度不同，分为 AES-128 算法、AES-196 算法、AES-256 算法。虽然这些算法密钥长度不同，但是同属于 AES 算法族，都有着相同的分组长度 128 比特[33]。

（2）非对称加密算法

非对称加密算法和对称加密算法的区别在于，非对称加密算法中加密密钥和解密密钥不同。如果加密密钥作为私有保密的私钥，那么解密密钥是公开的公钥，反之亦成立。非对称加密算法中发送者与接受者之间不需要密钥传输操作，从而保证了无密钥传输的保密通信。非对称加密算法的安全性依赖于数论中计算复杂度比较高的难题。作为非对称加密中被广泛使用的 RSA 算法由 RSA 公司发明[34]，是一个支持变长密钥的公共密钥算法，需要加密文件块的长度也是可变的。

非对称加密算法中公钥和私钥是一一对应的关系，即每一个公钥都有与之一一对应的私钥，反之亦然。并且所有的公钥私钥对都是不同的。使用公钥加密的信息可以被私钥解密，同时使用公钥也可以解密使用私钥加密的信息。可使用私钥生成公钥，但使用公钥推算与之对应私钥是基本不可能实现的[35]。

2.3 群签名

2.3.1 数字签名

信息安全主要解决保密性、完整性、有效性三大问题[36]。即保障信息在传输过程中不被泄露，信息在传输过程中不被篡改和验证信息使用者的合法性。加密算法用来

解决信息安全中的保密性问题，而数字签名技术解决信息安全中完整性和有效性问题。数字签名是指可以添加到文件的电子安全标记，一个数字签名通常定义了签名和验证两种互补运算，是非对称加密算法的逆应用[37]。

现实生活中，传统手写签名表示了纸上内容由签名者书写，数字签名与手写签名含义相同，在计算机世界中证明消息是由特定消息发送者发送，用来验证消息发送者身份。另外，数字签名还能证明传输过程中消息没有被其他任何人篡改，证明消息的完整性[38]。

数字签名和手写签名相比，虽然含义相同，但是如表 2.1 所示，其实有着很大的不同。

表2.1 数字签名与手写签名对比

	数字签名	**手写签名**
签名和信息是否隔离	是	否
签名验证方法	任何拥有签名者公钥的参与方都可以验证签名	比对签名样本进行验证
复制后签名有效性	有效签名的复制仍然是有效的签名	签名的复制是无效的
是否抗抵赖性	是	否

数字签名与手写签名相比具有很大优势：数字签名同时具有确认信息来源和保证信息完整性这两个功能，也就是说，数字签名更偏向于对消息整体上的验证。特别要注意地是，数字签名可以保证抗抵赖性。即信息发送者在发送信息后不能对发送信息行为做出抵赖。信息接收者收到信息后，可根据数字签名，通过第三方确认签名者身份。在双方就信息的发送者和信息内容有争议的时候，数字签名可以用来解决信息发送者和信息接收者之间的争端。

数字签名的使用条件[39]：

（1）信息接受者具备验证信息发送者签名的条件。

（2）信息发送者发送信息并签名后不能对抵赖签名行为。

（3）信息接收者只能验证签名，不能伪造签名。

如图 2.1 所示，数字签名具体过程为：

（1）甲作为信息发送方，乙作为信息接收方。甲乙各拥有公钥和私钥。

（2）甲乙作为通信两方。甲秘密保存自己私钥，已知乙的公钥，同时乙秘密保存自己私钥，已知甲的公钥。

（3）甲使用乙公钥对加密信息明文，生成密文消息。此时，甲可以发送密文给乙，在乙私钥没有泄露的前提下，只有乙可以解密密文，甲和乙之间安全发送信息。但是此时，乙无法确认信息发送方为甲。

（4）甲对信息明文使用 HASH 算法生成数字摘要 A。

（5）甲使用自己的私钥对数字摘要 A 进行加密，加密后的数字摘要即为数字签名。

（6）甲将密文信息和数字签名一起发送给乙。

（7）乙使用甲的公钥对数字签名进行解密，得到数字摘要，验证甲的身份。

（8）乙使用自己的私钥对密文信息进行解密得到明文。

（9）乙使用 HASH 算法对上述步骤得到的明文进行计算，得到数字摘要 B，将数字摘要 A 与数字摘要 B 进行对比，若结果一致，则解密后明文为要接收信息，否则认为，甲发送信息错误或者在发送过程中信息被恶意篡改。

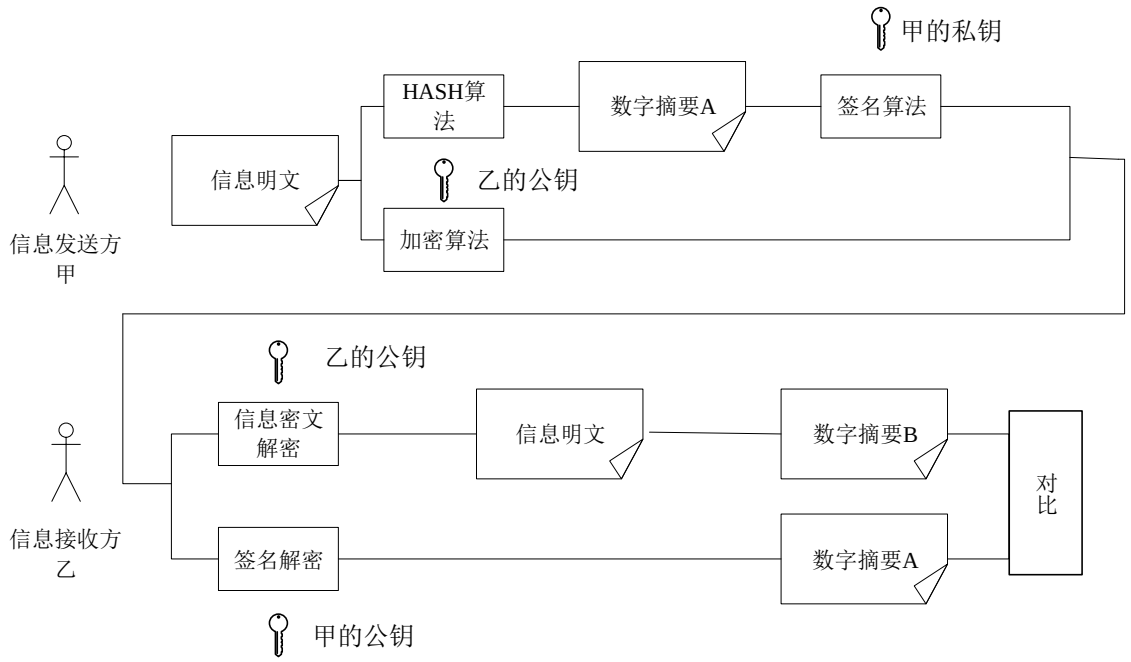

图2.1　数字签名过程

使用数字签名用来保证信息的真实性和完整性。如果没有使用有效的数字签名，那么无法保证接收方接收的信息是否真的来自发送方，也无法保证信息传输过程中是否被恶意篡改。利用数字签名和验签技术保证信息的抗抵赖，同时也起着身份识别的作用。

2.3.2 群签名

群签名属于数字签名，由 Chaum 和 van Heyst 提出，随后 Camenish、Stadler 等进行了修改和完善[40]。群签名使得群组内任意成员代表群组进行消息签名，所有知道群组公钥的人都可以验证群签名的正确性，但是又能保证群组内签名的具体成员身份不被验证者知晓，保证群组签名成员的匿名性[41]。

群签名的特点[42]：

（1）群体特性：只有群组内成员可以对消息签名。

（2）验证简单：所有知道群组公钥的人都可以验证签名的正确性。

（3）匿名性：信息的接收者可以验证签名的有效性，但是只能验证签名是否来自包含信息发送者的群组，不能判断群组中签名的具体成员。

（4）可追查性：具体签名者身份可由群组管理员有效追踪确认。

群签名的过程：

（1）群组初始化：首先要有群管理员，群管理员是主要操作者。

（2）成员入群：群成员申请入群，群管理员根据交互协议颁发证书给群成员。

（3）生成群签名：根据概率算法生成群签名，算法的输入为发送消息和成员私钥。算法的输出为生成的群签名。

（4）验证签名：根据算法对签名进行验证，算法输入为签名消息和群公钥。算法输出为签名是否来自某特定群组。

（5）打开群组：打开群组由群管理员操作，依据算法确定签名成员的具体身份，算法的输入为生成的群签名和群私钥。算法的输出为签名由群组内某特定成员签署的事实。

2.4 区块链

区块链技术可以实现从单系统内自理扩展到多个结构的完美合作。区块链最早在中本聪撰写的《Bitcoin:A Peer-to Peer Electronic Cash System》中提出，文中指出区块链是按时间顺序记录比特币交易账目历史的数据结构[43]。区块链技术从 2015 年开始真正繁荣。到目前为止，区块链技术经历了三次迭代：从比特币为代表的货币交易网络到以以太坊为代表的合约区块链技术，然后再到具有权限控制的 Hyperledger 项目。Hyperledger 主要由分布式共享账本，智能合约，相关密码学技术和共识机制四大部分组成，其中证书默认签名算法是 ECDSA，HASH 算法默认 SHA256。Fabric 是 HyperLedger 中的一个针对区块链技术实现的一个子项目[44]。

（1）区块链的特点

区块链并不是凭空诞生的新技术，而是各种技术发展到一定程度后相互交错的结

果[45]。从技术角度出发，区块链涉及分布式系统，数据存储，密码学知识，心理学，网络协议和博弈论等领域。但是，区块链主要拥有三个关键点。第一，区块链网络中所有数据被存储在人人都可以访问的、公开的、不可变的交易账本中。第二，区块链网络中交易数据不能删除或更改，所以交易数据是完整的并且不可否认的。第三，区块链是由网络中计算节点组成的去中心化网络。网络中所有数据在每个节点上都是完整存在的，所以区块链网络具备很强的鲁棒性。

（2）区块链的分类

区块链根据不同的分类标准有不同的分类。根据网络覆盖范围划分为公有链，私有链和联盟链[46]。公有链是一个任何结点都可以随意加入的公共区块链网络，网络中每一个节点都可以访问账本信息；私有链是网络中所有节点共有同一拥有者的区块链网络；联盟链允许授权的结点加入网络，具有身份认证和权限访问等功能。根据部署环境的差异，可以把区块链分为主链和测试链。主链由功能完备的正式客户端节点组成，测试链主要用于学习开发测试的区块链网络。根据对接类型，区块链又可以分为：单链和互联链以及侧链。普通独立运行的区块链网络一般是单链。侧链与单链类似，有自己的账本、共识机制、交易类型和智能合约的独立区块链系统。互联链由可以发行比特币的比特币系统为主链和不能发行比特币的侧链双向挂钩形成，当侧链流通属于主链比特币时，主链相对应比特币被锁定。

（3）区块链中常见隐私保护技术

区块链技术在互不信任的各方之间架起了信任的桥梁，具有时序性的分布式账本可以保证区块链网络中参与方的安全[47]。在比特币交易系统中，交易双方使用和真实身份无关的公钥哈希值作为自己交易的地址，这在一定程度上保护了用户的隐私，也就是所谓的区块链网络具备匿名性，虽然用户真实身份与交易地址之间无关联，但是，公钥哈希值与真实身份一一对应，分析多个交易过程，发现交易之间存在着特定的关系。实际上，区块链网络中参与方是化名存在的。在区块链网络中，参与方拥有与自己真实身份无关的身份标识，但是身份标识代表的虚拟身份在网络中所做的任何事“人人皆知”，区块链技术并不如广为流传所说可以提供强大的隐私保护机制[48]。

随着区块链技术应用市场的日益壮大，围绕区块链技术隐私保护问题的研究讨论从未止步。区块链涉及的隐私不只包括输入方和输出方的隐私，还应该包含所有涉及方的隐私，所以在保障隐私的同时还要考虑其他所有的因素[49]。由 R3 联盟发布的区块链交易隐私保护机制研究报告可知，交易隐私保护方案大致可分为，经许可和审核的方式，去链方式和身份混淆，零知识证明和隐形地址[50]。

经过许可和审核指的是区块链网络的参与方是经过深入资格审查后才能加入网络。组织中使用联盟链或者私有链就是使用经许可和审核的方式进行隐私保护。

隐形地址是指区块链网络中，交易双方进行交易时，交易发起者不需要提前已知

交易接受者地址，而是发起者创建新地址并且发起交易，使用某种技术保证接受者可以打开交易。这种方式可以实现交易双方交易前不必传播其他信息，就可以确定交易接收者的地址。

零知识证明在 Zcash 中有着很好的应用，在不知道数据具体内容的情况下不借助其他知识证明数据是合法有用的，但是零知识证明也有一些缺点。例如，计算速度缓慢，在交易吞吐量大的场景中不能发挥作用。

2.5 本章小结

本章对文中涉及的相关技术进行了简单的介绍。首先介绍了数字医疗和医疗数据标准，为医疗数据的高效共享提供指导，然后介绍了加密算法，数字签名和群签名。为接下来隐私保护机制的研究提供理论基础，最后概要讲述了文中使用的区块链技术。

第三章　基于默克尔树加密的医疗数据隐私保护研究

数据隐私保护手段主要有两种：一种是构建可信的执行环境，另一种是使用密码学手段[51]。对医疗数据使用加密的方法存储和发送是保护数据隐私有效且更常用的方法，加密数据借助区块链网络在共享双方之间流通。但是由于区块链网络中区块容量有限，所以需要对原始医疗数据进行预处理，同时为了提高医疗数据共享的效率和保障数据需求者安全访问数据，在基于区块链的医疗数据共享过程进行隐私保护的研究十分有必要。

3.1 总体思路

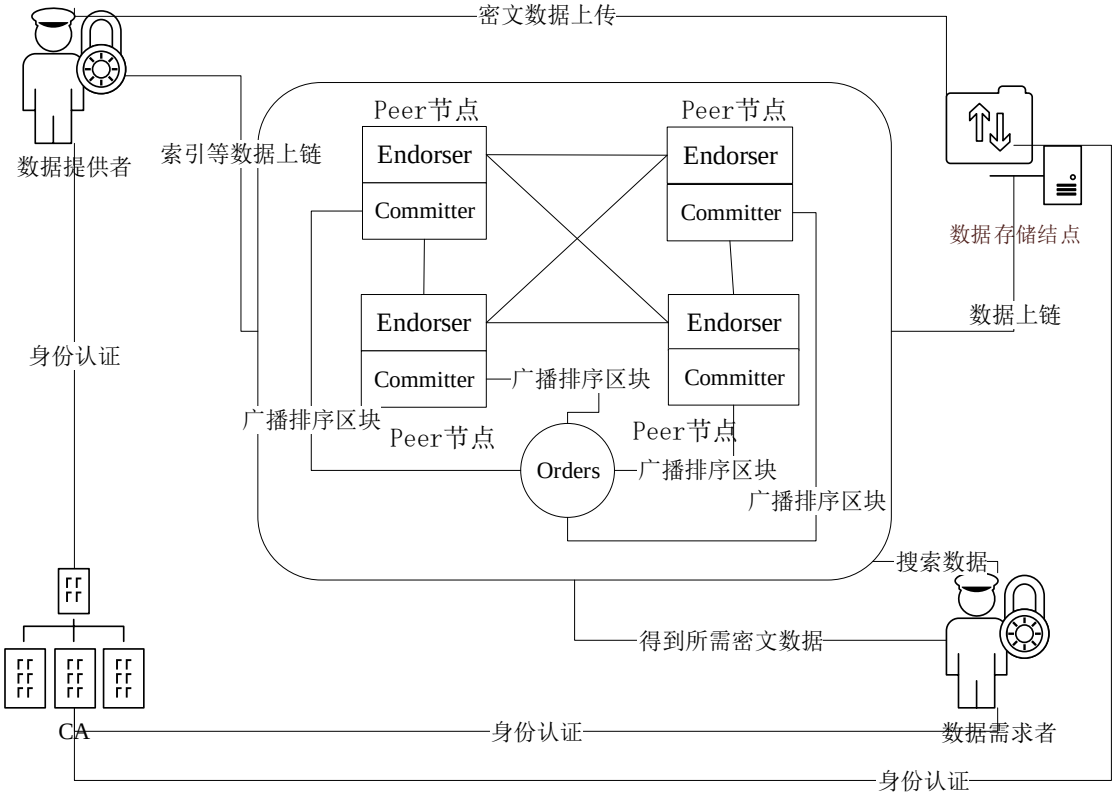

图3.1 医疗数据隐私保护研究过程

如图 3.1 所示，本章从应用层的角度出发进行医疗数据隐私保护研究。首先分析医疗数据，发现原始医疗图像数据所需存储空间较大，而区块链中区块容量有限，因此为了高效地共享数据，对数据进行标准化处理，生成结构化的医疗数据。其次，为了提高医疗数据的共享效率，对预处理之后数据进行关键字提取，构建医疗数据的关键字索引结构。同时将医疗数据及其关键字索引结构存储在区块链上，形成医疗数据

的默克尔树，保证医疗数据的完整性。最后，为保障数据需求者对数据的安全访问和数据提供者对医疗数据的完全控制，研究并实现医疗数据加密共享算法。共享过程分为两个阶段：第一阶段为数据提供者上传密文数据，第二阶段为数据需求者下载数据。医疗数据提供者上传标准化数据及索引和数据需求者下载所需医疗数据的流程图分别如图 3.2 和图 3.3 所示。

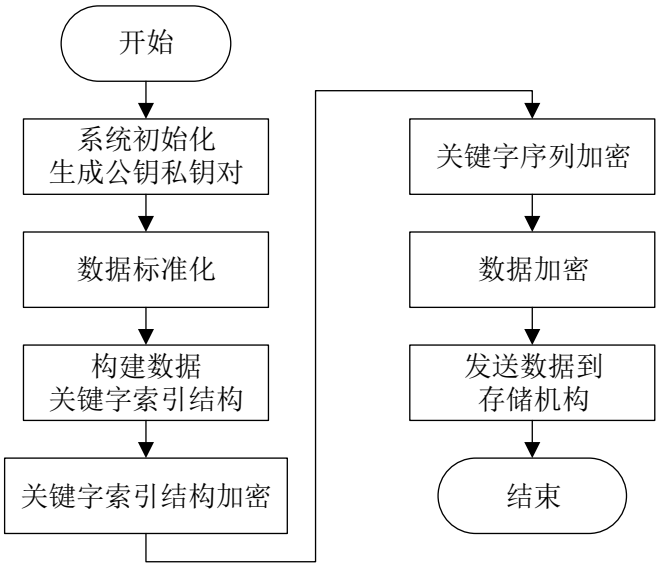

图3.2 医疗数据上传流程图

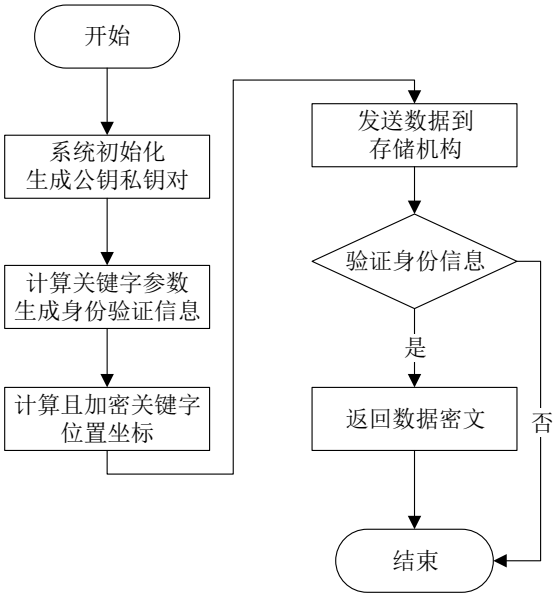

图3.3 医疗数据下载流程图

3.2 医疗数据标准化

医学科技创新研究依赖于大量标准化和高可信的医疗数据，同时富有价值的医学数据还会引发疾病治疗方式的变革。文中所用医疗数据来自不同数据源，由于医学数据在开发利用、开放共享和安全保护等方面存在明显不足，并且医疗数据本身结构复杂[52]，所以为使数据共享双方对数据有共同的理解，文中使用结构化形式对医疗数据进行描述，实现安全高效地数据共享，充分地共享和利用医疗信息的价值。

3.2.1 医疗数据预处理

医疗数据不仅仅包含表格形式的统计数据，还包括大量医疗图像数据。对于医疗图像数据，按照公共数据库 VOC 数据集格式对图像数据做处理，提取关键信息，生成可供研究使用的包含关键信息的 XML 格式。对于表格形式的统计数据，依据表格属性及其属性值生成格式化数据。医疗数据总体预处理流程如图 3.4 所示。

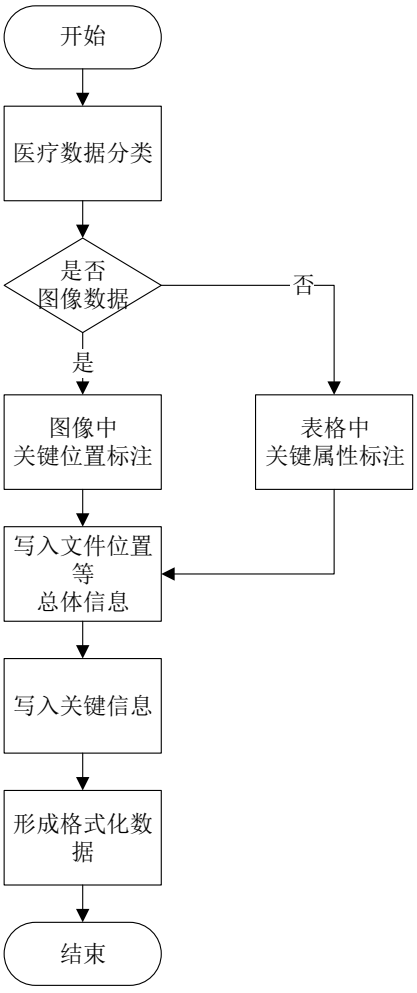

图3.4 医疗数据预处理流程图

医疗数据预处理流程图中数据标准化具体操作如图 3.5 所示。

函数名称：toXML

输入：分类框起始x坐标元组xmin_tuple，分类框结束x坐标元组xmax_tuple，分类框起始y坐标元组ymin_tuple，分类框结束y坐标元组ymin_tuple，图像名称image_name，文件所在包的名字docname

输出：生成满足需求的XML格式的数据

步骤：

```
1. node_root←Element('annotation')                    /*生成xml文件的根元素annotation*/
2. node_folder←SubElement(node_root,'folder')    /*为根元素annotation生成子元素folder*/
3. node_folder.text ← docname                          /*为子元素folder赋值*/
4. node_file←SubElement(node_root,'filename')    /*为根元素annotation生成子元素filename*/
5. node_file.text← filename                            /*为子元素filename赋值*/
6. node_path←SubElement(node_root,'path')        /*为根元素annotation生成子元素path*/
7. node_path.text← path                                /*为子元素path赋值*/
8. size()                                                  /*填充元素图像数据*/
9. for i form 0 to len(xmin_tuple)
10.   node_object←SubElement(node_root,'object')
11.   node_name←SubElement(node_object,'name')
12.   node_name.text ←Classname
13.   node_pose←SubElement(node_object,'pose')
14.   node_pose.text← 'Unspecified'
15.   node_trun←SubElement(node_object,'truncated')
16.   node_ trun.text←0
17.   node_diff←SubElement(node_object,'difficult')
18.   node_diff.text ←0
19.   node_bndbox ←SubElement(node_object, 'bndbox')
20.   node_xmin ←SubElement(node_bndbox, 'xmin')
21.   node_xmin.text← str(xmin_tuple[i])
22.   node_ymin ←SubElement(node_bndbox, 'ymin')
23.   node_ymin.text← str(ymin_tuple[i])
24.   node_xmax ←SubElement(node_bndbox, 'xmax')
25.   node_xmax.text ← str(xmax_tuple[i])
26.   node_ymax ← SubElement(node_bndbox, 'ymax')
27.   node_ymax.text ←str(ymax_tuple[i])
28.end for
                                                           /*填充object的子元素*/
29.xmltostring(node_root,pretty_print=True)       /*生成xml数据*/
30.dom←parseString(xml)
31. return dom                                         /*返回格式良好的xml数据*/
```

图3.5 医疗数据预处理图

结构化数据中包含原医疗数据的存储位置，文件大小，文件名称等概述信息，也包含医疗图像中关键信息和医疗文本数据的关键属性信息等。医疗数据提供者提供结构化数据，数据接收者可根据结构化数据从总体上了解医疗数据，同时原始医疗数据与共享通信数据隔离，起到了隐私保护的作用。

分析研究结构化医疗数据，对数据集中每一个标准化、结构化数据进行关键字提取，形成关键字集合，为下一小节做准备。

3.3 医疗数据的默克尔树

3.3.1 构建关键字索引结构

现有

（1）标准化、结构化医疗数据集合 x，n 为数据集大小。

（2）包含 m 个关键字的集合 k，m 为标准化医疗数据集合中互不相同的关键字总数。

（3）哈希函数 H：关键字与其位置坐标的映射关系。

（4）随机生成器 R：01 随机生成函数。

利用上述数据集和函数来构建关键字索引的默克尔树结构。首先构建具有以下特点的关键字索引树结构：

假设树中所有叶子节点数量为 n，则关键字索引树中所有节点数为 2n-1。每一个叶子节点与可唯一识别标准化医疗数据的标识符一一对应。

树结构中每一个叶节点存储的是 key-value 键值对，key 存储长度为 m 的 01 比特串，value 值是唯一识别标准化医疗数据的标识符。叶子节点中比特序列与关键字集合中关键字一一对应。对于每一个叶子节点来说，比特序列中某一比特值为 1，则说明标识符对应的数据中包含关键字集合中相应位置上的关键字。

树中所有内部节点由该节点的孩子节点运算生成，同样存储包含 m 个元素的 01 比特序列。整个关键字索引树结构从叶子节点到树根节点，由下而上，逐层形成。关键字索引树具体构造过程为：内部节点的比特序列由该节点的左孩子比特序列与右孩子比特序列做或运算得到。对于树中内部节点而言，若比特序列中某一位置值为 1，则该节点的左孩子节点与右孩子节点中至少有一个节点的该比特位值为 1。也就是说，从根节点开始搜索，经过该内部节点的搜索路径至少有一条可达叶子节点，找出所需医疗数据。

按照上述过程构建关键字索引结构并查找包含关键字k_i（i 为大于 0 小于 m 的整数）的医疗数据，若构建医疗数据关键字索引结构时，根据哈希函数 H 计算关键字k_i在叶子结点的 01 比特序列中的位置为从左往右的第二位，则查找过程如图 3.6 所示。

（1）从根节点开始，查看根节点的比特序列中k_i相对应位置上的值，如果该值为零，则说明现有医疗数据集中完全不存在关键字k_i，结束查询。否则，继续第二步。

（2）此时说明根节点的 01 比特序列中，关键字k_i相应位置上值为 1，由于该节点的比特序列值由该节点的左孩子节点和右孩子节点的 01 比特序列值做或运算得到，所以该节点的左孩子节点和右孩子节点并列检查搜索。

（3）若节点的比特序列中，该关键字k_i相对应位置上的值为 1，且该节点为叶子节点，则返回标识符所对应的医疗数据。如果该节点不是叶子节点，则按照第二步继续检查搜索。若节点 01 比特序列中，关键字k_i相对应位置上的值为 0，则该路径上结束检查。

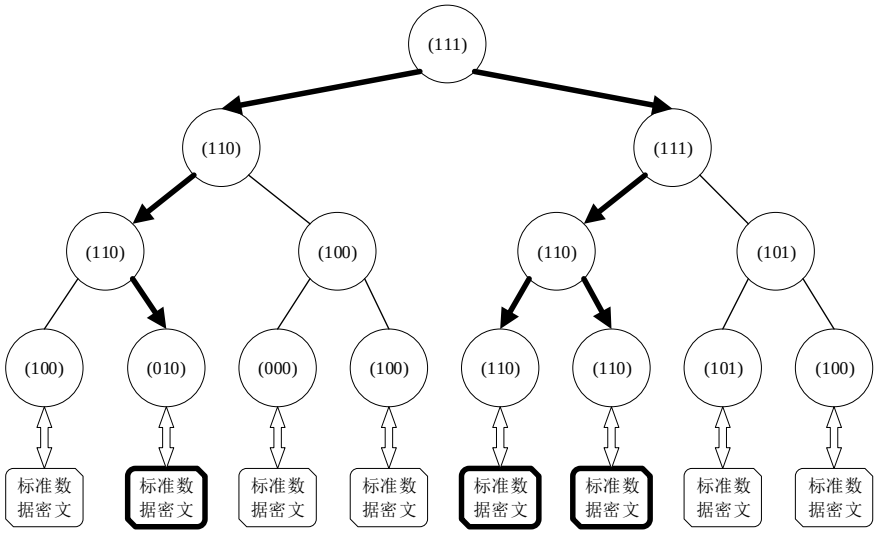

图3.6 关键字索引结构图

3.3.2 构建医疗数据的默克尔树

以标准化处理的包含 normal 和 lens 以及 dense 关键字的小儿白内障数据集为例，进行关键字集合抽取，进一步构建医疗数据的关键字索引结构。

（1）构建医疗数据的关键字结构

关键字集合由 normal 和 lens 以及 dense 构成，则索引树结构中所有节点的比特序列长度为 3。根据哈希函数 H 计算 normal 关键字在比特序列中位置为 2，lens 关键字在比特序列中位置为 1，dense 关键字在比特序列中位置为 3。按照上一小节处理数据集中每一个结构化数据，最终生成关键字索引结构如图 3.7 所示。

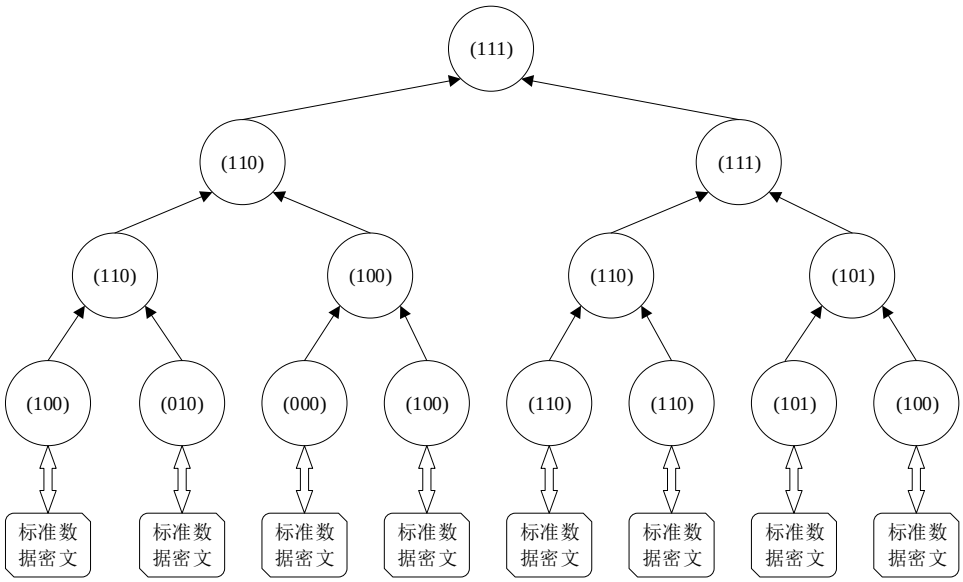

图3.7 医疗数据的关键字索引结构图

（2）构建医疗数据的默克尔树结构

默克尔树是由一个根节点，一组叶子节点和一些中间节点组成的存储哈希值的树结构。叶子节点中存储数据或者数据集的哈希值，中间节点由该节点的左右孩子节点生成，先把左孩子节点哈希值与右孩子节点哈希值合并成一个字符串，然后对该字符串再进行哈希计算得到，从下往上，逐层计算[53]。

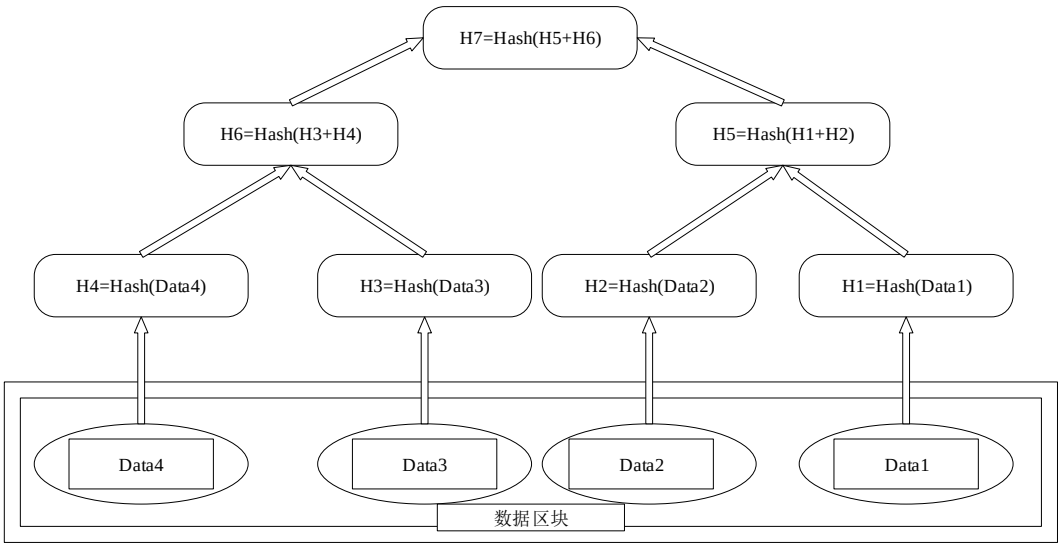

图3.8 默克尔树结构图

本文研究基于区块链的医疗数据共享，但是医疗数据包含医疗图像数据，区块链中区块块容量有限，原始医疗数据所需存储空间较大，所以文中区块链不存储原始医

疗数据，存储的是与原始数据对应的标准化数据。依据图 3.8 构建医疗数据的默克尔树，具有以下特点：医疗数据的默克尔树结构是二叉树，满足树结构的所有特点；默克尔树中叶子节点相对应的医疗数据是经过加密处理的，依据 3.2 小节进行标准化处理的结构化数据。医疗标准化数据与原始医疗数据一一对应；医疗数据的默克尔树从下往上，逐层计算。

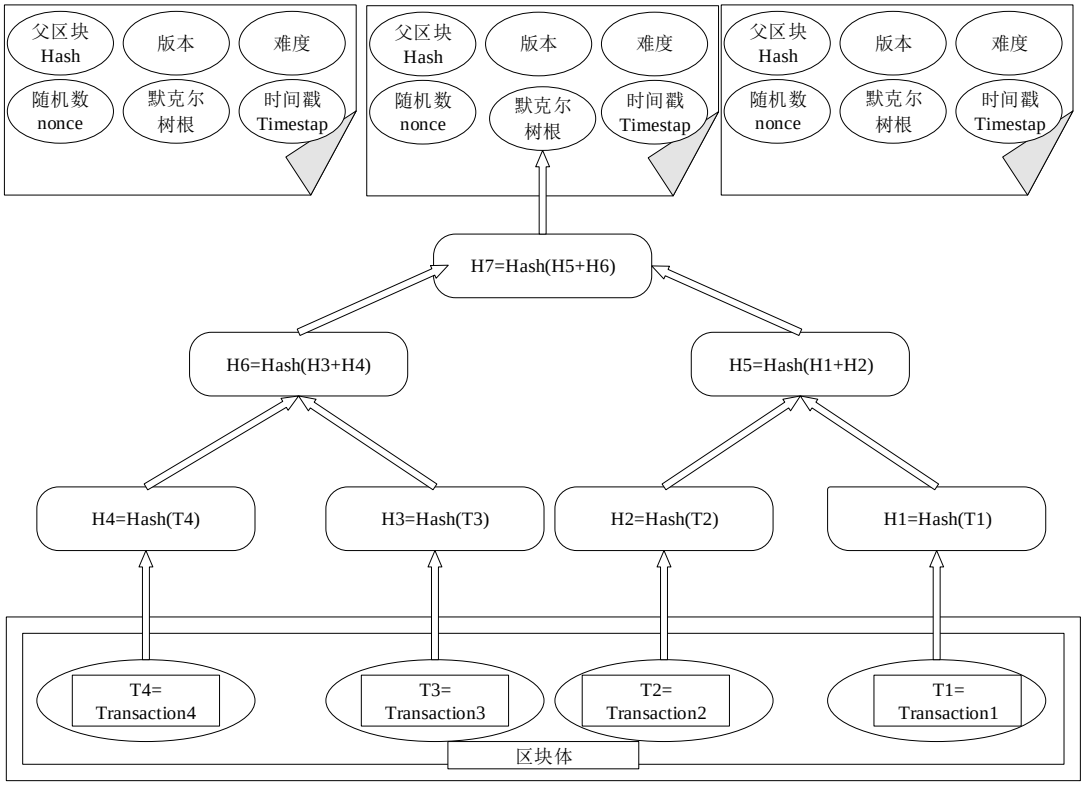

图3.9 比特币网络中的默克尔树

如图 3.9 所示，比特币网络中使用默克尔树存储每一个区块中所有交易数据[54]。一方面便于快速验证某一笔交易的存在，另外一方面便于比特币网络中轻客户端的使用，若使用轻客户端则只需下载区块链中区块头，不用下载完整的区块信息和交易信息[55]。文中为了保障医疗数据的完整性和不可篡改性，数据提供者将标准化数据密文，加密关键字索引结构存储在区块链上，每一个区块都是一组有序交易集合，通过 hash 值链接到父区块，最终在区块链上的具体存储形式如图 3.10 所示。

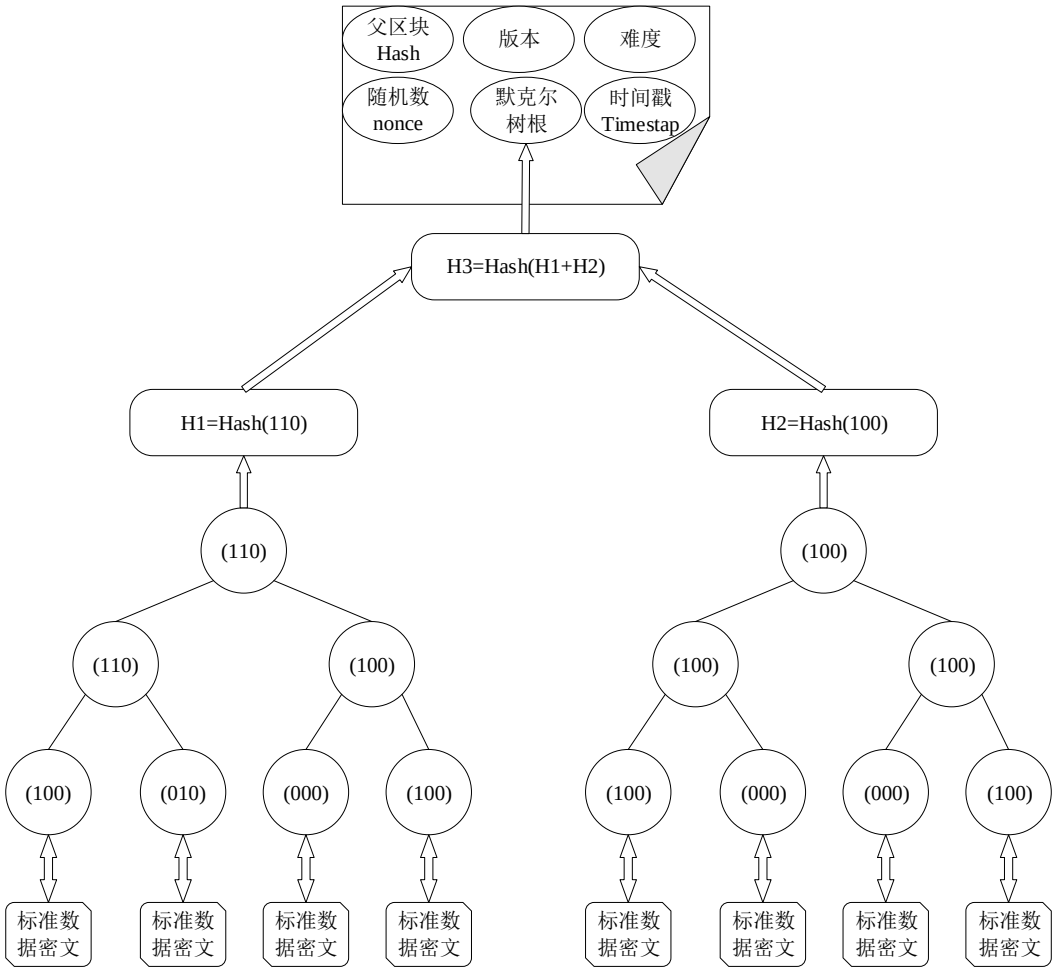

图3.10 区块链中医疗数据示意图

3.4 医疗数据的加密算法

在上一小节给出了基于医疗数据的可查找关键字索引结构的默克尔树，在本节研究实现了针对医疗数据的加密共享算法，使得既保证数据提供者对数据的完全控制，又保证数据需求者对数据的安全访问。

3.4.1 算法研究与实现

如图 3.11 所示进行医疗数据共享，具体过程如下：

（1）密钥生成

首先初始化 RSA 加密算法，随机选择两个不相等的质数p_1和质数 q_1（质数p_1和质数 q_1的选择越大越好），计算n_1=$p_1 * q_1$（n_1二进制长度为密钥的长度），计算n_1的欧拉公式$\varphi(n_1)$=$((p_1\text{-}1)(q_1-1))$。随机选择一个满足条件大于 1 且小于$\varphi(n_1)$的整数e_1,且所选整数e_1与$\varphi(n_1)$互质，计算e_1对$\varphi(n_1)$的模反元素d_1。则数据提供者的公钥为（n_1，e_1），私钥为（n_1，d_1）。同理可计算出数据存储节点公钥的（n_2，e_2），私钥

（n_2，d_2），数据接受者公钥（n_3，e_3），私钥（n_3，d_3）。

同时系统中公开，哈希函数H_1:{0，1}^{y} × {k_1, … , k_m} → {1, … , m}，随机生成器R:{0，1}$^{\log m}$ × {0，1}* → {0,1}。m 为文本集中所有关键字的个数。其中，{0，1}^{y}为使用随机函数产生长度为 y 的 01 比特串。

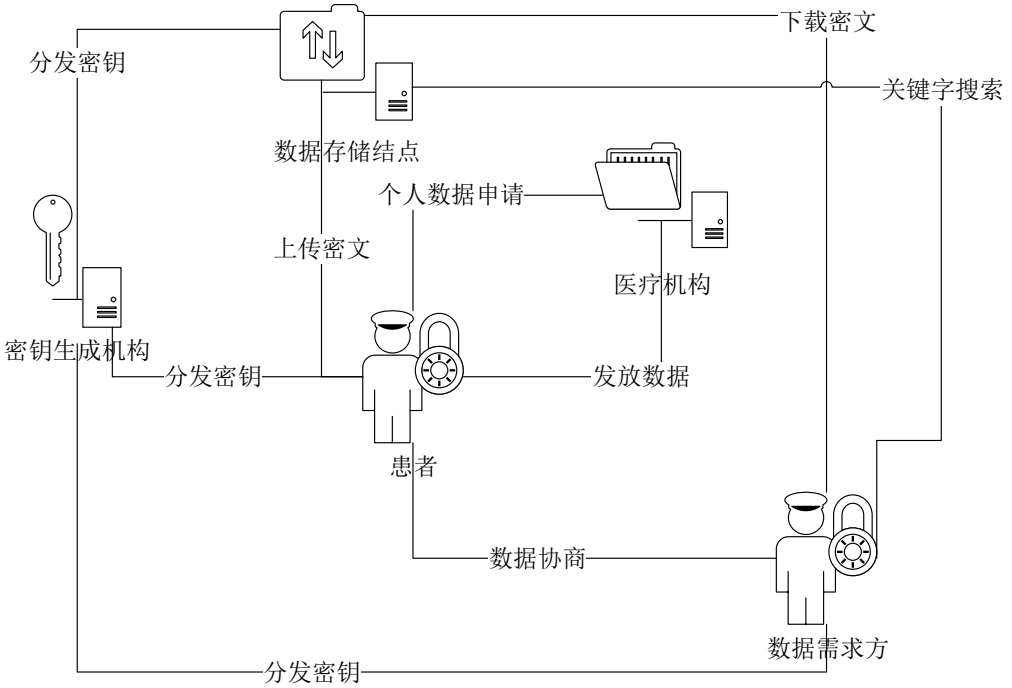

图3.11 医疗数据共享示意图

（2）加密计算

使用步骤（1）生成密钥进行加密计算。数据需求者向数据提供者申请数据，经过协商，数据提供者已知数据需求者所需数据格式和数据内容。假设数据需求者申请包含关键字k_i（i是大于 0 小于等于m的自然数）的医疗文本数据，首先对原始医疗数据进行标准化处理，得到标准化数据集 x。原始医疗数据另行存储，标准化数据集中包含原始医疗数据的存储路径等信息。使用数据需求者公钥对标准化数据集加密得到密文。如果同一数据需要发送给不同数据需求者，因需求者公钥不同，虽然数据明文一致，但此时要看作不同数据，分别计算得到不同密文。

这里以标准化数据集 x 中所有数据都发送给同一数据需求者为例，使用式（3-1）计算得到密文序列 X（i 是大于 0 小于等于 n 的自然数）。

$$X_i = {x_i}^{e_3} \ mod \ n_3 \tag{3-1}$$

（3）数据提供者构建关键字索引

数据提供者根据关键字索引生成算法，对标准化数据集中每个医疗数据进行处理，

构建关键字子索引结构。具体的算法如图 3.12 所示。

函数名称：KeyIndex	
输入：文档序列 x，关键字序列 k，文档序列总数 n，关键字序列总数 m。	
输出：关键字子索引树 T。	
步骤：	
1.T←BuildTree(n)	/*生成叶子节点数为 n，总个数为 2n-1 的完全二叉树*/
2.for i from 1 to n	/*初始化树的叶子节点*/
3. f_i←generate(i)	/*生成并存储每个文档的标识符 f*/
4. for j from 1 to m	
5. v←init(v_j)	/*初始化数组 v=($v_1, v_2, v_3 \dots v_m$)，v 的长度为关键字个数 m*/
	/*v_j ∈ {0,1},0<j<n，标识符 f 和数组 v 一一对应*/
6. for t from 1 to m	/*对于每一个关键字，计算每一个叶子节点*/
7. if k_t in f_i	
8. v_j←1	
9. else	
10. v_j←0	
11. end for	
12. end for	
13.end for	
14.for i from n-1 to 1	/*根据叶子节点计算树的内部节点*/
15. v_i←v_i.left ∩ v_i.right	/*该节点左孩子节点与右孩子节点或运算*/
16.end for	
17.return T	/*输出子索引结构*/

图3.12 关键字索引生成算法

数据提供者在构建关键字索引结构过程中，计算关键字在索引树结构节点中位置时，对于每一个关键字，使用哈希函数H_1计算关键字在 01 比特序列中的坐标。那么，关键字在节点存储的 01 比特序列中的坐标位置具有随机性，所以算法输入关键字序列与构建子索引树节点存储的 01 比特序列对应的关键字序列很大程度上不一致，则此时选择数据存储节点公钥按式（3-2）对数据提供者构建索引结构的关键字序列进行加密处理，并且发送到存储节点。但是索引结构树中每一个节点，计算关键字位置时都使用同一哈希函数H_1。相同函数，同一关键字输入，输出的关键字位置坐标一定相同。

$$K = k^{e_2} \ mod \ n_2 \tag{3-2}$$

选择无碰撞哈希函数H_2并计算关键字k_i的哈希值$H_2(k_i)$，同时选取非零自然数 s 计算

$$S = s^{H_2(k_i)} \ mod \ n_3 \tag{3-3}$$

（4）关键字子索引结构加密

数据提供者对子关键字索引结构进行加密处理，具体算法描述如图 3.13 所示。

函数名称：EncKeyIndex	
输入：用户本地子索引结构 T，关键字数量 y，m 与 y 一致，T 中包含比特串 v	
输出：加密后的子索引结构T'。	
步骤：	
1.for i from 1 to 2n-1	
2. Init(v')	/*初始化数组v',v'的长度与关键字数量相同*/
3. for j from 1 to m	
4. index←$\{0，1\}^y \times k_j$	/*计算关键字在数组中的坐标*/
5. $I_j \leftarrow \{0，1\}^{\log m} \times \{0，1\}^* \rightarrow \{0,1\}$	/*使用随机函数 R 生成该坐标的 01 值*/
6. $v_j' \leftarrow I_j$异或 v_j	
7. end for	
8.end for	/*对于树中每一个结点都做上述加密处理*/
9.return T'	/*输出加密后的索引结构*/

图3.13 关键字索引加密算法

数据提供者将标准化数据密文 X 和关键字序列密文K，关键字k_i计算结果S，加密子关键字索引发送给数据存储结点。

（5）数据存储结点生成全网关键字索引的默克尔树

数据存储结点在共享过程中担任共享数据存储和数据管理的任务。在医疗数据共享过程中，一方面接受数据提供者的存储请求，另一方面应答数据需求者的搜索并获取数据请求。基于区块链的医疗数据共享网络中，任意参与数据共享过程的数据提供者都先将标准化数据密文，关键字索引结构密文等，发送给数据存储结构存储，数据需求者通过区块链网络输入特定参数，调用链码，获取所需数据。

假设数据存储结点收到 z 个子索引结构，先使用自己私钥对关键字序列解密，然后运行子索引合并算法，构建所有共享数据的关键字索引结构。因为所有数据源用户使用同一个哈希函数H_1，所以各个数据源子关键字索引结构相同。

（6）数据需求者提交搜索申请。

按照数据提供者和数据需求者协议，数据需求者需要包含关键字k_i的医疗数据。则数据需求者先使用哈希函数H_1计算该关键字k_i坐标位置 index。根据哈希函数的特点，输入关键字与输出坐标一一对应，则得到关键字在索引树叶子节点的比特序列中相应位置，最后使用数据存储结点公钥对 index 进行加密。

$$Index = index^{e_2} \ mod \ n_2 \tag{3-4}$$

同时计算 A

$$A = H_2(k_i)^{-1} \ mod \ \varphi(n_3) \tag{3-5}$$

随机选取整数$\mathrm{r} \in Z_n^*$，计算B和D。

$$B = r \cdot A \ mod \ \varphi(n_3) \tag{3-6}$$

$$D = s^{\mathrm{r}} \ mod \ n_3 \tag{3-7}$$

发送计算结果 B，D和关键字密文坐标 Index 到数据存储结点。

（7）执行搜索过程。

函数名称：Search
输入：关键字 w 及位置 index，全局数据关键字的索引树 A，A 中叶子节点的个数 n，A 中包含密文 f 和比特串 v。 输出：包含关键字的密文集合 C 步骤： 1.C← ∅ /*初始化结果密文集合 C 为空集*/ 2. for i from 1 to $\log_2 n$+1 /*从根节点开始搜索关键字*/ 3. r← H(w) /*恢复关键字 w 在该节点生成的随机数 r*/ 4. temp←r 异或 index(w) /*随机数 r 与关键字 w 对应位置上值做异或运算*/ 5. if temp is 0 6. end /*该搜索路径结束*/ 7. else if temp is 1 and isTree(i) 8. C← f_i /*该叶子节点对应的密文f_i加入密文集合 C*/ 9. else if temp is 1 and !isTree(i) 10. Search(v_i.left) and Search(v_i.right) /*并列搜索该节点的孩子节点*/ 11.end for 12.return C /*返回搜索结果 C*/

图3.14 搜索关键字算法

数据存储结点首先计算等式$D = S^B$否成立，如果成立，则说明该获取数据请求来

源于特定数据需求者。其次使用自己私钥对密文$Index$解密，得到待搜索关键字在树节点的比特序列中位置$index$。然后从根节点开始，从上到下搜索关键字索引树结构，检查树中每一个节点。当非叶子节点的内部节点包含关键字的时候，左孩子节点和右孩子节点同时执行搜索过程，直至叶子节点，返回密文数据，否则返回空集。

（8）取消数据需求者对数据的访问权限

若数据提供者想取消某一数据需求者共享数据访问权限，则可发送消息给数据存储结点，数据存储节点按照要求，将关键字索引树中该需求者所需密文对应的叶子节点比特序列位全置为 0，再从下到上重新计算包含该叶子节点路径上的关键字索引结构，同时更新账本。这样当数据需求者发送获取数据请求，返回密文集合为空，如此取消数据需求者对共享数据访问权限。

3.4.2 加密算法分析

（1）正确性分析

数据提供者计算：

$$S = s^{H_2(k_i)} \tag{3-8}$$

使用数据需求者的公钥对数据加密：

$$X_i = x_i{}^{e_3} \tag{3-9}$$

数据需求者计算：

$$A = H_2(f_i)^{-1} \tag{3-10}$$

$$B = r \cdot A \tag{3-11}$$

$$D = s^{\mathrm{r}} \tag{3-12}$$

数据需求者使用自己私钥解密密文：

$$x_i = X_i{}^{d_3} \tag{3-13}$$

数据存储节点计算：

$$S^B = s^{H_2(k_i)\mathrm{r}H_2(f_i)^{-1}} \tag{3-14}$$

如果说k_i与f_i是同一个关键字，那么$S^B = s^{\mathrm{r}} = D$；如果k_i与f_i不是同一个关键字，那么$s^{H_2(k_i)\mathrm{r}H_2(f_i)^{-1}}$与$s^{\mathrm{r}}$不相等。

（2）隐私保密分析

文中将原始医疗数据保密存储，数据提供者与需求者之间共享标准化数据，在标准化数据中包含原始数据的存储位置等关键信息，这样原始医疗数据与数据双方之间的通信数据分离。

数据提供者使用数据需求者公钥，基于 RSA 算法对医疗标准化数据进行加密处

理，在数据需求者私钥没有泄露的前提下，任何没有私钥的参与者想要解密密文是非常困难的。并且，数据存储节点只是保存和管理数据，并不知道医疗数据具体内容，文中提供了密文搜索功能，数据存储节点在执行搜索过程中不能得到数据提供者的任何信息，这样在很大的程度上保障了数据隐私安全。更重要的是，在数据提供者和数据需求者结束共享过程后，数据提供者可以取消数据需求者对数据的访问权限，从访问控制的角度保护了数据隐私。

（3）安全性分析

数据提供者生成子关键字索引的过程中，使用了哈希函数和随机生成器，每次数据共享过程中，子关键索引结构不完全一致，即使同一数据共享双方，在一次共享过程中不慎泄露关键字序列，对下一次共享过程没有任何威胁。同时关键字索引加密过程中，每一个节点比特序列中每一比特位都由随机生成器 R 产生，叶子节点每位比特加密都具有随机性，也保障了安全。

（4）性能分析

文中的医疗数据隐私保护机制中选用了 RSA 公钥加密算法。RSA 中对极大整数做因数分解的难度决定了 RSA 算法的可靠性，RSA 算法产生的密钥长度越长，算法被破解的可能性就越小。基于大整数分解实现此加密方案，与基于双线性的加密方案相比具有较好性能。

3.5 本章小结

对数据进行加密是隐私保护机制中一种被广泛使用的方式。在这一节中，首先对医疗数据进行标准化处理，然后对标准化数据进行关键字提取，构建医疗数据的关键字索引结构，接着针对医疗数据研究实现了基于 RSA 公钥加密可搜索加密共享算法，最后从理论上分析医疗数据的加密共享方案可以很好地保护医疗数据隐私。

第四章　基于群签名和加密的交易隐私保护研究

如果说互联网的出现实现了信息交换，那么区块链的出现则是实现了价值转移[56]。联盟链中所有结点必须经过身份认证，才被允许加入网络，但是联盟链中所有授权节点共享交易账本[57]。使用联盟链进行医疗数据共享，医疗数据中包含大量隐私信息，为了保障医疗数据共享具有更高的安全性和更好的隐私保护，需要针对医疗数据具体共享过程研究实现交易隐私保护机制。

4.1 问题引述

4.1.1　超级账本 Fabric 交易过程研究

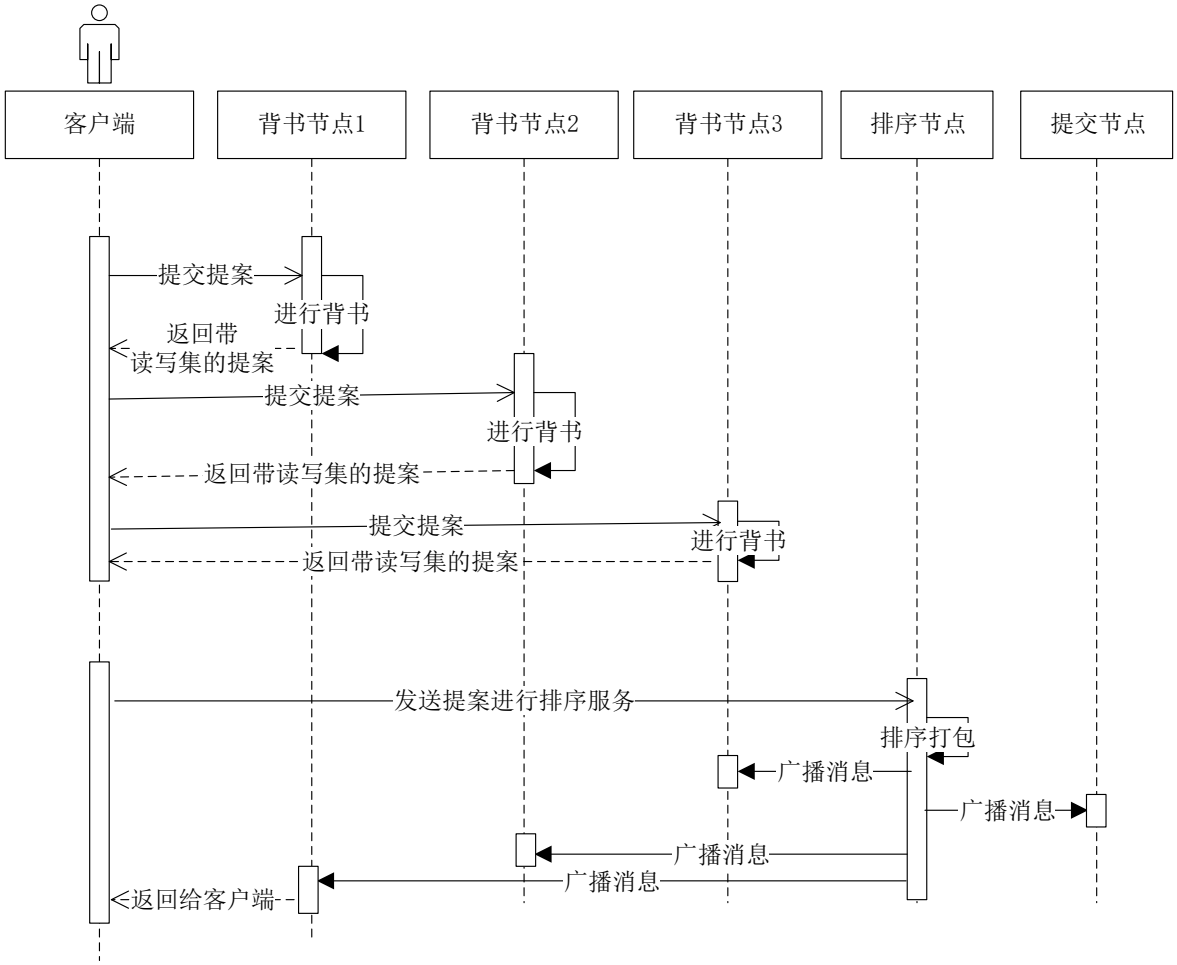

图4.1 Fabric 交易序列图

与其他区块链技术相同，超级账本 Fabric 自身使用数学知识和密码学原理确保

了所有交易稳定成功执行，但是研究图 4.1 的超级账本 Fabric 交易过程，发现在背书环节和排序环节可以改进。

背书节点进行背书时需要客户端的交易提案明文，以便调用状态数据库相关链码模拟执行交易，生成结果的读写集，并且对读写集签名后返回客户端。客户端根据背书结果构造带有读写集的交易提案发送给排序节点，排序节点在超级账本 Fabric 中仅提供交易的排序和打包服务，无需访问客户端交易提案的具体内容。

大量研究表明，通过分析区块链中交易数据，可以回溯交易创建者的信息[58]。所以为了保护交易数据隐私，在安全性更高的业务场景中，需要依据具体过程设计隐私保护机制。例如，在 Blindcoin 方案中，用户对输出地址采用盲签名和中心化的混币机制，使第三方能提供混币服务的同时，不会将用户的输入地址与输出地址相关联，从而保护用户交易信息[59]。本文研究实现了群签名和对称加密相结合的交易隐私保护机制，客户端创建交易提案时，使用加密算法对交易提案进行加密并且群签名，背书过程中发送交易提案密文和密钥，排序过程中仅传输交易提案密文。

4.1.2 群签名算法分析

群签名算法允许群成员代表群进行群签名，群签名验证者可以通过群公钥验证群签名，但是无法确定签名成员的具体身份，只有群管理员可以追溯到签名成员身份信息[60]。群签名算法具有匿名性的同时又保障了安全审核，将群签名加入超级账本交易创建过程具有非常重大的意义。比较分析群签名算法可知研究大致分为五类，具体如表 4.1 所示。

表4.1 群签名算法对比

群签名算法	优点	缺点
基于离散对数的 群签名算法	与之前群签名相比， 效率提高	群成员数量与群公钥及 群签名长度线性相关
基于知识签名的 群签名算法	群签名公钥 与群成员数量无关	群签名计算复杂， 效率低
基于身份的 群签名算法	安全性高	群管理员权限过大， 用户密钥存放安全问题
无证书 群签名算法	与基于知识签名群签名 算法相比， 效率提高	群公钥易被替换，存在 安全问题
基于证书 群签名算法	无密钥托管问题， 用户私钥参与签名	群成员撤销 存在问题

本节以文献[61]中提出的方案为前提，做如下分析。

（1）群签名方案

文献[61]中提出的群签名方案是在文献[62]中利用 Paillier 系统的基于身份的签名方案的基础上，运用文献[63]提出的将基于身份的签名方案转化成群签名的方法构造形成。该方案主要有 5 个算法。

群组初始化：该算法用于群签名的群组初始化。输入是安全参数 k，输出为群组公开参数（n，g）和秘密保存的密钥（p，q）。算法的执行者是可信的私钥生成机构。

成员入群：该算法用于描述成员入群的过程，该过程有两个子过程，第一个子过程的输入是用户的真实身份 ID，输出为$Q_{ID} = H_1(ID||r)$，该子过程由注册中心执行，输出的接收者是要加入群组的成员。第二个子过程的输入为上一个子过程的输出Q_{ID}，输出是（x，y_1）和（t，Z_{ID}）。第二个子过程由密钥颁发中心执行，（x，y_1）的接收者为用户，（t，Z_{ID}）的接收者为可信的中间机构。

生成群签名：该算法描述对消息进行群签名的过程，该过程有三个步骤。第一个步骤的输入是消息 m，输出是 c 和 R 以及（s_1，s_2），这个步骤的执行者是群成员，输出信息给中间机构进行二次签名。第二个步骤的输入是Q_{ID}和（R，c，s_1，s_2），输出是（R，c，Z_{ID}，s_1，s_3），该子过程的执行者是中间机构，输出由中间机构发送给签名接收者。

验证签名：该算法描述消息接收者对群签名进行验证的过程。输入是（R，c，Z_{ID}，s_1，s_3），输出是验证签名是否有效的结果，该算法由签名接收者验证公式$\mathrm{R} = Z_{ID}^{c} \cdot g^{s_1} s_3^{n} \ mod \ (n^2)$来完成。

打开群组：当签名接收者对签名有效性有争议时，该算法被执行。该算法首先由中间机构执行，最后由注册中心执行，输出为成员表中签名者身份信息。

（2）群签名方案分析

该方案是在应用 Paillier 机制的基于身份签名方案的基础上，利用基于身份的签名方案和群签名方案的相似性，使用特定的算法转换提出。基于身份的群签名方案安全性高，但是存在密钥无法托管和管理员权利过大的问题。该方案为避免上述问题，在群组织结构上来说，群管理员由用户注册中心和密钥颁发中心两部分构成，同时引入可信的中间机构。因为群组成员的私钥被分成两个部分，一部分由成员保存，另一部分由中间机构保存。所以，一方面避免密钥颁发中心伪造签名，另一方面中间机构对群组内成员的签名权有控制作用。在安全性保障方面，基于身份的群签名方案需要安全信道来传输用户私钥。

4.2 总体思路

本节从交易层出发进行隐私保护研究，整体思路如图 4.2 所示，用户生成交易提案发送给背书节点时，使用对称加密算法对交易提案加密，同时对加密后交易提案群签名，将交易提案密文和密钥发送给背书节点，背书节点使用群公钥验证群签名后，使用密钥解密获取交易提案明文，执行背书过程。但是用户客户端发送给排序节点交易提案时，发送群签名的加密交易提案，不发送密钥，从而在保证交易提案成功写入账本的同时具有匿名性，起到交易数据隐私保护的作用。

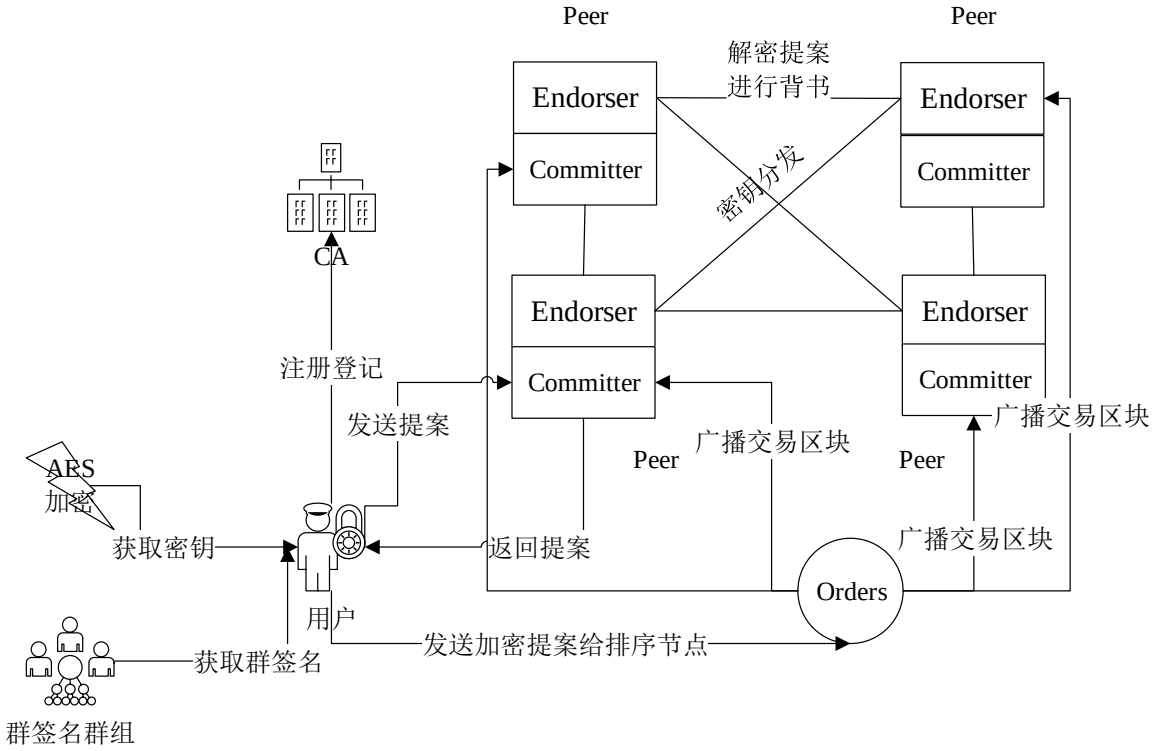

图4.2 交易隐私保护研究总体结构

4.3 方案设计

研究 Hyperldger Fabric 的 PKI 可以发现，如图 4.3 所示，证书体系把证书划分成三大类：一类是 E-Cert(Enrollment Cert)，另一类是 T-Cert(Transaction Cert)，还有一类是 TLS-Cert。E-Cert 携带参与方数字身份信息，用作身份证明。TLS-Cert 是通讯证书，当一方与另外一方通讯需要通过 Https 或者 Grpc 协议时，需要用到 TLS-Cert。T-Cert 与 E-Cert 最大的不同是，T-Cert 中不携带显式身份信息。E-Cert 与 T-Cert 是一对多的关系，并且 E-Cert 与 T-Cert 之间的关系通过特定的算法保护，因此在 Fabric 中我们不能追溯交易方的具体数字身份信息，并且 T-Cert 是服务于交易的，每一次交易使用不同的 T-Cert。

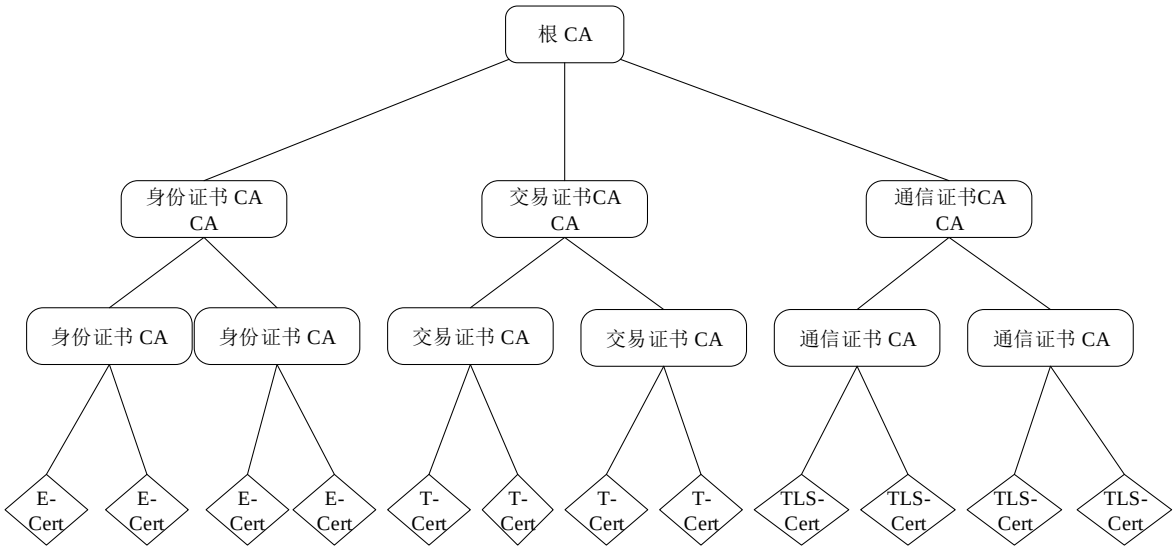

图4.3 Fabric PKI 结构图

4.3.1 方案改进

超级账本 Fabric 中 MemberService 组件提供了采用传统的公钥基础结构（PKI）的身份证书管理服务，采用分层的 PKI 结构来实现对用户证书的管理[64]。分析 4.1.2 中群签名方案的加入群组和群成员密钥颁发过程可以发现：用户入群时需要将自己的身份信息通过秘密通道发送给注册中心，同时密钥颁发中心也需要通过安全信道将密钥颁发给群成员和可信的中间结构。针对这个问题，进行改进。在文献[61]的基础上结合超级账本 Fabric 的 MemberService 组件，对加入群组和群成员密钥颁发过程中的关键信息进行加密处理，从而使群签名方案也适用于不可信环境。

4.3.2 结合超级账本 Fabric PKI 的群签名方案

在超级账本 Fabric 中一次交易一个交易证书，具有可跟踪性和不可伪造性，注册证书和交易证书的结合使得用户在网络中具有匿名性[65]。在文献[61]群签名方案中，管理员分权为注册中心和密钥颁发中心，这种思想与 Hyperledger Fabric 中用户注册时有注册证书，交易时会有一系列交易证书的思想不谋而合。所以本文研究实现了基于 Paillier 的群签名方案与 Fabric 中的 PKI 机制相结合的一个新的群签名方案。

方案假设：数据提供者申请成为群成员之前，加入超级账本 Fabric 网络，拥有注册证书，当发起交易时，会有一系列交易证书。一个交易证书就是一对公钥和私钥对。

与其他群签名方案类似，和 Hyperledger Fabric 的 PKI 体系相结合的基于 Paillier 的群签名方案也包括群组初始化，成员入群，生成群签名，验证签名，打开群组五个过程，如图 4.4 所示。

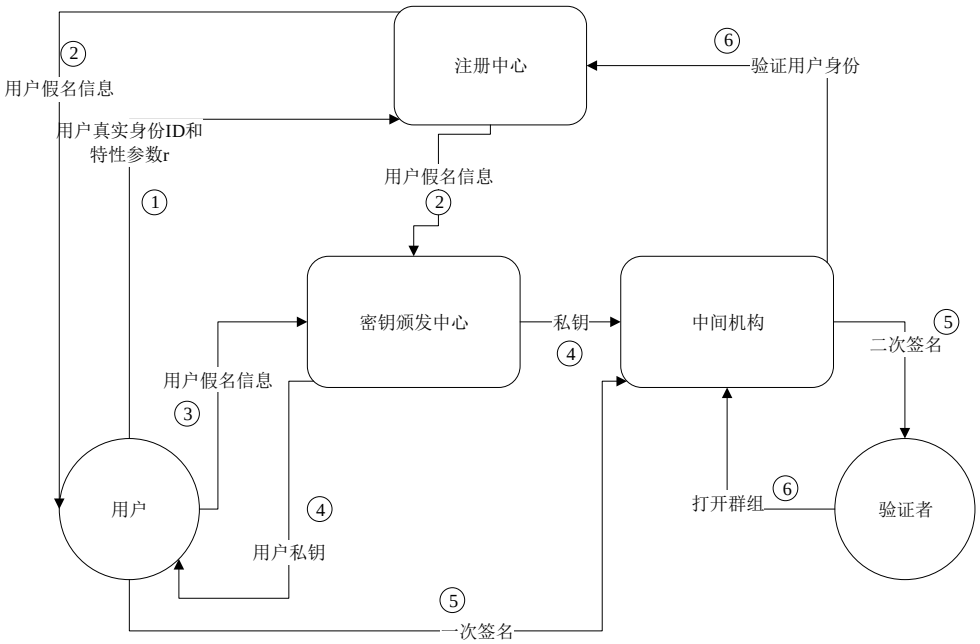

图4.4 群签名过程图

（1）形成群组

用户节点向医疗机构节点发送医疗数据申请请求，医疗机构将间隔时间为 t 内的所有数据请求节点组成一个群组，同时对按照请求顺序进行编号。由于医疗机构天然的政府资质背书，可作为可信中间机构。

（2）群组初始化

此过程输入为安全参数 k，输出为群公钥（n, g）和用来生成群成员私钥的主密钥 λ 或者（p,q），由可信机构 PKG（密钥生成机构）执行。具体过程是：根据安全参数 k 获取大质数 p 和大质数 q。使用式（4-1）和式（4-2）得到 n 和（p-1）与（q-1）的最小公倍数 λ。选取参数$g \in Z_{n^2}^*$，即 g 属于小于 n 的平方的整数组成的集合。同时满足条件$n|ord_{n^2}(g)$；选择分别满足式（4-3），式（4-4）和式（4-5）的密码学哈希函数H_1，哈希函数H_2和θ。选取一系列符合$u < n^2$，且u=1mod n 条件的元素u组成集合，对于集合中的每一个元素满足式（4-6）。

$$n = pq \tag{4-1}$$

$$\lambda = \mathrm{lcm}(p-1, q-1) \tag{4-2}$$

$$H_1: \{0,1\}^* \to \{0,1\}^k \subset Z_{n^2}^* \tag{4-3}$$

$$H_2: Z_{n^2} \times Z_{n^2}^* \to Z_n \tag{4-4}$$

$$\theta: Z_{n^2}^* \to Z_{n^2}^* \tag{4-5}$$

$$\mathrm{L(u)} = \frac{u-1}{n} \tag{4-6}$$

其中，Z_{n^2}为小于n^2的整数集合，$Z_{n^2}^*$为Z_{n^2}中与n^2互质的整数的集合。$ord(u)$表示 u 的阶次。

（3）用户入群

群管理员逻辑角色上分为注册中心和密钥颁发中心。用户与密钥注册中心交互：首先用户随机选取整数r作为自己的特性参数，使用注册中心证书对特性参数r和真实身份ID进行加密作为申请信息，再使用自己证书私钥对申请信息签名后发送给注册中心。其次，注册中心首先使用用户证书验签，若通过验证，使用自己的证书私钥解密密文得到成员特性参数r和用户真实身份ID。计算$Q_{ID} = H_1(ID||r)$，因为Q_{ID}为用户的假名信息，所以将Q_{ID}发送给用户和密钥颁发中心与通道是否可信无关，同时将用户（ID, Q_{ID}, r）信息存储群成员列表中，以便验证管理用户身份。

用户与密钥颁发中心交互：用户将假名信息Q_{ID}发送给密钥颁发中心，密钥颁发中心将收到的用户信息和上一步中注册中心发送的信息进行对比验证，结果一致时密钥颁发中心选取随机整数$\mathrm{t} \in Z_n^*$，使用式（4-7），式（4-8），式（4-9）和式（4-10）分别计算Z_{ID}，x，y，y_1。密钥颁发中心使用用户证书对该用户群签名私钥（x, y_1）加密后发送给用户，且由中间机构保存（t, Z_{ID}）。用户收到加密信息后用证书私钥解密得到明文信息，用户成功入群。

$$Z_{ID} = t^n Q_{ID} \ \ mod \ \ n^2 \tag{4-7}$$

$$\mathrm{x} = \frac{L(Z_{ID}^{\lambda} \ mod \ n^2)}{L(g^{\lambda} \ mod \ n^2)} \ \ mod \ \ n \tag{4-8}$$

$$\mathrm{y} = (Z_{ID} g^{-x})^{n^{-1} mod \ \lambda} \ \ mod \ \ n \tag{4-9}$$

$$y_1 = \frac{y}{t} \ \ mod \ \ n \tag{4-10}$$

（4）签名生成

群成员使用（x, y_1）签名，用户随机选取整数 a 和正整数 b 即$\mathrm{a} \in Z_n, \mathrm{b} \in Z_n^*$。使用式（4-11）计算得 R，输入 R 和消息 m 利用式（4-12）计算得 c。输入成员私钥得到（s_1, s_2）。

$$\mathrm{R} = g^a \cdot b^n \ \ mod \ \ n^2 \tag{4-11}$$

$$\mathrm{c} = H_2(m, R) \tag{4-12}$$

$$s_1 = a - cx \ \ mod \ \ n \tag{4-13}$$

$$s_2 = \ by_1^{-c} \ \ mod \ \ n \tag{4-14}$$

群成员将身份信息和上述计算得到的（ $\mathrm{R}, \mathrm{c}, s_1, s_2$）发送给中间机构，中间机构

先将成员身份信息Q_{ID}发送给注册中心，注册中心在群成员列表中查找，如果查找不到，则忽略发送信息，如果注册中心返回结果，则输入Q_{ID}，t和Z_{ID}，计算$Z_{ID} = t^n Q_{ID} \ mod \ n^2$等式是否成立，如果成立，则使用$t$计算得$s_3$，且将（$\mathrm{R}, \mathrm{c}, Z_{ID}, s_1, s_3$）发送给签名接受者。

$$s_3 = s_2 \cdot t^{-c} \ mod \ n \tag{4-15}$$

（5）接受者验证签名

接收者输入（$\mathrm{R}, \mathrm{c}, Z_{ID}, s_1, s_3$），利用式（4-16）是否成立来验证签名的有效性。

（6）打开签名

中间机构先使用式验证s_3的有效性，若s_3有效，则利用式验证Q_{ID}的身份，若确认存在，则将式（4-17）等式右部分发送给注册中心。注册中心在成员列表中查找满足式（4-17）的条件记录，得到签名者的真实身份。

$$\mathrm{R} = Z_{ID}^{c} \cdot g^{s_1} s_3^n \ mod \ n^2 \tag{4-16}$$

$$\mathrm{H}(ID||r) = \frac{Z_{ID}}{t^n} \tag{4-17}$$

4.3.3 方案分析

为保证匿名性，申请入群的用户将身份信息加密传输给注册中心，注册中心构造假名信息发给用户和密钥颁发中心，之后过程使用用户的假名信息，所以，在整个群签名过程中，只有注册中心知道用户的真实身份信息，假设注册中心和其他密钥颁发中心等不会相互勾结，这样群成员在此方案中具有匿名性。用户假名信息由用户随机选取整数构造，群成员私钥生成时随机选取整数，所以，每次群签名具有不可关联性。用户真实身份信息由注册中心保存，当争议发生时，由可信机构和密钥注册中心依打开签名的算法打开签名，找出真正的签名者，具有可监管性和可追踪性。该方案中群签名由群成员签名和可信中间机构签名两部分完成，群成员只掌握自己的部分私钥，群成员之间私钥相互保密，所以群签名不可伪造。在用户发送身份信息的过程中使用用户证书对身份信息进行处理，使得适用于非安全信道。

4.4 结合群签名和加密的交易过程

在联盟链中，只有被授权的节点才可以加入网络，参与 Fabric 交易过程的角色从逻辑上划分为客户端，Endorser 节点，Committer 节点，Order 节点和 CA 五种[66]。大体上来说，这五种逻辑角色也可划分为记账节点，参与节点和交易请求节点。peer 节点（逻辑上划分为 Endorser 节点和 Committer 节点）担任记账节点的角色，执行链码，

更新账本。Order 节点对交易进行排序服务，参与交易过程，但是无权过问记账过程。

假设交易双方已经通过客户端 SDK 注册且已通过 CA 中心认证，获取证书。网络中的记账节点都已经安装好链码且在网络中实例化，对链码操作生成交易。

新方案中具体的交易流程如图 4.5 所示。

（1）形成群组

根据目前现状，医疗数据都是存储在医疗机构数据库或者内部云服务中，患者对医疗机构提交数据申请请求，医疗机构对患者身份认证后，医疗机构将间隔时间为 t（t 由医疗机构决定）内所有提交数据请求的用户组织成一个群组。使用上述改进的群签名方案，形成群组，公开群公钥（n,g），患者获取自己的群私钥（x,y_1），同时医疗机构担任可信的中间机构，保存$(\mathrm{t},\ Z_{ID})$。

医疗机构与患者要求相结合，对数据标准化处理且将所需医疗数据用证书签名发送给患者。

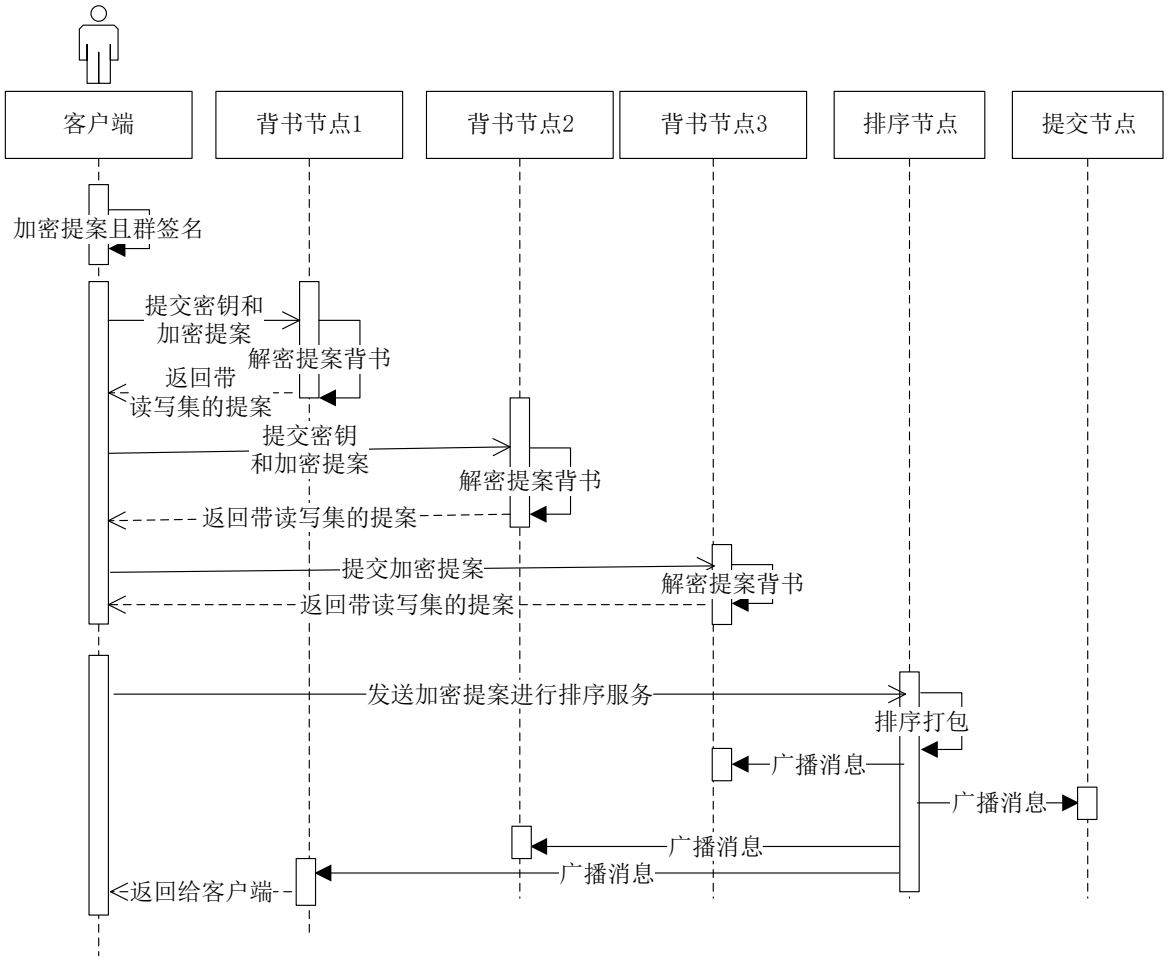

图4.5 改进交易序列图

（2）创建交易提案

患者初始化 AES 对称加密算法，得到密钥。为防止上一次交易中密钥丢失影响下一次交易，不同交易生成不同密钥，互不影响。

患者收到数据后验证医疗机构身份，使用 3.3 小节处理数据，然后将密文数据，关键字索引结构密文，关键字序列密文等作为链码输入参数。再加上使用授权背书节点公钥对 AES 算法密钥加密，在本地生成交易提案。

交易请求节点先用证书私钥对交易提案签名，然后使用 AES 对交易提案加密，接着使用自己成员私钥（x, y_1）对加密提案签名，将信息发送给可信的医疗机构进行二次签名，最后将交易提案密文和密钥密文一同发送到背书节点。交易数据在区块链中匿名和安全传播。该交易提案包括群签名（R,c,Z_{ID}, s_1, s_3）信息，链码标识信息，链码的输入参数等，具体结构如图 4.6 所示。

<table>
<tr><td colspan="2">AES对称加密</td><td rowspan="2">群签名：
用户签名，
可信机构签名</td></tr>
<tr><td>交易提案：
链码参数，链码标识，提案标识，
加密AES密钥等</td><td>用户
证书签名</td></tr>
</table>

图4.6 背书节点交易提案结构图

（3）背书节点验证签名和模拟交易。背书节点先使用等式 $R = Z_{ID}^{c} \cdot g^{s_1} s_3^n \ mod \ n^2$ 验证客户端群签名的合法性，然后使用自己私钥解密 AES 密钥，使用被解密密钥解密交易提案。检查交易发起者是否具有写账本的权限和检查该交易是不是新交易等，验证通过后，与链码的 docker 实例通信，模拟执行交易（此时账本的状态不会有任何改变），将 State Database 读写集使用证书签名和是否背书的结果返回交易发起者，此时还没有进行正式交易。若按背书策略有多个背书节点，背书节点之间通过 gRPC 消息传播 AES 密钥。

<table>
<tr><td colspan="2">AES对称加密</td><td rowspan="2">群签名：
用户签名，
可信机构签名</td></tr>
<tr><td>交易提案：
链码参数，链码标识，提案标识等</td><td>用户
证书签名</td></tr>
</table>

图4.7 排序节点交易提案结构图

（4）交易发起者收到背书节点的返回信息后，若按背书策略收到背书结果，且交易提案模拟结果一致，则将交易提案密文，模拟的 State Database 读写集等打包成交易提案，加密并且签名后发送给 Order 节点。为保障交易隐私，使用 AES 对交易提案加密且群签名，签名过程与发送给背书节点时签名过程相同，不再赘述。最大的不同是，不会把 AES 密钥发送给排序节点。具体交易提案结构如图 4.7 所示。

（5）Order 节点作为普通参与节点，独立于 peer 节点流程之外，以先到先得的方式为所有的交易做共识排序，所以收到交易提案密文后，使用群公钥验证交易信息，若验证成功，则排序打包输出区块，并且向所有 peer 节点（包括 Committer 节点）广播，否则忽略该交易提案。

（6）Committer peer 节点接收到 Order 节点的区块后，先对区块数据进行 VSCC（validator system chaincode）校验，检查区块整体数据是否正确，然后验证块中每一个交易的 State Database 读写集与 State Database 的数据版本是否一致。若区块中所有交易都验证结束，则该区块加入区块链，且将结果一致的 State Database 读写集写入 State Database，同时在 LevelDB 中建立索引。至此区块链中完成了执行链码且更新账本。若在交易过程中对群签名有争议，则由管理节点（包括可信机构和密钥注册中心）依方案中打开群签名，找出真正的签名者，具有可监管性。

4.5 本章小结

群签名是利用公开的群公钥签名和验签的方案。本节首先研究超级账本 Fabric 交易过程，发现交易提案在背书和提交环节隐私性不足的问题。其次，对比研究现有群签名算法，根据超级账本 Fabric 特点使用基于 Paillier 的群签名方案，针对原群签名方案中必须通过可信通道下发群成员私钥的问题，结合超级账本 PKI 机制研究实现了新的群签名方案。最后使用新群签名方案和对称加密算法对交易提案进行改进，可使交易请求节点发起交易请求时，匿名保护身份隐私和交易提案在区块链中安全传播。并且当交易提案的群签名出现争议时，支持群管理员打开群组，追踪签名者身份，达到监管的目的。

第五章 基于区块链的医疗数据共享实现

5.1 问题引述

5.1.1 常见医疗数据共享模型

不论对患者、医生还是同生态圈的医疗机构，常见的医疗数据共享都十分不便[67]。如表 5.1 所示，常见医疗数据共享系统存在共享动力不足，标准不一，隐私安全等诸多问题。

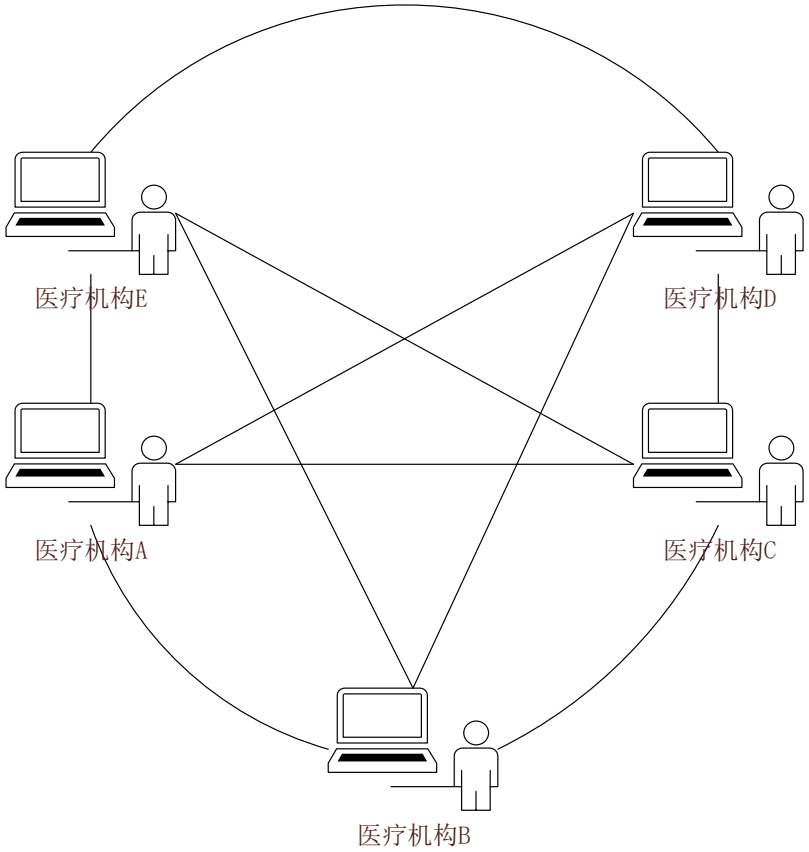

图5.1 医疗数据共享结构图

对于患者来说，患者转诊时需携带纸质记录，或转诊医生通过传真转发医疗记录。同样，新的医疗服务提供者必须手动键入或者扫描记录到 EHR。

对于医生来说，申请病人相关健康史的过程耗时耗力，甚至无法开展。即使最终获取患者所有相关健康记录，还需要相关人员将历史纪录手动键入或扫描至当前医院系统。医生还必须定期将患者记录转发给其他医疗服务提供者。

对于同生态圈的医疗机构来说，医疗机构采用的数字化信息平台都是为单个医院或者医院内某部门服务，如果业务逻辑涉及两个医院，目前被广泛采用的方案为这两

个机构分别设计接口，如果遇到涉及更多机构的场景，为不同医疗机构设计一对一的接口，那么很可能出现如图 5.1 所示情况。系统中有C_5^2个接口，参与方越多，系统的互操作性和稳定性越差，效率越低。

表5.1 医疗数据共享关键问题研究

类型	关键问题	描述
缺乏动力	共享成本高	需要大量人力物力资源投入共享系统的建设和维护
	缺乏激励机制	现有系统对于共享行为无相应激励机制
	消极共享	医疗数据作为医疗机构核心，被拒绝共享
	数据主体确权	医疗数据由医疗机构产生，但是包含患者隐私，共享给第三方后，数据难以确权
标准不一	医疗数据结构不一	由于医疗机构科室不同，数据结构不一
	语义不一	对于同一事物，医学表达标准不一
	信息交换标准不一	医疗机构信息化建设程度不一，导致信息交换标准不一
	数据质量不一	医疗设备和医疗技术人员水平不一，导致医疗数据质量参差不齐
隐私安全	患者无知情同意权	患者对包含自己隐私的数据无知情控制权
	数据泄露	数据共享过程中无相应隐私保护机制
	技术风险	采用有安全隐患的硬件或软件设备
	相关人员无隐私保护意识	患者或医务人员因无隐私保护意识造成隐私泄露
	法律法规不完善	国家角度上无完善层级法律法规

5.1.2 Fabric 的隐私保护机制

超级账本 Fabric 使用分布式账本的思想提供了一个去中心化网络，参与网络的所有节点之间公开透明共享数据，但在医疗数据共享过程中，数据提供者不需要和数据需求者共享所有数据。这就需要系统提供数据隐私保护机制，超级账本 Fabric 使用了通道机制来进行隐私保护[68]。

初始化区块链网络的同时创建通道，并且根据协商将所需 Peer 节点和 Order 节

点加入通道。如果一个 Peer 节点加入不同通道，则该 Peer 节点同时维护不同通道的不同账本，这些账本之间状态不一致但互不影响，以此实现数据访问控制。但是研究上述过程发现：

（1）一般情况下，联盟链在初始化网络的同时创建通道的配置交易文件，以这样的方式来创建通道。

（2）联盟链网络中节点无法动态退出通道。

（3）通道间账本状态互不影响，数据隔离，无法共享。

所以为了更好地在基于联盟链的医疗数据共享过程中进行隐私保护，文中从医疗数据加密和交易隐私保护这两个方面给出解决办法。

5.2 总体思路

从目前区块链技术的应用场景来看，利用区块链技术的应用开发大致分为三类：以太坊、超级帐本和其他[69]。由于超级账本 Fabric 的身份认证功能，可以从访问控制的角度保护数据隐私，所以选用联盟链进行医疗数据的共享，同时将上述数据隐私保护研究和交易隐私保护研究应用于联盟链的医疗数据共享中。

在医疗领域中数据共享的是数据的使用权，而不是数据本身。本文目的在于保障医疗数据在数据提供者和数据需求者之间共享的同时保证隐私，这里的隐私包括医疗数据隐私和交易隐私。

针对医疗数据的隐私保护，研究保护患者隐私安全的条款发现，不同机构对于患者隐私和数据授权范围都有着不同的管理规定。文中根据患者与数据需求者的协商结果，使用 3.4 小节处理数据，实现数据提供者对数据的完全控制和数据需求者对数据的安全访问。同时在区块链网络中锚定相应数据和操作记录生成交易提案，上链数据无法更改，从而也确保了记录的完整性。

对于交易隐私保护，文中使用 4.3 小节群签名和 AES-128 对称加密算法，与超级账本 Fabric 交易过程相结合，使用 AES-128 对交易提案进行加密并且群签名。在保证交易匿名性的同时通过加密算法保护交易数据隐私。

5.3 系统模型设计与流程

5.3.1 模型参与方定义

医疗数据共享过程中的参与者有：医疗数据提供者，医疗机构，基于联盟链的共享网络，医疗数据需求者，数据存储结点。现实生活中，大多医疗数据都是存储在医疗机构中，被视为宝贵的资产。本文为简化医疗数据共享过程，假设患者在医疗机构诊断治疗，原始医疗数据保存在医疗机构，但是该医疗数据是否参与共享、与谁共享

以及以怎样的形式共享等，医疗机构没有决定权，而是由患者本身决定。

医疗机构 A。医疗数据的原始存储机构，在患者和数据需求者之间数据共享时，患者向医疗机构申请数据，医疗机构在保护患者隐私的前提下，依据相应医疗标准对数据标准化处理，发送给患者。同时作为可信中间机构，对患者的交易提案进行二次签名。

数据提供者 B。医疗数据的拥有者和提供者，对医疗数据进行关键字提取且对数据加密处理，以便和数据需求者进行安全地数据共享。

数据需求者 C。需要数据进行以医疗数据为支撑的科研活动，与数据提供者协商所需数据内容和格式，同时在接受数据的时候可以发起搜索请求，验证数据提供者的身份。

数据存储结点 D。由患者同意共享且标准化处理的医疗数据密文和索引存储在数据存储结点。由数据存储结点提供关键字索引支持，以便数据需求者能够快速地获取所需数据。同时维护账本状态并运行容器内链码对账本进行读写操作。

可信的密钥生成机构。提供加密算法所需要的密钥。

联盟链。医疗数据属于行业内数据，所以选择联盟链作为数据共享网络，加入联盟链的成员必须通过身份认证。联盟链在本文中作为节点的共享交易日志记录。数据提供者的原始医疗数据散列值、上传数据操作，数据需求者获取数据操作等都存储在链上。

5.3.2 模型建立

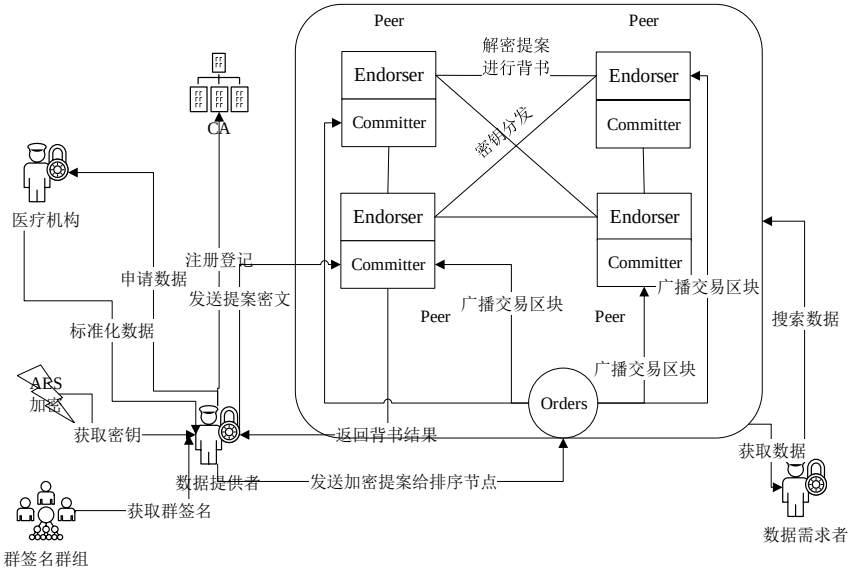

图5.2 基于 Fabric 的医疗数据共享总体架构

如图 5.2 所示，基于 Fabric1.0 搭建网络，数据提供者，数据需求者等使用客户端通过 SDK 向 CA 请求注册和登记，获取注册证书。HyperLedger Fabric CA 采用双密钥对 PKI 体系，即用户注册时，数字证书中心颁发签名密钥对和加密密钥对。签名用途的私钥严格保密，只能用户本人持有。链码已经被安装在 peer 节点上，并且在通道上实例化。设置背书策略，指定所有 peer 节点必须对网络中所有交易进行背书。

数据共享涉及的操作有三类，分别是数据提供者上传数据，数据存储和基于关键字检索获取数据。通过上传和获取完成医疗数据共享，且在共享过程中可实现关键字的数据检索。

使用第四章的结合超级账本 Fabric PKI 和基于 Paillier 的群签名方案来执行系统的初始化，依据 3.4 小节算法编写链码，链码已经被安装在 peer 节点上，并且在通道上实例化，数据提供者和数据需求者通过客户端 SDK 对链码操作，对链码的操作对发送交易，进而生成新的区块。

依据 3.4 小节使用密钥对医疗数据加密，同时提取关键字，按照 3.1 小节构建数据集的关键字索引，将密文数据和索引结构发送给数据存储节点的同时上链。数据存储节点管理所有共享数据，整合所有数据的索引结构，更新账本。数据需求者输入搜索关键字，调用链码，返回符合搜索条件的密文集合。具体链上链下数据对比如表 5.2 所示。

表5.2 医疗数据链上链下对比

角色	链下数据	链上数据
数据提供者	原始医疗数据	标准化、结构化的数据密文 关键字索引结构 原始医疗数据 HASH 值 原始医疗数据存储位置 数据提供者上传数据操作
存储节点	原始标准化数据密文	所有关键字索引结构 数据需求者获取数据操作

5.3.3 医疗数据标准化

医疗数据本身结构复杂，存在数据格式不统一、数据医疗标准不统一等问题，同时区块链中区块存储容量有限，无法直接将原始医疗数据上链共享，所以为了更加高效地进行数据共享，并且保护数据隐私，需要对医疗数据进行标准化、结构化处理。

按照 3.2 小节对医疗数据进行预处理，生成的结构化数据与原图像数据相比，所需存储空间大大减少，但是包含原始医疗数据的关键信息。以小儿白内障的医疗图像数据为例，该结构化数据如图 5.3 所示。

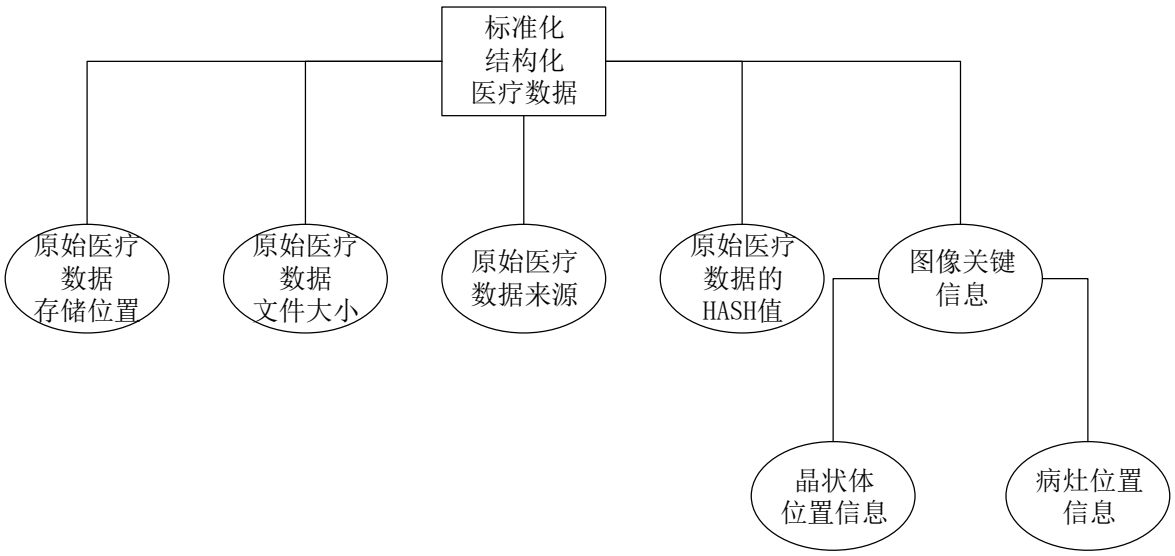

图5.3 标准化的小儿白内障数据结构

5.3.4 数据加密存储

医疗数据共享过程中的数据存储包括：原始医疗数据的存储，预处理之后医疗数据的存储，医疗数据关键字索引结构的存储，医疗数据关键字序列的存储。上述医疗数据的存储集中在申请医疗数据和上传医疗数据两个环节。

申请医疗数据：原始医疗数据存储在医疗机构。患者发送数据申请请求到医疗机构，医疗机构对申请信息做验证处理。验证通过后，根据医学标准对患者的医疗数据做标准化处理，然后使用患者公钥对结构化数据加密，对数据关键词做签名附着于加密数据，表明标准化处理的合法性和可靠性。最后发送给患者。

上传医疗数据：患者拿到医疗数据后，先用医疗机构的公钥验签，然后再使用自己私钥对加密数据解密。根据 3.4 小节的加密算法，患者使用数据需求者公钥对医疗数据进行加密，根据 3.3 小节构建关键字的索引结构，对索引结构加密，使用存储节点的公钥对关键字序列加密等。

5.3.5 数据共享过程

患者作为数据提供者，将结构化数据密文、关键字索引等存储于联盟链网络，根据协商使用不同数据需求者的公钥处理数据，只有真正的数据需求者通过存储节点验证后，才能获取密文且解密数据。患者将上传数据到存储节点，存储节点合并密文数

据的关键字索引，数据需求者获取所需数据等操作都是通过执行链码来推进。具体过程如下：

（1）共享数据请求

数据需求者 C 向患者 B 发送数据共享的请求，请求中包括数据类型，数据中关键字等。

（2）共享数据发送

数据提供者 B 根据协商结果，用 C 的公钥对数据加密、构建关键字索引结构、生成数据获取认证信息等，发送到数据存储节点 D。存储节点验证信息后，合并索引结构。

（3）共享数据获取

数据需求者将关键字密文，数据获取凭证等发送给数据存储机构 D，D 验证数据需求者获取凭证，若验证成功，使用自己私钥解密关键字，依据关键字搜索数据，成功则返回密文集合，否则返回空集。

（4）访问所需数据

数据需求者获取密文后，使用自己私钥解密得到医疗数据对应的标准化数据。从标准化数据可以获取原始医疗数据的存储位置等信息，访问存储位置获取原始数据。通过链码中的访问控制逻辑限制医疗数据访问权限，只有特定的访问者可以获取数据实现数据的共享。

5.3.6 加密关键词检索

本文数据共享过程中关键字检索获取数据分为两种情况：一种是指定的数据需求者检索且获取数据，另一种其它参与者检索数据。根据 3.4 小节，数据提供者上传数据医疗数据时，提供了关键字的搜索验证，并且随着密文数据发送给数据存储节点。只有数据需求者的关键字凭证和数据提供者的关键字验证一致，数据需求者才能获取数据，否则，只能获取链上数据集是否存在包含该关键字的数据，并不能获取密文。

5.3.7 Fabric 中交易提案加密

数据提供者与数据需求者都是通过 Fabric SDK 中的客户端实例和本地密钥参与联盟链网络，如图 5.4 所示，客户端实例提供设置加密模块方法 set_crypto_suite，具体加密模块由 CryptoSuite 接口实现。该加密模块中默认实现了 ECDSA 系列签名算法，AES128、AES256 的对称加密算法和哈希算法等，所以在数据提供者客户端创建交易提案时，调用默认提供的 AES-128 算法对交易提案进行加密，客户端实例存储 AES-128 的密钥，同时将密钥随交易提案发送给背书节点，背书节点之间通过 gossip 协议传递密钥。背书节点接收到 SDK 的交易提案后也使用 AES128 解密提案密文，

然后执行背书过程，返回背书结果给客户端实例。

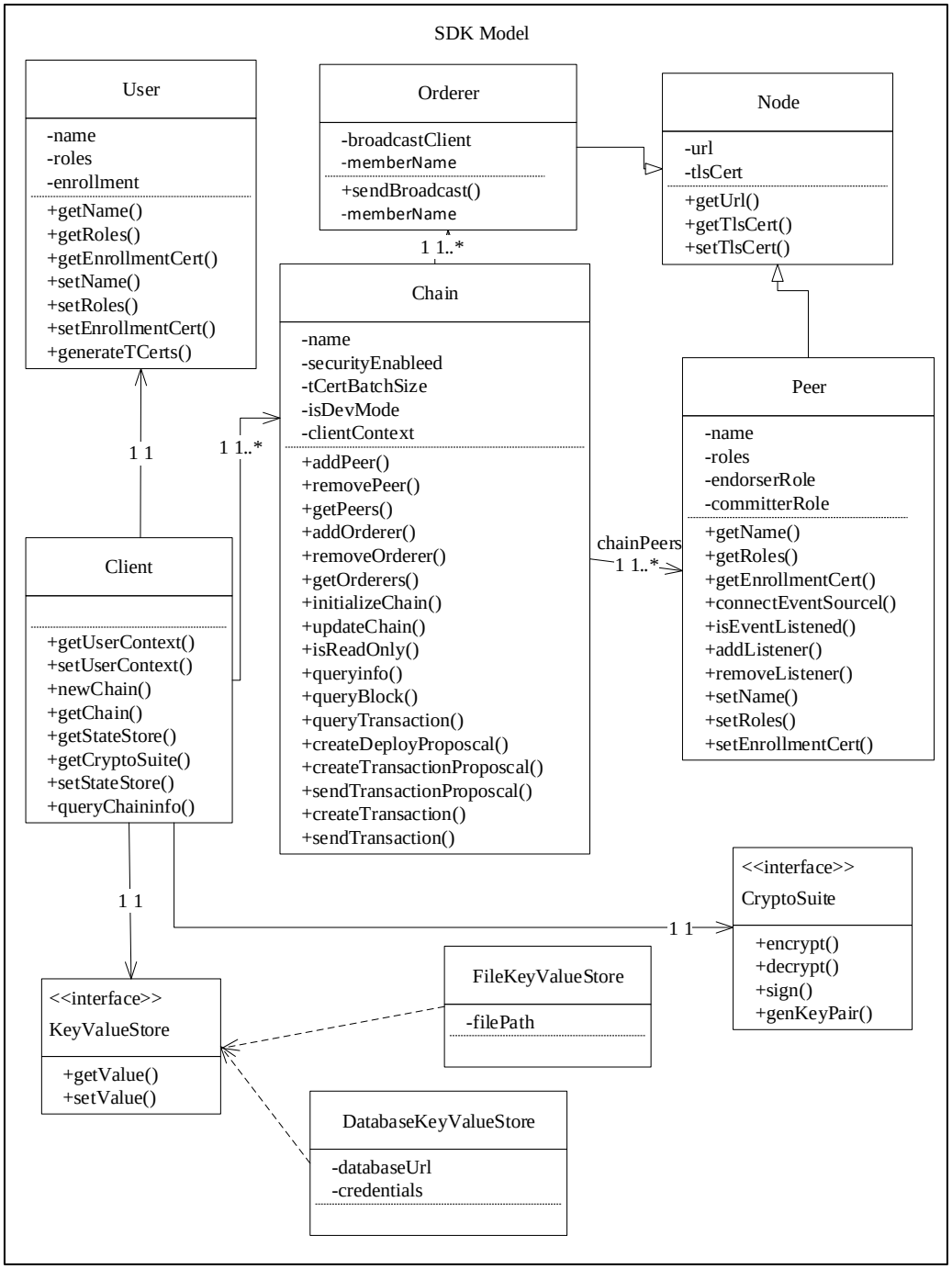

图5.4 Fabric SDK 结构图[70]

5.4 实验与分析

5.4.1 实验环境

本文使用如表 5.3 所示测试环境进行测试验证。联盟链中包含一个 orderer 节点，

三个组织，每个组织各自包含 peer 节点，和一个 fabric-ca 服务。

表5.3 基于超级账本 Fabric 的医疗数据共享实验环境

类别	名称	配置/版本	数量
硬件环境	服务器	Intel Xeon CPU E5-2620 8 核、64G 内存、1TB 磁盘	3
	笔记本	Intel(R)Core(TM)i5-2450M 处理器，4G 内存	1
软件环境	操作系统	Ubuntu16.04	1
	Hyperledger Fabric	V1.0	1
	Docker	18.03.0-ce	1

5.4.2 医疗数据加密共享分析

表5.4 医疗数据集

数据集名称	样本量/个	关键字数量/个
医疗数据 1 （小儿白内障裂隙光源图像数据）	1292	329
医疗数据 2 （小儿白内障红反光源图像数据）	839	574
医疗数据 3 （小儿白内障弥散光源文本数据）	2037	290

如表 5.4 所示，白内障数据集来自实验室与中山大学眼科中心合作项目，按照小儿白内障数据分类标准进行处理。原始图像数据大小差异较大，有些图像数据 200 多 KB，有些图像数据甚至达到了 2MB。按照 3.2 小节进行预处理，预处理之后单个图像数据所对应的标准化数据的大小为 1KB 左右，大大减少了存储空间。图 5.5 为预处理之后的数据。

医疗数据共享测试主要是通过超级账本 Fabric 来共享数据，主要看数据共享参与方是否可以正确获取数据，医疗数据共享测试用例表如表 5.5 所示。

```
<annotation>
    <folder>dense</folder>
    <filename>20181008-5456-62145</filename>
    <path>D:\cataract\dense\20181008-5456-62145.jpg</path>
    <source>
        <database>Unknown</database>
    </source>
    <size>
        <width>4752</width>
        <height>3168</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
        <name>dense</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>2803</xmin>
            <ymin>149</ymin>
            <xmax>3703</xmax>
            <ymax>1229</ymax>
        </bndbox>
    </object>
    <object>
        <name>lens</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>2623</xmin>
            <ymin>14</ymin>
```

图5.5 dense 分类小儿白内障数据预处理结果

表5.5 基于超级账本 Fabric 的医疗数据共享测试

用例名称	基于超级账本 Fabric 的医疗数据共享测试	
用例说明	数据提供者上传数据，数据存储节点合并索引数据，数据需求者下载数据，实现数据共享。	
前置条件	数据提供者拥有数据。	
测试情况		
序号	**输入及步骤**	**测试结果**
1	数据提供者对医疗数据分类，对数据标准化处理。	生成结构化数据且内容正确
2	数据提供者对结构化数据提取关键字。	生成关键字集合且内容正确
3	数据提供者对医疗数据加密，并构建基于关键字的数据索引结构。	数据关键字索引结构正确，且可以使用。

表 5.5 基于超级账本 Fabric 的医疗数据共享测试（续）

4	数据提供者对关键字加密	生成关键字密文且内容正确
5	数据提供者发送数据索引结构给存储节点	存储节点接收数据且格式正确
6	存储节点提供基于关键字索引获取数据	关键字索引结构正确且可以使用
7	数据需求者对关键字加密	生成关键字凭证
8	数据需求者发送关键字凭证给数据存储节点	数据存储节点接收数据且数据内容正确
9	数据存储节点验证数据需求者的关键字凭证	验证成功
10	数据需求者下载数据	数据包含关键字，格式和内容正确
测试结论	测试通过	

根据 3.3 小节的算法编写链码，对链码的操作会发送交易，进而产生新的区块。在基于超级账本 Fabric 进行医疗数据共享的过程中，为保障数据需求者对数据的安全访问，文中对数据需求者通过关键字进行身份验证，验证通过后，基于关键字搜索数据需求者的数据，如图 5.6 所示，随着区块链中医疗数据样本量的增加，根据关键字获取医疗数据的时间越来越长，但是可以看到，医疗数据样本量越多，搜索耗时增长率越低，说明构建医疗数据的关键字索引结构在医疗数据共享中，起到了很大的作用，可以提高数据共享的效率。

根据 3.3 小节的构建关键字索引结构算法可知：

索引结构树中存储比特序列的叶子节点与标准化、结构化密文数据一一对应。

索引结构树中所有节点都存储比特序列，序列长度与关键字数量一致。

分析以上可知，存储于区块链的索引结构所占空间与数据源的数量无关，与数据种类和数据集中所有互不相同关键字数量有关。数据越多，关键字索引结构中的叶子节点就越多，依据从下到上的索引树构造过程，关键字索引树越高，所占空间就越大。另外，因节点中存储比特序列长度与关键字集合大小一致，所以关键字的种类越多，总体关键字索引结构不发生改变，但是每个节点增加占据空间。

从表 5.4 可以看到，三类医疗数据中医疗数据 1 与医疗数据 2 相比，数据样本量相差不大，医疗数据 2 中关键字数据量将近是医疗数据 1 中关键字数量的一倍。医疗数据 1 与医疗数据 3 相比，数据中包含的关键字数量相差不大，但是医疗数据 3 的样本量是医疗数据 1 的一倍。如图 5.7 所示，文中对三类医疗数据分别进行共享耗时分

析。分析结果，可以看到，在基于超级账本 Fabric 的数据共享中，三类医疗数据随着样本量的增加，耗时趋势一致，数据样本对数据共享的影响更大。

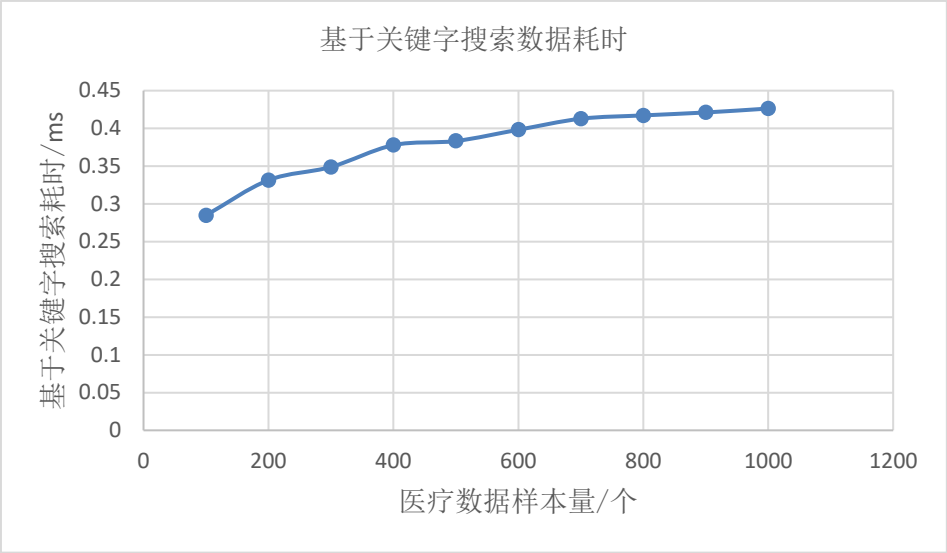

图5.6 基于关键字获取医疗数据耗时

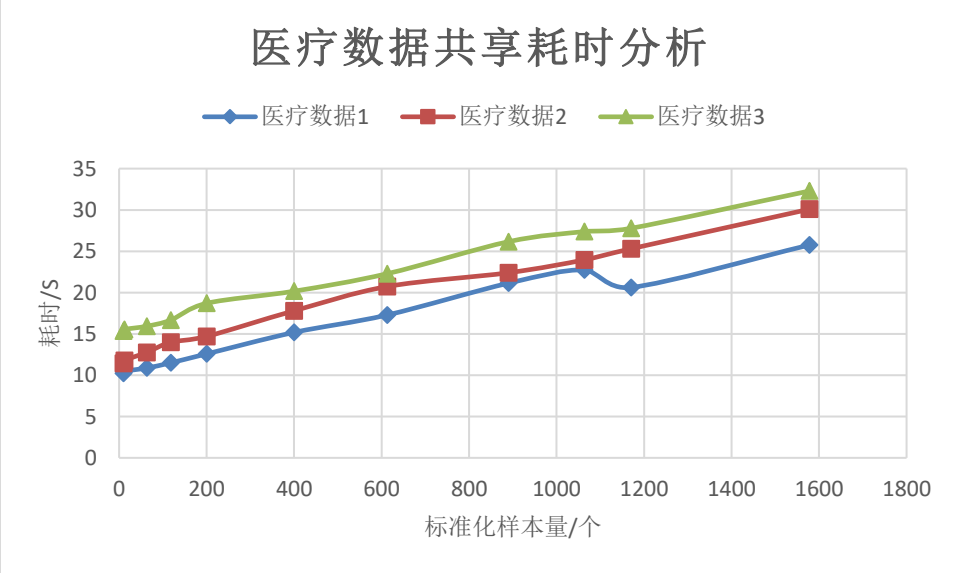

图5.7 基于超级账本 Fabric 的不同医疗数据共享耗时对比

文中为保护医疗数据隐私，同时考虑到医疗图像数据所需存储空间较大和区块链中存储容量有限，对原始医疗数据进行处理后，研究且实现加密方案。分析基于区块链的医疗数据加密方案，为了达到医疗数据共享的目的，首先由数据提供者按照协议将数据密文上传到存储节点，然后数据需求者使用进行密文关键字搜索验证，验证成功后获取密文，最后使用自己私钥解密获取数据，这个共享过程可以看出系统中关键

操作集中在数据提供者上传数据，数据需求者关键字搜索并获取数据。

文中医疗数据共享网络中，数据共享由数据提供者使用数据需求者公钥将数据传送到数据存储节点，数据需求者向数据存储节点验证身份后得到标准化结构化数据。以上过程中需要将相关操作和数据构建成交易发送到联盟链上，由记账节点记账且同步交易信息到全网账本中。

比特币系统每 600 秒产生一个区块，以太坊每 15 秒产生一个区块。而在超级帐本 Fabric 提供了依据业务逻辑配置网络中的出块时间和每个交易区块的大小。实验中不断进行数据上传下载操作，原始医疗数据不存储在区块链网络中，上链数据为交易操作以及标准化数据密文及其哈希值等，每一个交易提案包含的数据较小，所以如图 5.8 所示，修改 configtx.yaml 文件设置参数 BatchTimeout 指定每 10 秒产生一个区块，设置参数 MaxMessageCount 指定每一个交易块包含的数据为 300 个交易。依据超级账本设计的出块策略，若网络中交易频繁只要交易数大于 300，不管是否达到设定时间 10 秒都会产生区块，同理若网络中交易较少只要达到设定时间也会产生区块。

```
# Batch Timeout: The amount of time to wait before creating a batch.
BatchTimeout: 10s

# Batch Size: Controls the number of messages batched into a block.
BatchSize:

    # Max Message Count: The maximum number of messages to permit in a
    # batch.
    MaxMessageCount: 300

    # Absolute Max Bytes: The absolute maximum number of bytes allowed for
    # the serialized messages in a batch. If the "kafka" OrdererType is
    # selected, set 'message.max.bytes' and 'replica.fetch.max.bytes' on
    # the Kafka brokers to a value that is larger than this one.
    AbsoluteMaxBytes: 10 MB

    # Preferred Max Bytes: The preferred maximum number of bytes allowed
    # for the serialized messages in a batch. A message larger than the
    # preferred max bytes will result in a batch larger than preferred max
    # bytes.
    PreferredMaxBytes: 512 KB

# Max Channels is the maximum number of channels to allow on the ordering
# network. When set to 0, this implies no maximum number of channels.
MaxChannels: 0
```

图5.8 超级账本 Fabric 出块时间与区块大小配置信息

本文主要研究医疗数据共享的隐私保护机制，为保护数据隐私，实现数据提供者对数据的完全控制和数据需求者对数据的安全访问，文中的基于医疗数据的隐私保护研究中加密环节使用了 RSA。RSA 与 AES 相比，RSA 算法原理简单，易于使用，密钥长度越长，安全性越高，但是密钥长度越长，加解密的速率越低。但是由于使用非对称加密算法前，对原始医疗数据进行了标准化和结构化处理，处理之后实验数据量较小，如图 5.9 所示基于区块链的数据共享中，当共享数据量较小时，该方案有着适用性。

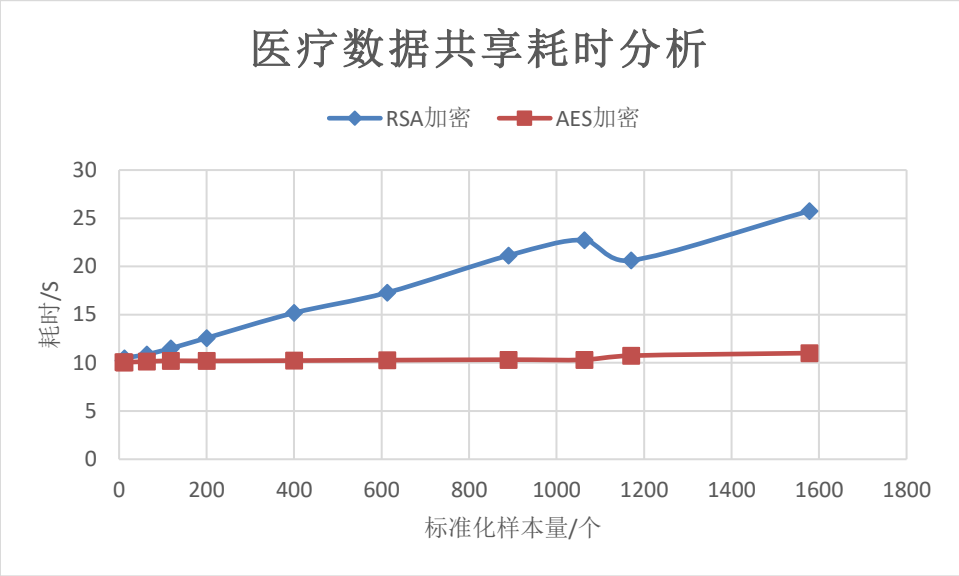

图5.9 基于超级账本 Fabric 的医疗数据共享耗时对比

在第三章医疗数据隐私保护研究中的加密环节分别使用 AES-128 和 RSA 对医疗数据加密共享。如表 5.6 和表 5.7 所示，在相同数据量时，在加密耗时、解密耗时和加密后文件大小三个方面进行对比分析，并且通过联盟链网络共享耗时分析如图 5.9 所示。随着医疗数据量的增大，与使用 AES-128 进行医疗数据共享相比，使用 3.4 小节算法进行数据共享耗费的时间越来越长，两者的时间差越来越大。但是若使用 AES-128 加密进行医疗数据共享，存在加密后文件所需存储空间增大，密钥的分发和管理复杂，代价高昂等问题。首先，在基于区块链的医疗数据共享过程中，参与方越多，需要保存的密钥个数越多，如果参与方数为 n，那么需要 n(n-1)/2 个密钥，当参与方越多，密钥的分配和保存成为很大的问题。其次，使用 AES-128 不能进行数字签名，无法满足医疗数据共享过程中身份认证的作用。最后，与使用 RSA 加密相比，使用 AES-128 加密后，医疗数据所需存储空间增大。

表5.6 非对称加密医疗数据共享耗时

数据量	加密耗时	解密耗时	加密后文件大小
10KB	123ms	126ms	13KB
12KB	238ms	243ms	17KB
118KB	294ms	1200ms	139KB
1063KB	1811ms	10930ms	1569KB
1170KB	1238ms	9387ms	1810KB

表5.7 对称加密医疗数据共享耗时

数据量	加密耗时	解密耗时	加密后文件大小
10KB	54ms	22ms	16KB
12KB	47ms	14ms	23KB
118KB	137ms	83ms	160KB
1063KB	203ms	111ms	1810KB
1170KB	508ms	237ms	2130KB

5.4.3 交易隐私保护分析

在 4.3 小节为了保护交易隐私，研究实现了结合超级账本 Fabric 的基于 Paillier 的群签名方案，在系统设计中使用该群签名方案初始化系统。在这里，以患者客户端为例，首先形成群组织，公开公钥，选出群管理员。然后将结合该群签名方案的 CA 集成到数据共享的 Fabric 网络中，患者使用 CA Client 加入网络，提交交易，验证身份，执行链码，得到结果。

在交易隐私保护方案中，使用了对交易数据加密然后再使用群签名，所以在文中对交易账本加密测试。超级账本 Fabric 中提供过了加密模块，该加密模块中默认实现了 ECDSA 系列签名算法，AES128、AES256 的对称加密算法和哈希算法等。所以本文中直接使用提供的 AES128 对交易提案加密，同时使用授权背书节点的身份证书对密钥进行加密。

使用 cat 命令来看区块的信息，数据明文如图 5.10 所示。

图5.10 加密前交易数据

加密之后的区块信息如图 5.11 所示。

图5.11 加密后交易数据

对比发现原来区块中明文信息变成了密文信息，可证明交易隐私保护方案有效。

超级账本 Fabric 中使用 MSP 管理用户成员身份，利用双层 PKI 体系来发放和管理用户证书，其中证书默认使用签名算法 ECDSA。所以将交易隐私保护方案中群签名与超级账本 Fabric 证书默认签名从签名计算量和验证计算量进行比较分析。如表 5.8 所示，对比可以看出两种签名方案性能相近。

表5.8 文中签名方案与 ECDSA-256 性能对比

算法	签名性能（次/秒）	验签性能（次/秒）
本签名方案	2013	628
ECDSA-256	2708	969

5.5 本章小结

本小节将医疗数据的隐私保护机制和交易数据的隐私保护机制应用于医疗数据共享，给出了基于超级帐本 Fabric 的医疗数据共享的具体过程，证实了方案的正确性，并且从身份验证、访问控制、隐私保护等方面对医疗数据共享过程进行安全性分析。最后对文中隐私保护机制进行了测试和验证。搭建超级账本 Fabric 环境进行医疗数据共享，统计分析医疗数据加密隐私保护方案中交易隐私保护机制中对交易数据加密功能进行功能性测试，并且将交易隐私保护方案群签名与超级账本 Fabric 证书默认签名算法 ECDSA-256 从签名计算量和验证计算量进行比较分析。

第六章 总结与展望

6.1 工作总结

本文首先研究分析了常见医疗数据共享模型，发现了医疗数据共享难，医疗数据提供者无法参与医疗数据共享过程，同时参与方增多，导致效率低、互操作性和稳定性差等问题，基于区块链进行医疗数据共享。同时从应用层和交易层对共享中隐私保护问题进行研究与实现，确保在医疗数据正常高效共享的前提下，达到隐私保护的目的。

为了在应用层达到隐私保护的目的，首先研究分析医疗数据，发现原始医疗数据所需存储空间较大，同时考虑到数据高效共享和区块链中区块容量有限，依据医疗数据标准对原始医疗数据进行预处理，使预处理后的医疗数据标准化和结构化；其次，构建基于默克尔树的关键字索引结构，在高效获取数据的同时验证访问者身份；最后，针对标准化、结构化医疗数据和关键字的索引树结构，研究实现了隐私保护算法，保证数据提供者对数据的完全控制和数据需求者对数据的安全访问，以起到医疗数据共享过程中隐私保护的目的。

为了从交易层起到隐私保护的作用，首先研究超级账本 Fabric 的底层架构，分析 Fabric 交易过程，发现 Fabric 在交易处理流程中的背书过程和排序过程环节可以改进以确保隐私性。其次，研究对比群签名方案，结合超级账本 Fabric 交易过程，使用基于 Paillier 的群签名方案，同时针对群签名方案中用户和管理员交互时身份信息易泄露的问题，结合超级账本 Fabric PKI 对交互中关键信息进行加密处理。最后，文中将群签名和对称加密融入超级账本 Fabric 的交易过程以确保医疗数据共享过程中交易数据隐私保护的目的。

最后，基于医疗数据隐私保护和交易数据隐私保护的研究，将隐私保护方案使用超级账本 Fabric 具体实施医疗数据共享过程，并进行测试验证。

6.2 工作展望

区块链的热潮吸引了众多行业的目光，随着区块链与行业的深度融合，具体业务场景千差万别，依据具体业务场景制定隐私保护机制是十分关键且重要的事。但是目前来说区块链中隐私保护机制的研究仍在初步测试阶段，还没有成熟且统一的隐私保护手段。文中从应用层将加密应用于医疗数据共享过程和交易层使用群签名和加密进行交易隐私保护两方面来进行隐私保护，测试分析证明有效，但是同时也证明文中所提隐私保护机制有不小的改进空间，主要包括：

（1）实现电子医疗数据的共享，对降低患者的就医成本和医疗机构的误诊率等都有着深远的意义。但是一些卫生系统以及医疗设备的 IT 供应商为了保护自身利益，拒绝实现医疗数据共享。所以需要有关政府部门采取强制措施，同时需要投入更多的精力确保数据共享过程的标准化，鼓励数据共享，确保整个共享过程更加地高效和透明。

（2）区块链中隐私保护问题十分重要，但是没有完美的隐私保护机制。文中针对医疗数据的隐私保护机制的加密环节使用了 RSA 算法，由于文中对数据进行了标准化处理，相比于处理之前，数据所需存储空间缩小很多，所以本文中总体数据量较小，为确保安全和隐私保护，应用层选择了 RSA-1024 实现，但是随着共享数据量的增大，该算法加密速率不理想。在后期研究中可以选择效率更好更快的公钥算法，例如运算速度快，资源消耗低的椭圆曲线算法。并且使用更多不同种类的医疗数据进行研究。同时，在以后的研究中，结合超级账本 Fabric 自身的基于属性的访问控制，可以更快地进行隐私保护。

（3）区块链非常适合解决医疗领域中医疗数据分散化带来的隐私泄露等诸多问题，但是随着科学技术的进步，区块链结合加密以实现隐私保护机制的安全级别越来越低，这也是今后研究的一个重点。

参考文献

[1] Mccarty J L, Gołofit P, Tigges S, et al. Digital Medical Illustration for the Radiologist.[J]. Radiographics, 2018:170088-.

[2] 薛艳杰. 探析信息技术在医院卫生经济管理中的应用[J]. 财经界, 2016(29).

[3] Nordin J D , Kasimow S , Levitt M J , et al. Bioterrorism Surveillance and Privacy: Intersection of HIPAA, the Common Rule, and Public Health Law[J]. American Journal of Public Health, 2008, 98(5):802-807.

[4] Lang L . HIPAA Privacy Rule and Negative Influence on Health Research[J]. Gastroenterology, 2008, 134(1):6-6.

[5] 郭珉江. 美国卫生部建议修改 HIPAA 隐私安全保密规定[J]. 医学信息学杂志, 2010(6):92-92.

[6] 崔光悦, 庞静, 姚俊鑫. 医疗信息交换平台(HIE)现状分析及展望[J]. 信息系统工程, 2018, No.293(5):136.

[7] Report P I. Benchmark study on patient privacy and data security.[J]. Journal of Healthcare Protection Management Publication of the International Association for Hospital Security, 2011, 27(1):69.

[8] Lincke S. Planning for Incident Response[M]// Security Planning. 2015.

[9] 焦仃. 2018 年医疗机构将面临网络安全威胁[J]. 计算机与网络, 2018(7).

[10] Funk E , Riddell J , Ankel F , et al. Blockchain Technology:, A Data Framework to Improve Validity, Trust, and Accountability of Information Exchange in Health Professions Education[J]. Academic Medicine, 2018, 93.

[11] Abdullah N , Hakansson A , Moradian E . Blockchain based approach to enhance big data authentication in distributed environment[C]// 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN). IEEE, 2017.

[12] Zelmer J, Hagens S. Advancing Primary Care Use of Electronic Medical Records in Canada[J]. Health Reform Observer - Observatoire des Réformes de Santé, 2014, 2(3).

[13] 李赞梅, 胡志民, 孙海霞, et al. 国家人口与健康科学数据共享平台资源综合评价指标构建研究[J]. 中国数字医学, 2018.

[14] 宁艳阳. 健康大数据迎来新局面[J]. 中国卫生, 2017(5):52-53.

[15] 邓明华. 代谢综合征风险预测问题评述[J]. 数学建模及其应用, 2017, 6(1):70-71.

[16] Azaria A , Ekblaw A , Vieira T , et al. MedRec: Using Blockchain for Medical Data Access and Permission Management[C]// 2016 2nd International Conference on Open and Big Data (OBD).

IEEE, 2016.

[17] Esposito C , Santis A D , Tortora G , et al. Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?[J]. IEEE Cloud Computing, 2018, 5(1):31-37.

[18] 赛迪顾问. 中国网络信息安全发展白皮书(2018)[N]. 中国计算机报,2018-09-24(008).

[19] Kuo T T , Ohno-Machado L . ModelChain: Decentralized Privacy-Preserving Healthcare Predictive Modeling Framework on Private Blockchain Networks[J]. 2018.

[20] Yue X , Wang H , Jin D , et al. Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control[J]. Journal of Medical Systems, 2016, 40(10).

[21] 陈敏, 牟海燕, 秦健. 健康医疗大数据标准体系框架研究[J]. 中国数字医学, 2018, 13(4):14-16.

[22] 任倬辉. 基于差分隐私保护的医疗数据分析系统的设计与实现[D]. 2018.

[23] 洪建, 李锐, 徐王权. 医疗健康数据隐私保护技术综述[J]. 中国数字医学, 2015(11).

[24] Jiang S , Cao J , Wu H , et al. BlockHIE: a BLOCkchain-based platform for Healthcare Information Exchange[C]// 4th IEEE International Conference on Smart Computing. IEEE, 2018.

[25] Peng Z , Jules W , Schmidt D C , et al. FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data[J]. Computational and Structural Biotechnology Journal, 2018, 16:267-278.

[26] 张圣垚. 基于区块链的电子病历系统的设计与实现[D].哈尔滨工业大学,2018.

[27] Gkoulalas-Divanis A , Loukides G , Sun J . Publishing data from electronic health records while preserving privacy: A survey of algorithms[J]. Journal of Biomedical Informatics, 2014, 50(Sp. Iss. SI):4-19.

[28] Hsin-Te W , Chun-Wei T . Toward Blockchains for Health-Care Systems: Applying the Bilinear Pairing Technology to Ensure Privacy Protection and Accuracy in Data Sharing[J]. IEEE Consumer Electronics Magazine, 2018, 7(4):65-71.

[29] Omar A A , Rahman M S , Basu A , et al. MediBchain: A Blockchain Based Privacy Preserving Platform for Healthcare Data[C]// International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage. Springer, Cham, 2017.

[30] Aiqing Z , Xiaodong L . Towards Secure and Privacy-Preserving Data Sharing in e-Health Systems via Consortium Blockchain[J]. Journal of Medical Systems, 2018, 42(8):140-.

[31] Rao P K S, Krishna D M, Ravi D. Multivariate Public Key Cryptography and Digital Signature[J]. 2018.

[32] Mironov I, Pandey O, Reingold O, et al. Incremental Deterministic Public-Key Encryption[J]. Journal of Cryptology, 2018, 31(1):134-161.

[33] Kumar V, Pandey P S, Ranjan P. A High-Throughput FPGA-Based Architecture for Adv anced Encryption Standard: AES-512 Using Pre-ciphered Lookup Table[M]// Intelligent C

ommunication, Control and Devices. 2018.

[34] Cao Y , Fu C . An Efficient Implementation of RSA Digital Signature Algorithm[C]// International Conference on Intelligent Computation Technology & Automation. IEEE Xplore, 2008.

[35] Singh G , Supriya, Singh G , et al. A Study of Encryption Algorithms (RSA, DES, 3DES and AES) for Information Security[J]. International Journal of Computer Applications, 2013, 67(19):33-38.

[36] Samy G N, Shanmugam B, Maarop N, et al. Information Security Risk Assessment Framework for Cloud Computing Environment Using Medical Research Design and Method[J]. Advanced Science Letters, 2018, 24(1):739-743.

[37] Bi W, Jia X, Zheng M. A Secure Multiple Elliptic Curves Digital Signature Algorithm for Blockchain[J]. 2018.

[38] Hertlein M, Manaras P, Pohlmann N. Smart Authentication, Identification and Digital Signatures as Foundation for the Next Generation of Eco Systems[M]// Digital Marketplaces Unleashed. 2018.

[39] Zhang H, An X B, Zhang C H, et al. High-efficiency quantum digital signature scheme for signing long messages[J]. Quantum Information Processing, 2019, 18(1):3.

[40] Chaum D, Heyst E V. Group Signatures[M]// Advances in Cryptology — EUROCRYPT '91. 1991.

[41] Hung D T, Minh N H, Hai N N. A Hybrid Threshold Group Signature Scheme with Distinguished Signing Authority[M]// Information Systems Design and Intelligent Applications. 2018.

[42] Eom S , Huh J H . Group signature with restrictive linkability: minimizing privacy exposure in ubiquitous environment[J]. Journal of Ambient Intelligence and Humanized Computing, 2018.

[43] Jiang S , Cao J , Wu H , et al. BlocHIE: A BLOCkchain-Based Platform for Healthcare Information Exchange[C]// 2018 IEEE International Conference on Smart Computing (SMARTCOMP). IEEE Computer Society, 2018.

[44] Androulaki E, Barger A, Bortnikov V, et al. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains[J]. 2018.

[45] Liang X , Zhao J , Shetty S , et al. Integrating Blockchain for Data Sharing and Collaboration in Mobile Healthcare Applications[C]// The 28th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE PIMRC 2017). IEEE, 2017.

[46] Hyv?Rinen H , Risius M , Friis G . A Blockchain-Based Approach Towards Overcoming Financial Fraud in Public Sector Services[J]. Business & Information Systems Engineering, 2017.

[47] Katuwal G J, Pandey S, Hennessey M, et al. Applications of Blockchain in Healthcare: Current Landscape & Challenges[J]. 2018.

[48] Zhang P, White J, Schmidt D C, et al. Applying Software Patterns to Address Interoperability in Blockchain-based Healthcare Apps[J]. 2017.

[49] Hussein A F , Arunkumar N , Gustavo R G , et al. A medical records managing and securing

blockchain based system supported by a Genetic Algorithm and Discrete Wavelet Transform[J]. Cognitive Systems Research, 2018, 52:1-11.

[50] Kamel B M N , Wilson J T , Clauson K A . Geospatial blockchain: promises, challenges, and scenarios in health and healthcare[J]. International Journal of Health Geographics, 2018, 17(1).

[51] Nagasubramanian G , Sakthivel R K , Patan R , et al. Securing e-health records using keyless signature infrastructure blockchain technology in the cloud[J]. Neural Computing and Applications, 2018(6).

[52] Heston T F . Why Blockchain Technology Is Important for Healthcare Professionals[J]. Social Science Electronic Publishing, 2017.

[53] Kakushadze Z , Russo , R P . Blockchain: Data Malls, Coin Economies and Keyless Payments[J]. Social Science Electronic Publishing.

[54] Watanabe H , Fujimura S , Nakadaira A , et al. [IEEE 2016 IEEE International Conference on Consumer Electronics (ICCE) - Las Vegas, NV, USA (2016.1.7-2016.1.11)] 2016 IEEE International Conference on Consumer Electronics (ICCE) - Blockchain contract: Securing a blockchain applied to smart contracts[J]. 2016:467-468.

[55] G?Bel J , Krzesinski A E , Keeler H P , et al. Bitcoin Blockchain?dynamics: The selfis h-mine strategy in the presence of propagation delay[J]. Performance Evaluation, 2016:S0 16653161630089X.

[56] Mettler M . Blockchain technology in healthcare: The revolution starts here[C]// 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom). IEEE, 2016.

[57] Gupta H, Hans S, Mehta S, et al. On Building Efficient Temporal Indexes on Hyperledger Fabric[J]. 2018:294-301.

[58] Zhou J , Tang F , Zhu H , et al. Distributed Data Vending on Blockchain[J]. 2018.

[59] Valenta L, Rowan B. Blindcoin: Blinded, Accountable Mixes for Bitcoin[J]. 2015.

[60] Esposito C, Castiglione A, Palmieri F, et al. Integrity for Event Notification Within Industrial Internet of Things by Using Group Signatures[J]. IEEE Transactions on Industrial Informatics, 2018, PP(99):1-1.

[61] 魏文燕. Paillier 同态密码在隐私保护中的应用研究[D].河南理工大学,2017.

[62] Man H A,Wei V K.ID-based Cryptography from Composite Degree Residuosity.[J].Iacr Cryptology Eprint Archive,2004,2004.

[63] Castelluccia C. How to convert any ID-based Signature Schemes into a Group Signature Scheme[J].HAL-INRIA,2006,2002.

[64] Sousa J, Bessani A, Vukolić M. A Byzantine Fault-Tolerant Ordering Service for the Hyperledger

Fabric Blockchain Platform[J]. 2017.

[65] Benhamouda F, Halevi S, Halevi T. Supporting Private Data on Hyperledger Fabric with Secure Multiparty Computation[C]// 2018 IEEE International Conference on Cloud Engineering (IC2E). 2018.

[66] Roman D , Stefano G . Towards a Reference Architecture for Trusted Data Marketplaces: The Credit Scoring Perspective[C]// 2nd International Conference on Open and Big Data (OBD), 2016. IEEE, 2016.

[67] Smith B K , Nachtmann H , Pohl E A . Improving Healthcare Supply Chain Processes Via Data Standardization[J]. Engineering Management Journal, 2012, 24(1):3-10.

[68] Hull R , Batra V S , Chen Y M , et al. Towards a Shared Ledger Business Collaboration Language Based on Data-Aware Processes[C]// International Conference on Service-Oriented Computing. Springer International Publishing, 2016.

[69] Ethereum whitepaper. A next-generation smart contract and decentralized application platform[EB/OL].[2018-04-10].https://github.com/ethereum/wiki/wiki/White-Paper.

[70] Hyperledger Fabric documentation, Getting started and developer guides. [EB/OL].[2018-04-10].http://hyperledger-fabric.readthedocs.io/en/latest/

致谢

不知不觉，三年的研究生时光已经进入了尾声，同时也意味着我的在校学生生涯就要暂时地划上句号。回想在西电七年的点点滴滴，仿佛发生在昨天一般。在这七年里，我学到了专业知识，结交了很多好朋友，遇到了非常多的、可爱的人和事。在此，我要向我生命中出现的所有的人和事奉上最真挚的感谢。

感谢我的导师覃桂敏老师，在和覃老师的相处中，我感受到了她对科研的认真负责，对学生的细心耐心。覃老师不仅在学业上给予悉心的指导，在我论文撰写困难时给予帮助，还在生活中及时地提醒我注意相关通知。从老师那里学到的专业知识和为人处世的道理，都将在我日后的学习生活工作中非常受用。

感谢实验室的刘西洋老师和王黎明老师，研究生期间开展的科研项目是在两位老师的认真指导下完成，刘老师的学识渊博和王老师的视野广阔、一丝不苟都对我留下了深深的印象，在论文的撰写和修改过程中，两位老师也给予了悉心的指导。同样感谢实验室的霍秋艳老师、高海昌老师和范磊老师，为我们营造了良好的学术氛围。在此，衷心地祝福所有的老师工作顺利，身体健康！

感谢实验室所有的师兄师姐师弟师妹和我的舍友们。刚进实验室，杜婷师姐带领我开展科研项目，在师姐的帮助下，我的技术水平有了很大的提升。同时，我也在张馨月师姐、邓钊师兄、雷倩师姐、李玲红师姐、王晓东师兄、李佳伟师兄、韩佳浩师兄等师兄师姐身上学到了很多，他们不仅在我平时学习生活中给予帮助，还在找工作时给予指导。感谢实验室的小伙伴们，陈敏、杨璐琼、王帅、孙赛、李琳、杨军、杨俊、杨洲、原富强，我们一起进行科研项目，一起出去玩，和你们在一起的日子会成为我生命中浓墨重彩的一笔。

最后，我要特别感谢我的家人。我的父母不仅在物质上一直支撑着我，还不断地鼓励我去尝试、要果敢自信，在我遇到问题时耐心开解，让我一步一步变得更好。衷心地祝福所有人平安喜乐，诸事顺遂！

作者简介

1. 基本情况

王旭，女，山西运城人，1994 年 8 月出生，西安电子科技大学计算机科学与技术学院软件工程 2016 级硕士研究生。

2. 教育背景

2012.09 ～ 2016.07 西安电子科技大学，本科，专业：软件工程
2016.09 ～ 西安电子科技大学，硕士研究生，专业：软件工程

3. 攻读硕士学位期间的研究成果

3.1 参与科研项目及获奖

[1] 越秀区健康素养平台, 2016.10~2017.09, 已上线, 项目核心开发人员.
[2] 智能红娘平台, 2016.10~2017.06, 已上线, 项目核心开发成员.

西安電子科技大學
XIDIAN UNIVERSITY

地址：西安市太白南路2号
邮编：710071
网址：www.xidian.edu.cn