

电子科技大学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

专业学位硕士学位论文

MASTER THESIS FOR PROFESSIONAL DEGREE



论文题目 基于区块链的电子健康记录隐私保护方案

研究与实现

专业学位类别 工 程 硕 士

学 号 201722220433

作 者 姓 名 王 靖

指 导 教 师 曹 晟 副教授

分类号_____密级_____

UDC^{注1}_____

学 位 论 文

基于区块链的电子健康记录隐私保护方案研究与实现

(题名和副题名)

王 靖

(作者姓名)

指导教师

曹 晟

副教授

电子科技大学

成 都

(姓名、职称、单位名称)

申请学位级别 硕士

专业学位类别 工 程 硕 士

工程领域名称

软件工程

提交论文日期 2020.03

论文答辩日期 2020.05

学位授予单位和日期

电子科技大学

2020 年 6 月

答辩委员会主席

评阅人

注 1: 注明《国际十进分类法 UDC》的类号。

Research and Implementation of Privacy Preservation Scheme of Electronic Health Records Based on Blockchain

A Master Thesis Submitted to
University of Electronic Science and Technology of China

Discipline: **Master of Engineering**

Author: **Wang Jing**

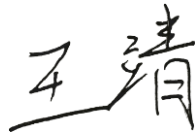
Supervisor: **Prof. Cao Sheng**

School: **School of Information and Software Engineering**

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，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。与我一同工作的同志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谢意。

作者签名：



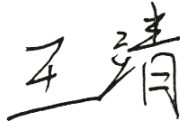
日期：2020 年 5 月 26 日

论文使用授权

本学位论文作者完全了解电子科技大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权电子科技大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后应遵守此规定）

作者签名：



导师签名：



日期：2020 年 5 月 26 日

摘 要

电子健康记录（Electronic Health Records, 简称 EHR）已被广泛应用于基于云的医疗平台中，包含患者的隐私数据，患者和医院分别是电子健康记录的所有者和管理者。因此，电子健康记录所有权与管理权的分离，导致其存储在医疗云平台中时存在泄漏和滥用的风险，探索电子健康记录隐私保护新方案就显得尤为重要。区块链技术以其去中心化、不可篡改、可追溯等特性，可在电子健康记录隐私保护中起到重要作用。

本文主要做了以下工作：

（1）研究了电子健康记录的应用场景，梳理了区块链用于电子健康记录隐私保护的主要方法，提炼出电子健康记录在不同医院之间流转对区块链跨链技术的需求，对跨链场景中电子健康记录涉及的角色和功能进行了分析与设计，提出了基于 Polkadot 中继链的电子健康记录跨链架构。

（2）提出了一种基于区块链跨链技术的电子健康记录隐私保护方案，简称 CEPS。给出了基于 Polkadot 的电子健康记录跨链流程，从 51% 攻击、不诚实节点和恶意共谋等三个方面论证了方案的安全性。采用 Substrate 区块链开发框架，运用 Node.JS、React、WebAssembly 等技术，从应用层、数据层和区块链层等三个层面实现了 CEPS 方案跨链系统。

（3）展示了电子健康记录跨链系统主要界面，搭建了 Substrate 区块链测试环境，对电子健康记录账户和区块生成、增加和删除记录等主要功能模块进行了测试，从出块速度、计算开销两个方面测试了系统性能，验证了本文提出的方案对电子健康记录隐私保护的有效性。

本文将区块链跨链技术应用于电子健康记录隐私保护，实现了患者电子健康记录在不同医院之间的可信共享，提升了患者就医效率，对促进医疗领域以医院为中心向以患者为中心的转变，具有重要的推动作用。

关键词：电子健康记录，隐私保护，区块链，跨链技术，Polkadot

ABSTRACT

Electronic Health Records (EHR) have been widely used in cloud-based medical platforms, which includes patient's privacy data. Patients and hospitals are the owners and managers of the electronic health records, respectively. Therefore, the separation of ownership and management of electronic health records leads to the risk of leakage and abuse when electronic health records are stored in the cloud-based medical platforms. Thus, it is particularly important to explore new privacy preservation schemes for electronic health records storage. Due to the features of decentralization, immutability, and traceability, blockchain technology is a promising way to play an important role in the privacy preservation of electronic health records.

The contributions of this thesis are listed as follows:

(1) Describe the application scenarios of electronic health records, investigate the state-of-the-art methods applying blockchain technology to protect the privacy of electronic health records, and refine the demand of cross-chain technology for the circulation of electronic health records between different hospitals. The roles and functions involved in electronic health records in cross-chain scenarios are analyzed and designed, and a cross-chain electronic health records architecture based on the Polkadot relay chain is proposed.

(2) An electronic health record privacy preservation scheme based on blockchain cross-chain technology (CEPS) is proposed. An electronic health record cross-chain process based on Polkadot is given. Security analysis of the scheme is demonstrated from three aspects, including 51% attacks, dishonest nodes, and malicious collusion. We adopt Substrate as the blockchain development framework, combine Node.JS, React and WebAssembly technologies together, then implement the CEPS solution cross-chain system from three levels: the application layer, the data layer, and the blockchain layer.

(3) The layout of the electronic health record cross-chain system was demonstrated. We establish the Substrate blockchain test environment, carry out the functional testing on the generation of electronic health record account and block, as well as the addition and deletion of records. The performance is tested in two aspects, namely, block production speed and communication costs. And the effectiveness of the proposed scheme for the privacy preservation of electronic health records is verified.

ABSTRACT

This thesis applies the cross-chain technology to the privacy preservation of electronic health records, realizing trusted sharing of patient electronic health records across different hospitals. Then the patient medical treatment could be more cost-effective. It will be beneficial to promoting the variation from hospital-centric to patient-centric in the medical field.

Keywords: Electronic Health Records (EHR), privacy preservation, blockchain, cross-chain, polkadot

目 录

第一章 绪 论	1
1.1 背景与意义	1
1.2 国内外研究现状	2
1.2.1 电子健康记录隐私保护	2
1.2.2 区块链技术	3
1.2.3 区块链与电子健康记录	5
1.3 本文组织结构	6
第二章 相关技术研究	8
2.1 电子健康记录	8
2.2 区块链技术	9
2.2.1 区块链特点	9
2.2.2 区块链层次架构	10
2.2.3 时间戳	11
2.2.4 哈希函数	12
2.2.5 Merkle 树	13
2.3 区块链跨链技术	13
2.4 Polkadot 中继链	15
2.5 Substrate 开发框架	16
2.6 本章小结	18
第三章 基于区块链的电子健康记录隐私保护需求分析	19
3.1 区块链用于电子健康记录	19
3.2 需求分析	20
3.2.1 角色分析	21
3.2.2 功能分析	21
3.2.3 跨链流程分析	23
3.3 本章小结	23
第四章 基于区块链的电子健康记录隐私保护方案	24
4.1 方案概述	24
4.2 基于 Substrate 的电子健康记录隐私保护方案	26
4.2.1 方案概述	26

4.2.2 方案流程	28
4.2.3 电子健康记录脱敏设计	29
4.2.4 电子健康记录索引	31
4.3 基于 Polkadot 的电子健康记录跨链方案	32
4.3.1 基于 Polkadot 的医疗区块链隐私保护方案流程	32
4.3.2 Polkadot 角色	33
4.3.3 Polkadot 跨链	34
4.3.4 算法设计	37
4.4 安全性分析	38
4.4.1 51%攻击	38
4.4.2 不诚实节点	38
4.4.3 恶意共谋	38
4.5 花费分析	40
4.6 本章小结	40
第五章 系统关键功能实现与测试	41
5.1 整体架构	41
5.2 关键功能实现	42
5.2.1 开发环境	42
5.2.2 基于 Substrate 的区块链实现	43
5.2.3 电子健康记录操作实现	46
5.2.4 数据脱敏实现	52
5.2.5 跨链实现	53
5.3 测试	55
5.3.1 测试环境	55
5.3.2 功能测试	55
5.3.3 性能测试	61
5.4 本章小结	64
第六章 总结与展望	65
6.1 总结	65
6.2 展望	65
致 谢	67
参考文献	68
攻读硕士学位期间取得的成果	71

第一章 绪 论

1.1 背景与意义

电子健康记录(Electronic Health Records, 简称 EHR)^[1]是将个人的医疗信息电子化存储, 包括个人的身份信息、诊断信息、病历信息、用药信息、治疗信息和住院信息等。相较于传统的纸质健康记录有易存储、易操作、易获取的特性, 正是由于这些特性, 使得电子健康记录容易被攻击, 这些攻击包括黑客窃取、医生篡改等。同时随着云计算, 大数据等技术的飞速发展, 医疗行业也在逐渐融合信息技术, 趋向于数字化。然而信息技术在医疗行业的使用也势必带来信息安全风险。2017 年, 据《法制日报》报道, 某部委医疗服务信息系统遭黑客入侵, 超过 7 亿条公民信息遭泄露, 8000 余万条公民信息被贩卖; 2018 年全年美国共发生 18 起涉及医疗记录的数据泄露事件, 被泄露的医疗记录数量超过 10 万份; 2019 年, 外媒 Securityaffairs 报道, 几百个未受保护的服务器暴露于互联网, 这些服务器中包含大量医疗放射图像, 其中有超过 7.37 亿个放射图像, 涉及 2000 多万人和 52 个国家的患者。由此可见, 提高医疗行业信息系统的安全性迫在眉睫。

随着以比特币为代表的数字货币不断发展, 其底层区块链技术也越来越受到人们的关注。在我国, 2018 年由中国工业和信息化部信息中心发布的《2018 中国区块链产业白皮书》^[2]中指出, 目前我国区块链产业处于高速发展早期阶段, 区块链应用呈现多元化, 从金融延伸到实体领域, 医疗领域正是区块链可以拓展的一个重要领域。区块链整合现有技术, 将数据构造成区块, 并以链式结构按时间顺序连接, 后一个区块中存储前一个区块的哈希值来达到连接的目的, 类似链表。这样的结构造就了区块链上的数据不可篡改, 可溯源的特性, 十分契合医疗场景。然而区块链结合医疗领域的研究相对较少, 方向主要有医疗信息保护、医疗支付、医疗数据应用、医疗数据存储分享、医疗信息交易、预测分析等。IBM 在 2017 年发布了名为《医疗保健业集结于区块链—以患者为中心》的报告, 报告中称国外医疗保健组织已经采用区块链技术, 区块链技术会在临床试验记录、监管合规性和健康监控记录领域发挥巨大价值。同时区块链在健康管理、药物治疗、计费 and 理赔、医疗资产和合同管理等方面都能充分发挥作用。

目前, 将区块链用于电子健康记录隐私保护的方法已经较多, 可以有效保护电子健康记录的隐私, 这些方法针对的都是单个区块链上的隐私保护, 没有考虑电子健康记录在不同的区块链之间流转时的隐私保护需求。比如一个患者在医院 A 就诊, 当需要转院到医院 B 时, 很多情况下需要在医院 B 重新进行各项检查, 既

增加了开销，又延误了就诊时间。如果患者到了医院 B 就诊之后，能直接调用医院 A 的诊断记录，则在减少患者花费的同时又能为患者节约时间，还促进了不同医院间的数据共享，为了达到这种目的，就需要使用区块链跨链技术。

想象一个场景，每个医院都构建自己的私有区块链网络（基于以太坊或比特币等），患者先在医院 A 就诊，后面因病情或其他原因，需转移到医院 B 就诊。此时，医院 B 可以通过区块链跨链技术调用存储在医院 A 中的就诊病例（电子健康记录），从而加快就诊速度，减少患者就诊开销。

然而不同的区块链网络相对独立，不互通，像是一个个独立的“局域网”，没有连接成“互联网”。比如医院 A 和医院 B 分别有各自的私有区块链，采用不同的区块链构建医疗系统存储电子健康记录，就会在整个医疗行业中形成一个个的信息孤岛，不同医疗机构间的信息难以通过不同的区块链互联互通，这就需要跨链技术将不同医院的私有区块链连通起来。

目前，区块链跨链已成为一个新的研究热点，可以将区块链互联互通起来，使得不同的区块链也能信息交换，形成真正的区块链“互联网”。为了实现电子健康记录隐私保护的目的，本文提出了一种基于区块链的电子健康记录隐私保护方案，该方案采用波卡币的 Polkadot 中继链思想，使用 Substrate 区块链开发框架实现了医疗区块链的跨链系统，使得电子健康记录数据可以在不同医院的私有区块链间流转，同时能使患者通过区块链得知自己的电子健康记录的流向并可对电子健康记录进行操作，构建以患者为中心的医疗系统。

本文提出的基于区块链的电子健康记录隐私保护方案，将区块链技术与电子健康记录结合，不仅保护了患者的隐私，还解决了医疗领域的信息孤岛问题，实现了以医院为中心向以患者为中心的转变。

1.2 国内外研究现状

1.2.1 电子健康记录隐私保护

S. Chenthara 等人介绍了关于电子健康记录的网络安全的研究挑战 and 方向^[3]，他们指出在私有、公共或混合云环境中，确保电子医疗数据的完整性，机密性，可靠性以及真实性非常重要，并提出了电子医疗系统中安全性和隐私性要求，最后还总结基于密码学和非密码学方法的电子健康记录隐私保护技术。

Qi Wang 等人结合同态加密和代理重新加密技术来实现医疗保健系统中的外包计算方案^[4]，在该方案中有几个具有不同公钥的客户端，一个电子医疗云平台和辅助云服务器。电子医疗云平台可以为患者提供服务，并定期分析数据以提供更好的服务。

A. Sahi 等人指出许多医疗保健行业的公司逐渐将电子数据移至云中^[5], 减少内部存储, 这可能会导致严重的存储, 安全和隐私问题, 他们提出了两种方法, 即安全保护方法和隐私保护方法, 以及灾难恢复计划。安全保护方法是为了确保电子健康记录的安全性和完整性, 而隐私保护方法是保护个人健康记录隐私的有效身份验证方法。

陈虹云等人为解决医疗数据匿名发布的同步问题, 提出了一种建立在 (α, k) -匿名数据基础上的支持数据动态更新的算法^[6], 该算法通过对语义贴近度的计算, 在 (α, k) -匿名数据集中选择最贴近的等价类, 再进行相应的更新操作, 更新后的匿名数据集满足 (α, k) -匿名约束, 可有效地保护患者的隐私信息。

由此可见, 目前电子健康记录的隐私保护大多是基于密码学技术的研究, 但随着区块链技术的发展, 使得基于区块链的电子健康记录的隐私保护

1.2.2 区块链技术

从 2008 年中本聪发表论文《Bitcoin: A Peer-to-Peer Electronic Cash System》^[7]以来, 越来越多的学者、研究人员投入到区块链技术的研究中。区块链被誉为是计算范式的第五次颠覆式创新, 是人类信用进化史上第四个里程碑。区块链作为前沿技术, 各国也陆续在区块链领域加大了投入^[8]。

2016 年 10 月工信部发布了《中国区块链技术和应用发展白皮书(2016)》, 总结了国内外区块链发展现状以及区块链典型应用场景, 介绍了国内区块链发展路线, 以及未来国内区块链发展的方向; 2017 年 1 月, 工信部发布了《软件和信息技术服务业发展规划》, 提出区块链领域积极创新以及达到国际先进水平等要求; 2018 年 3 月, 工信部又发布了《2018 年信息化和软件服务业标准化服务要点》, 提出推动组建全国区块链和分布式记账技术标准化委员会等。2019 年 12 月由工信部指导出品的《2019 中国区块链底层技术平台发展报告》, 从基础技术能力、应用能力以及社会认可度三个维度对国内多个区块链技术平台进行了评估。

区块链科学研究所(Institute for Blockchain Studies)创始人 Melanie Swan 在著作《Blockchain: Blueprint for a New Economy》^[9]中将区块链分为了 1.0、2.0 和 3.0 三个阶段。在区块链 1.0 阶段, 区块链应用仅仅还是货币和支付系统; 区块链 2.0 是货币市场以及金融应用; 3.0 将区块链扩展到金融以外的一些应用, 如慈善、个人健康、公共卫生等。

袁勇, 王飞跃等人对区块链技术发展的现状进行了分析并展望了区块链未来的发展^[10]。他们根据区块链技术应用的现状, 将区块链目前的主要应用归纳为数字货币、数据存储、数据鉴证、金融交易、资产管理和选举投票六个场景, 概述了除数字货币外的五大应用场景, 并分析了区块链存在的安全问题、效率问题、

资源问题和博弈问题。

韩璇等人从区块链安全角度出发，对区块链安全进行了缜密的分析^[11]。他们分析了区块链的安全目标，并将区块链体系架构分为六个层次，对每个层次区块链所面临的安全性问题进行了描述，同时提出了未来区块链在安全方面应当研究的重点。

区块链发展中最重要一个方向是区块链跨链技术。跨链技术可以将不同的区块链网络互连，实现真正的价值互联网。

路爱同等人对区块链主流跨链技术进行了总结^[12]，介绍了区块链跨链技术的特性以及技术难点，并提出了参考解决方案；李芳等人分析了跨链技术的需求及面临的技术难点^[13]，总结了正在发展的跨链技术，并介绍了主流跨链技术的原理与实现思路，综合分析了跨链技术存在的安全性风险，最后总结探讨了跨链技术的未来发展趋势；魏昂基于哈希锁定技术，设计了一种改进的区块链跨链技术^[14]，该技术通过将主链资产双向锚定到侧链，并利用哈希时间契约锁协议防止信息篡改，限制交易时间，设置违约强制保障协议，实现资产跨链交易。

Michael Borkowski 等人提出了名为 Dextt 的跨区块链传输协议^[15]，该协议可用于以分散方式同时记录任意数量的区块链上的代币传输，它可以在任何数量的区块链上使用，并且其交易以分散的方式在区块链之间自动同步；以太坊创始人 Vitalik Buterin 对区块链跨链技术进行的详细的阐述^[16]，将跨链技术分为了七类并详细描述了主要的三类跨链技术的技术细节，分别是公证人技术、哈希锁定技术和侧链/中继链技术，并对这三类跨链技术从不同的角度进行了比较。

其中中继链跨链技术中，较为突出的三个技术是 BTC Relay、Cosmos 和 Polkadot。三者各有优劣，各有适用的场景。三种技术的特征比较如表 1-1 所示。

表 1-1 中继链跨链技术比较

	BTC Relay	Cosmos	Polkadot
架构	简单	较复杂	复杂
扩展性	弱	强	中
灵活性	弱	强	中
安全性	弱	中	强

BTC Relay 是由 ConsenSys 团队推出的被认为是区块链上的第一个侧链^[17]，其主要原理是 BTC Relay 把以太坊网络与比特币网络以一种安全去中心化的方式连接起来。BTC Relay 通过使用以太坊的智能合约功能可以允许用户在以太坊区块链上验证比特币交易。正是由于 BTC Relay 的原理限制了其应用场景，只能在比特

币区块链网络和以太坊区块链网络间进行价值交换。

Cosmos 是 Tendermint 团队推出的一个支持跨链交互的异构网络^[18]，它最终的目标是创建一个区块链互联网，允许大量自主且易开发的区块链互相扩展和交互。基于 Tendermint 开发，采用的 Tendermint 共识算法，是一个类似实用拜占庭容错共识引擎，具有高性能、一致性、具备拜占庭容错等特点。

Cosmos 与 Polkadot 的架构类似，但是各有侧重^[19]。在架构上 Polkadot 较于 Cosmos 相对更复杂，在医疗场景下相对复杂的架构能更好的保障系统的稳定性；在安全性上，因为 Cosmos 有较为简洁的架构，所以在一定程度上安全性略有下降，而 Polkadot 较为复杂的架构也带来了更高的安全性，这在电子健康记录的隐私保护中十分重要。灵活性方面，Cosmos 具有较低的接入门槛，但是接入 Cosmos 的区块链需要自行负责其安全性，而 Polkadot 接入门槛更高，但是 Polkadot 也为接入的平行链提供了安全性保障。

在医疗场景下，使用区块链首先需要较高的安全性以保证病人的电子健康病历的安全，不被侵犯隐私，为了在医疗区块链间跨链，又需要一定的扩展性和灵活性以实现不同医院间电子健康记录的跨链，但是过于灵活会导致整个医疗跨链架构过于庞大，一些不需要的区块链也可能接入整个架构中，所以在三种主流中继链技术中，Polkadot 最适合本论文的研究。

1.2.3 区块链与电子健康记录

近年来，关于基于区块链的电子健康记录隐私保护的研究很多，并证明使用区块链技术保护和管理云存储中的电子健康记录数据具有巨大的潜力和价值^[20]。

薛腾飞等人提出了一个基于区块链的医疗数据共享模型^[21]，该模型具有去中心化，安全可信，集体维护和不可篡改的特性，适用于解决各医疗机构数据共享的难题。徐文玉等人提出了基于区块链和同态加密的电子健康记录隐私保护方案^[22]，该方案解决了电子健康记录隐私保护问题、患者与其他角色交互时的安全问题，以及保险公司侵犯患者隐私的问题。Jingwei Liu 等人提出了一种基于区块链的电子病历隐私保护数据共享方案^[23]。在该方案中，原始电子病历安全地存储在云中，索引存储在防篡改联盟区块链中。Hao Wang 等人提出了将基于身份的属性加密与区块链相结合的方案，以确保电子健康记录数据的完整性和可追溯性^[24]。S. Cao 等人提出了一个基于以太坊的医疗数据共享模型^[25]，该模型具有去中心化，安全可信，集体维护和不可篡改的特性，适用于解决各医疗机构数据共享的难题。

Drew Ivan 讨论了区块链作为一种保护健康数据存储实施障碍的新方法^[26]，以及从现有技术向区块链解决方案逐步过渡的计划。Asaph Azaria 等人提出了一种基

于区块链技术处理电子病历的分散式医疗数据管理系统^[27]。该系统利用区块链的属性、身份验证、加密机制、问责机制进行了模块化设计，提升了互操作性和适应性。Vidhya Ramani 等人将区块链技术视为一种分布式方法来保护医疗保健系统中的数据^[28]，并为医疗系统中的患者和医生提出了一种基于区块链的安全有效的数据访问机制，同时能够有效保护患者的隐私。Marcela T.等人提出了一种基于区块链的方法来确保电子健康记录的安全^[29]，其中访问控制以患者为中心并将加密的电子健康记录保存在区块链中，患者仅与他/她信任的医疗专业人员共享解密密钥，该方案无需任何可靠的中介并且具有很好的扩展性。

Christian Esposito 等人指出，区块链在电子健康记录的应用中存在两个主要问题^[30]，一个是它不能在区块链中存储大量的电子健康记录数据。另一个是无法删除存储在区块链中的任何数据。电子健康记录中包含大量的患者隐私数据，这些隐私数据将受到隐私法的保护，其中许多法律不允许将个人数据永久存储^[31]。因此，基于区块链的电子健康记录隐私保护方案应赋予患者删除私人数据的权利。

以上研究均未解决在不同区块链之间传输时患者的电子健康记录数据的隐私保护问题。对于前述的患者从医院 A 转到医院 B 就诊的场景，目前的研究并未解决这类问题。为了访问医院 A 中患者的电子健康记录数据，需要使用跨区块链技术。因此，本文将在区块链跨链技术和区块链数据删除两个方面提出新颖的解决方案。

1.3 本文组织结构

第一章：绪论。阐述了研究工作的背景及意义，介绍了目前电子健康记录隐私保护存在的问题，描述了区块链技术，以及国内外将医疗与区块链相结合的研究现状，指出利用区块链跨链技术可以使电子健康记录在不同的医疗机构中流转共享。最后对本文的研究内容与结构进行了阐述。

第二章：相关技术研究。介绍了电子健康记录的基本概念和特点；接着介绍了 Merkle 树、哈希函数和区块链层次架构等相关技术，对区块链跨链技术中的中继链 Polkadot 进行了描述，并对由 Polkadot 发展出的区块链开发框架 Substrate 进行了阐述。

第三章：基于区块链的电子健康记录隐私保护需求分析。分析了传统电子健康记录系统的结构，描述了传统电子健康记录在数据安全、数据追溯和数据共享等领域存在的问题，对本文提出的基于区块链的电子健康记录隐私保护方案的角色进行了划分，对方案进行了功能模块和跨链流程分析。

第四章：基于区块链的电子健康记录隐私保护方案。依据需求分析，提出了

基于区块链跨链技术的电子健康记录隐私保护方案，该方案利用 Substrate 区块链开发框架构建了医疗区块链，并将医疗区块链作为 Polkadot 中继链的平行链，接入 Polkadot 网络，实现电子健康记录在不同的医疗区块链间跨链流转。

第五章：系统关键功能实现与测试。对基于区块链的电子健康记录隐私保护方案的关键功能进行了实现，并对方案进行了功能测试、性能测试和安全分析。

第六章：总结与展望。总结论文中涉及的工作，提出本文的不足以及未来可能的研究方向。

第二章 相关技术研究

本章将对论文所使用的相关技术进行介绍，包括电子健康记录、区块链及跨链技术、Substrate 区块链开发框架。

2.1 电子健康记录

随着信息技术的发展，医疗记录已从手写记录格式转换为电子健康记录格式。电子健康记录是以个人健康、保健和治疗为主要内容的数字记录，是以人为本的数字化健康档案，可以反映患者在一段时间内的临床诊断过程记录。它以数字化方式管理患者的健康状态和医疗保健信息，涉及患者信息的采集、存储、传输、处理和使用。

电子健康记录的发展历经的三个阶段：第一阶段是计算机化病案，是指医院在病人出院后，通过数码拍照或扫描方式将整本病案输入，集文字、表格、图像于一体，把病人的基本信息、临床诊断、用药、治疗方案、手术处置等信息综合在电脑上，单纯地由纸质病案向电子媒体移植，形成围绕病人的多媒体综合信息；第二阶段为电子病历，是医院信息化发展的趋势，包含了病人整个医疗过程，储存了病人全部的医疗信息，包括病史、各种检验、检查和影像资料以及医嘱改动的情况，是对个人医疗信息及其相关处理过程综合化的体现；第三阶段为电子健康记录，是指对于健康相关活动的电子化记录，不仅包括人们接收医疗服务的记录，还包括疫苗接种、接受保健服务、参与健康教育活动等的记录，是深度数字化的、上下文关联的病人终身医疗记录，覆盖个人从出生到去世的整个生命周期，强调个人信息。

电子健康记录的发展，旨在通过医疗卫生信息的高度共享，达到以下三个方面的目的：第一，减少并最终杜绝医疗差错；第二，解决医疗过度问题；第三，降低医疗成本，提高医疗水平。

患者在就医过程中，医务人员通过使用医院的信息系统生成文字信息、符号信息、表格信息、图像信息、视频信息等各种数字化信息，并对这些数字化信息进行存储、管理、传输和重现。电子健康记录是医疗信息的一种记录形式，而非狭义的纸质病历的电子化，其结构如图 2-1 所示。随着电子健康记录的普及，各大医院都建立了自己的电子健康记录，既可以帮助医生快速了解患者的病史，也可以帮助患者记录自身身体状况。

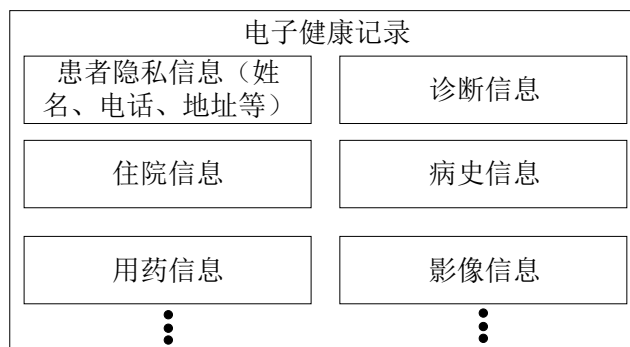


图 2-1 电子健康记录结构

2.2 区块链技术

2.2.1 区块链特点

区块链起源于 2008 年一个名为“中本聪”的人发表的《Bitcoin: A Peer-to-Peer Electronic Cash System》中，该文阐述了基于 P2P 网络技术、加密技术、时间戳技术、区块链技术等电子现金系统构架理念，引出了比特币这一虚拟货币。区块链正是这一虚拟货币的底层存储结构。区块链本质上一个共享的、去中心化的、分布式的数据库。区块链整合了哈希函数、时间戳技术、Merkle 树、共识算法等技术，再结合对称加密、非对称加密、数字签名、零知识证明等密码学技术，构建了一种全新的分布式架构。参与区块链的节点，把想要存储在区块链上的数据打包成区块，通过共识机制，将自己的区块竞争上链，并包含上一个区块的哈希值，以此将所有的区块串联起来，形成一条以时间为顺序的链式数据结构。

区块链本身的技术特点和结构特点使得区块链具备了多样特性，包含去中心化、不可篡改、可追溯、匿名、开放、容灾性等^[32]。

（1）去中心化。本质上讲，区块链是一种分布式数据库，区块链上的所有操作需要加入区块链网络的所有节点参与，不同于中心化的系统，区块链依靠共识算法建立的共识机制来达到可信的目的，避免了中心化系统使用的中心化服务器发生单点故障的问题。在区块链中，任何节点的交易不需要第三方机构颁发的证书，所有节点的权利对等。同时任何节点的加入或退出区块链网络都不会影响整个区块链网络的运行。

（2）不可篡改。通过验证并达成共识的数据将会存储在区块链上，并且所有参与区块链网络的节点共同维护区块链上的数据。若要篡改区块内的信息，必须控制网络中至少 51% 的节点来对这次修改达成共识，否则单个或几个节点的修改是完全无效的或者导致区块链分叉，因此存储在区块链中的数据极具可靠性。

（3）可追溯。在区块链中，所有的区块都包含当前区块上链的时间戳，这些

时间戳都唯一，且每个区块包含前一个区块的哈希值，这就使得每一个区块独一无二且所有区块按时间顺序有序排列，数据就有了合理的溯源路径。

（4）匿名。在区块链网络中任何用户在交易过程中都不会暴露身份，除了账户地址，交易中的所有数据都不包含任何与用户个人信息有关的数据。

（5）开放。区块链网络中除了包含交易的私有信息不会公开外，链上的所有数据对参与区块链网络的人来说都是开放的，任何人都可以在区块链上进行查询数据、进一步开发应用等操作。

（6）容灾。由于区块链网络是一个去中心化的分布式的点对点网络，为了使所有节点都能同步维护区块链网络，就需要所有节点维护的区块链网络数据完全一致。所有节点都持有整个区块链网络的所有账本数据，如果一个节点意外下线，后续上线后能从其他节点获取区块链网络全部数据。

区块链分为公有链、私有链和联盟链^[33]。公有链上的各个节点可以自由加入和退出区块链网络，并参加区块链上数据的读写，所有节点以扁平的拓扑结构互联互通，其中不存在任何中心化的服务端节点；私有链中各个节点的写入权限收归网络内部控制，而读取权限视需求选择性地对外开放，私有链仍然具备区块链运行的通用结构，适用于特定机构的内部数据管理与审计；联盟链的各个节点通常有与之对应的实体机构组织，通过授权后才能加入与退出网络，各机构组成利益相关的联盟，共同维护联盟链网络。

2.2.2 区块链层次架构

区块链架构可以分为网络层、数据层、共识层、激励层、合约层和应用层，其中核心的三个层次为网络层、数据层和共识层。各个技术应用于区块链不同的层次，形成了独具特色的架构^[34]，区块链层次架构图如图 2-2 所示。各层技术的合理有效使用为区块链网络提供了安全稳定的运行环境。

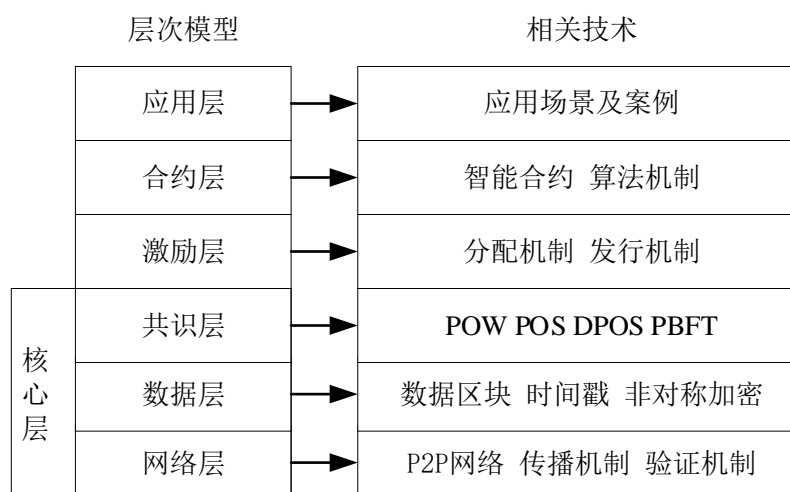


图 2-2 区块链层次架构

网络层：区块链网络本质上是一个点对点网络，网络上的资源和服务分散在各个节点上，信息的传输和服务的实现都在节点之间进行，无需中间环节或中心化服务器的介入，也无需可信第三方。

数据层：数据层是一个区块与链表的复合数据结构，包含数据区块、时间戳、非对称加密等设计。

共识层：共识层的作用是让分散的节点对区块的数据达到共识，主要的机制有 PoW、PoS、DPoS 和 PBFT 等共识算法。

激励层：激励层设计经济激励模型，鼓励节点参与区块链的安全验证工作，其中包括发行机制和分配机制。

合约层：包含智能合约和算法机制。合约层的实现使得区块链的应用领域更加的广泛。

应用层：是区块链的展示层，包含各种应用场景和案例。

区块链的层次结构中，共识层、数据层和网络层是区块链网络中的三个核心层次，是保证区块链网络正常运转的基础。

2.2.3 时间戳

时间戳是指 1970 年 01 月 01 日格林尼治时间 00:00:00(北京时间 1970 年 01 月 01 日 08:00:00)到现在的总秒数^[35]。时间戳是一段完整的、可验证的数据，它表示某个数据存在于某个特定的时间点。通常是一个字符序列，唯一地标识某一刻的时间。时间戳是一份完整的且可验证的电子凭证，它能够证明一份数据在某个特

定的时间点已经存在。在区块链系统中，当有新区块生成时，都会在这个区块中追加时间戳，然后依照区块生成时间的先后顺序，通过区块哈希值将区块连接构成区块链，每个节点又通过点对点网络建立连接，这就为信息数据的记录形成了一个去中心化的分布式时间戳服务系统。

时间戳服务系统主要包括可信时间源、签名系统和时间戳数据库三个部分：

(1) 可信时间源：是时间戳系统的时间来源，系统中所有时间的来源都必须以这个可信时间源为标准，在发布的时间戳中填写的时间必须严格按照可信时间源填写。

(2) 签名系统：负责接收构造区块时的时间戳申请，验证申请的合法性以及产生和分发时间戳，最后将时间戳存储于时间戳数据库中。用户发起时间戳申请，签名系统获取用户的数据摘要，并验证其申请的合法性，最后将时间戳和文件摘要绑定后追加签名并返回到用户。

(3) 时间戳数据库：负责保存时间戳，并且定期备份，用户可以在需要时向时间戳数据库申请从中取得时间戳。

2.2.4 哈希函数

哈希函数又叫散列函数^[36]，是一类数学函数，它可以在有限的时间内将任意长度的数据生成为固定长度的二进制串，相同的数据生成的二进制串完全相同，不同的数据哪怕差别很细微也会生成完全不同的二进制串，称为哈希值或散列值。哈希函数的输入和输出是一一对应的，如果存在没有唯一对应的情况，例如两个哈希值相同，但两个输入不同，这种情况被称作“哈希碰撞”或“散列碰撞”。因此，找到合适的哈希函数以尽可能的减少哈希碰撞十分必要。哈希函数具有单向性，任意数据不能通过其哈希值反向获取原数据，即一个从明文到密文的不可逆映射。

哈希函数在错误校正、语音识别、文件校验、数字签名等领域应用十分广泛。哈希值能够唯一地识别机密信息，这个应用场景是基于属性“无冲突”的特性，能够保护数据。在信息传递过程中，发送者通过将原消息和消息的哈希值一起发送给接受者，能够保证消息的真实性。使用哈希函数可以很直观的检测出数据在传输时是否发生错误，检测方法如下：数据发送方对将要发送的数据应用哈希函数，并将计算的结果即哈希值同原始数据一同发送到接收方。数据接收方用同样的哈希函数再一次应用到接收到的数据上，如果两次哈希函数计算出来的结果不一致，就说明数据在传输的过程中数据的某些地方有错误了，这就叫做冗余校验。

散列表也是哈希函数的主要应用之一，是由数组和链表组成的数据结构，使

用散列表能够快速的按照关键字查找散列表中的数据记录。例如，英语字典中的关键字是英文单词，和它们相关的记录包含这些单词的定义、解释等。这种情况下，哈希函数必须把按照字母顺序排列的字符串映射到为散列表的内部数组所创建的索引上，这样能够保证可以直接快速的访问表中的每一个数据。

2.2.5 Merkle 树

Merkle 树一般是一颗二叉树，由一组叶子节点、一组中间节点和一个根节点构成^[37]。Merkle 树一般用来快速比较大量的数据，当两个 Merkle 树根相同时，则意味着所代表的数据必然相同。Merkle 树对每个数据进行哈希运算后作为 Merkle 树的叶子节点，相邻两个叶子节点的哈希值拼接后再做哈希运算其结果作为两个叶子节点的父节点，然后把两个相邻的父节点做上述同样操作，直到最后只生成了一个哈希值，将这个哈希值作为 Merkle 树的根节点。其结构如图 2-3 所示。

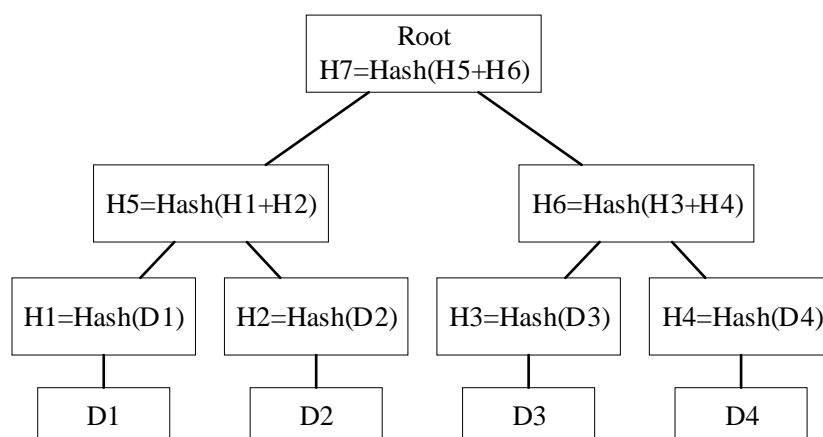


图 2-3 Merkle 树结构

2.3 区块链跨链技术

目前已运行的区块链网络有很多，例如比特币、以太坊、EOS、Fabric 等。各区块链网络之间是相互独立的，极大的限制了区块链间的互通性。为了将不同的区块链网络连接起来实现价值互联网，区块链跨链技术成了关键。目前主流的跨链技术包括公证人机制、哈希锁定、侧链和中继链。

（1）公证人机制：当跨链的交易双方互不信任且信息不对称时，最简单的方法是寻找双方都信任的中介^[38]。公证人机制也称见证人机制，是通过选举一个或一组可信节点作为公证人，对区块链 Y 上是否发生了特定事件进行验证，并向区块链 X 上的节点进行证明。公证人群体通过特定的共识算法对事件是否发生达成共识。公证人模式是目前应用最广泛的一种模式，最大的单一公证人就是交易所。

公证人机制是实现区块链之间互操作性的方案中较易实现的一种，无需进行复杂的工作量证明或权益证明，易于对接现有的区块链系统。

（2）哈希锁定：哈希锁定是系统之间进行原子交易的基本框架^[39]，能保障跨链交易的原子性，可拓展应用于中心化账本或去中心化账本的系统之间。然而，哈希锁定只能实现跨链的资产兑换，即各区块链资产总量保持不变的情况下，资产的持有人变化，无法真正将资产转移至另一条链上。对于资产转移，还需要配合其他跨链技术方可实现。

（3）侧链：侧链是相对于主链而言的一个概念，Blockstream 对“侧链”的正式定义是“侧链是验证来自其他区块链数据的区块链”^[40]。侧链协议本质上是一种特殊的区块链跨链解决方案。这种解决方案可以实现从链 X 到链 Y 的价值转移和稍后从链 Y 回到链 X 的价值转移。通常将链 X 称为主链，将链 Y 称为侧链。当主链性能出现瓶颈或者某些功能无法扩展时，把资产转移到侧链上，相关交易就可以在侧链上执行，从而达到分担主链压力、扩展主链性能和功能的目的。

（4）中继链：中继链模式适用于链接两个区块链，是实现区块链互操作性更为直接的方式^[41]。该模式不完全依赖于可信第三方的验证判断，仅通过中间人收集两条区块链的数据状态进行自我验证，其验证方式依据自身结构不同而存在显著差异。

各大跨链技术差异比较如表 2-1 所示，从表中可以看出，中继链技术虽然在实现难度相对偏大，但是在功能和特性上却是最完善的，尤其是在医疗、政务等公共服务环境与场合，具有独特而重要的应用价值，因此本文选用中继链技术来实现并测试方案。

表 2-1 跨链技术比较

	公证人机制	中继链/侧链	哈希锁定
支持跨链类型	所有	所有	部分
共识机制	大多数公证人诚实	抗 51%攻击	抗 51%攻击
可用于跨链交易	是	是	是
跨链资产的可移植性	是	是	否
支持跨链数据预言机	是	是	否
支持资产留置或抵押	是	是	部分可以
实现难度	中	难	易

2.4 Polkadot 中继链

Polkadot 是由 Web3 基金会发起的一个跨链项目^[42]，由 Parity Technologies 公司负责开发。Parity 认为尽管区块链正在如火如荼的发展着，但仍未看到区块链如同云计算一样在现实世界中大量的部署。Parity 总结造成这个现状的原因有以下五点：可扩展性、隔离性、可开发性、治理性和适用性。基于这五个原因，Parity 提出了 Polkadot 协议，并将协议实现为一个可伸缩的异构多链系统。Polkadot 提供了中继链(Relay-chain)，在中继链上可以接入大量的相同或不同结构的相对平行的区块链，称这些平行的结构化的区块链为平行链 (Parachain)。Polkadot 使开发者和企业能够利用其协议建立平行链，该协议称为 Substrate，将在下一节介绍。所有平行链都和中继链无缝连接。Polkadot 的网络架构如图 2-4 所示。

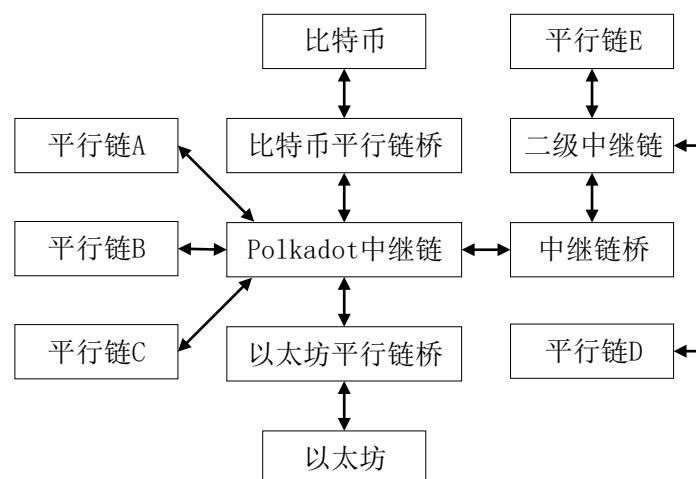


图 2-4 Polkadot 网络架构

Polkadot 架构以 Polkadot 中继链为中心向外延伸，可以连接多个基于 Substrate 框架开发的平行链，其他异构区块链例如比特币、以太坊可以通过基于 Substrate 框架开发的平行链桥连接 Polkadot。当 Polkadot 中继链网络负载过大，性能降低时，可以在 Polkadot 网络中通过中继链桥接入二级中继链，新加入的平行链可以接入二级中继链以减轻 Polkadot 中继链的负担。

交易由一个平行链到另一个平行链的工作流如下：

- (1) 用户在平行链 A 上创建一个交易以向平行链 B 发送信息。
- (2) 该交易被发送至平行链 A 的一个收集人。
- (3) 该收集人确保该交易有效，并将其包含在一个区块中。
- (4) 该收集人向平行链 A 的一个验证人展示这个区块以及状态转变证明。
- (5) 验证人验证接收到的区块，只有验证通过，确保区块中只包含有效的交

易，验证人就会抵押出他们的代币，为这个区块上链做准备。

(6) 当有足够的提名人抵押他们的代币并提名验证人时，向中继链广播其区块将得到授权。

(7) 该交易被执行，同时来自平行链 A 的数据被发送到平行链 B。

由于上述流程皆需保持无需信任，一旦核对人传输错误信息，该行为的证明会生成，随后该核对人可能遭受惩罚或清除。同时，验证人处于钓鱼人的监视下，为了保住代币，提名人会倾向于提名行为好的验证人。

2.5 Substrate 开发框架

Substrate 是从 Polkadot 抽象出来的区块链开发框架，可以基于 Substrate 框架开发出能接入 Polkadot 作为平行链的区块链系统。区块链开发人员需要了解密码学、分布式一致性、数据库、共识、经济学等诸多知识。区块链网络的复杂性也给区块链开发带来了难度，区块链网络由各个节点组成，这些节点在对等 (P2P) 网络上连接在一起。节点是运行区块链软件的网络上的各个计算机，为了使区块链网络正常运行，区块链节点需要以下部分：数据库、P2P 网络、共识引擎、交易处理、状态转移函数等。这些技术涉及计算机科学的广度较大，因此通常需要专家团队来开发。

但是，大多数区块链项目并不是从零开始开发的，他们并不注重这些基础技术，这些项目是从现有的区块链中分叉，创造出“新的”区块链，例如：比特币被分叉为：Litecoin, ZCash, Namecoin, Bitcoin Cash 等；基于以太坊存储库创建的区块链：Quorum, POA Network, KodakCoin, Musicoin 等。以这种方式构建区块链存在严重的局限性，因为这些区块链平台并非是对区块链结构进行改进，没有对区块链的进步起到推动作用。为了能使区块链开发更高效，开发出的区块链更具有特点，Substrate 应运而生。

Substrate 是一个用于构建区块链的开源的、模块化的和可扩展的区块链开发框架。Substrate 从根本上设计了区块链基板，为区块链开发者提供了一个灵活的框架来设计和构建区块链网络。Substrate 提供了构建自定义区块链节点所需的所有核心组件。这些组件包括：Substrate Runtime、Substrate Front-end、Substrate ink。

(1) Substrate Runtime: Substrate Runtime 可以是开发者自定义区块链的业务逻辑。Substrate Runtime 将定义用户可以调度的存储项目和功能。Substrate Runtime 提供了一组可以构建和配置的模块，称为 Pallets。Substrate Runtime 还提供了必要的支持库，以使这些 Pallets 与区块链用户交互。每个 Pallets 包含特定领域的逻辑和存储项目。在 Substrate Runtime 中，还可以构建自定义的 Pallets，用以扩充区块

链功能。可以使用标准的 pallets 界面添加自己的 Pallets，并访问其他 Pallets 的公共方法和特征。Pallets 构成 Runtime 的方法如图 2-5 所示，在 Runtime 中可以根据区块链的业务逻辑来自由的选择 Pallets，已达到实现区块链功能的目的，这些 Pallets 都是可插拔的。

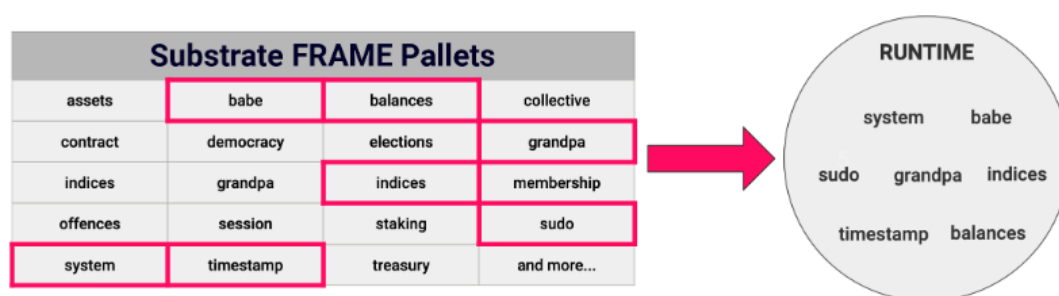


图 2-5 Substrate 中 Pallets 构建 Runtime

(2) **Substrate Front-end**: 是 Substrate 的前端界面，也是客户端，通过 substrate 开发的区块链可以使用 Substrate Front-end 与之交互。Substrate Front-end 用前端 React 框架开发，运行在 Node.JS 为后端的程序上，提供了与 Substrate Runtime 各个 pallets 交互的接口，无需开发人员再自行编写接口。

(3) **Substrate Ink**: 即 Substrate 中的智能合约。传统的智能合约平台允许用户在一些核心区块链逻辑之上发布其他逻辑。由于智能合约逻辑可以由任何人（包括恶意行为者和经验不足的开发人员）发布，因此围绕这些公共智能合约平台构建了许多安全防护措施。这些措施包括：

费用: 确保对智能合约开发者收取的费用是强制他们执行合同的计算机上进行的计算和存储的费用，并且不允许滥用块创建者。

沙箱: 限制智能合约的功能，使其无法直接修改区块链核心存储或者其他智能合约的存储。它的功能仅限于修改自身的状态，以及对其他协定或运行时函数进行外部调用的能力。

州租 (State Rent): 智能合约占用了区块链上的空间，因此应当根据智能合约所占的链上存储空间收取费用，这样可以确保智能合约开发人员不会利用“免费，无限存储”的漏洞。

还原: 智能合约可能容易出现导致逻辑错误的情况。不能保障智能合约开发人员的能力值很高不会出现错误，因此添加了额外的开销来支持交易失败时还原交易，当出现问题时不会更新任何区块链状态。

这些不同的措施使运行智能合约的速度变慢且成本上升，但是在一定程度上

牺牲性能来提高安全性是十分有必要的。

Substrate 可以做到运行时自动升级, 无分叉, 真正做到 Code Is Law。Substrate Runtime 可以编译成 Wasm 版本和 Native 版本。Wasm 版本是基于 WebAssembly 编译的可以直接在浏览器中运行并且区块链上每个节点的 Wasm 版本都统一; Native 版本节点修改代码后编译在本地的二进制可执行文件, 并且运行速度大于 Wasm 版本。Substrate 链上有 Wasm 版本的 Runtime 备份。当 Runtime 接收到外部输入准备运行改变的时候, 它会从链上获取 Wasm 版本的 Runtime, 并将它与 Native 版本的 Runtime 版本进行比较。如果相等, 那么就会运行 Native 版本的 Runtime, 这样运行效率更高; 如果不等, 就会运行链上 Wasm 版本的 Runtime, 保证整个区块链网络运行的代码完全一致。因此, 基于 Substrate 构建的区块链中的节点总是可以运行在正确版本上面, 永不分叉。而链上 Runtime 的升级是通过社区公投的形式来达成共识, 并通过 Runtime 的最高权限 root(sudo)交易来更新链上 Wasm 版本的 Runtime 代码, 强制执行升级, 因此可以做到对链的无缝升级。

本文提出的基于区块链的电子健康记录隐私保护方案中的医疗区块链就是基于 Substrate 框架开发的, 并用 Polkadot 中继链网络实现电子健康记录在不同的医疗区块链间共享。

2.6 本章小结

对本文中用到的区块链相关技术进行了阐述。接着介绍了区块链跨链技术, 包括公证人机制、哈希锁定、侧链和中继链, 然后介绍了中继链技术中的 Polkadot 中继链, 最后概述了 Substrate 区块链开发框架。

第三章 基于区块链的电子健康记录隐私保护需求分析

3.1 区块链用于电子健康记录

电子健康记录发展至今，虽然已经趋于成熟和稳定，并在一定程度上满足了隐私性和安全性，但仍然在如下几个方面表现出不足：

(1) 隐私泄漏：电子健康记录中存储的患者数据都是由医生创建、查看、调用和修改的，数据存储于云服务提供商中，然而这些数据的所有者却是患者，医生和云服务提供商可能在未经患者允许的情况下随意操作患者数据，同时云服务提供商遭受网络攻击也可能导致电子健康记录泄露。患者的隐私数据可能在不知情的情况下就被泄露和修改。

(2) 数据追溯：电子健康记录追溯分两个方面，一是对电子健康记录操作的追溯，及对电子健康记录从创建到删除的生命周期的追溯；二是对电子健康记录内容的追溯，对电子健康记录的修改、更新可能会导致部分数据的丢失，恶意的篡改可能会导致数据的丢失，无法追溯到电子健康记录的历史数据。

(3) 数据共享：现存医疗系统中，各个医院之间相互隔离，很难做到共享电子健康记录，数据不互通，形成了一个信息孤岛。当患者转院时，所有的诊断和检查都需要重新做一次，这不仅造成了资源浪费，还可能耽误治疗时间。

对于上述电子健康记录存在的不足与缺陷，选择更好的解决方案显得十分必要。区块链技术因其不可篡改、可追溯等各种性质，在数据隐私保护方面的能力尤为突出，可以用于电子健康记录的隐私保护。

不少学者研究了区块链在医疗领域的应用价值，基于这些研究^[43,44]，同时分析了电子健康记录存在的不足与缺陷，以及对现有基于区块链的电子健康记录隐私保护方案的分析，可以发现已有的基于区块链的电子健康记录隐私保护方案中仍然存在一些未解决的问题。例如：区块链在解决医疗领域的信息孤岛问题时，要求所有医院接入同一条区块链，但是区块链上的数据透明、公开，数据一旦上链，所有机构都能看到，一些医院不能被外部看到的数据也会被公开；两个医院使用两个不同的区块链，当患者需要在两个医院之间转院时，因为是不同的区块链不具备互通性，不允许数据跨区块链传输，所以患者所有的诊断信息需要重新进行。

如果医院使用区块链构建医疗系统，可以有效的保护电子健康记录的隐私，但是不同医院使用不同的区块链，更加难以做到数据共享，需要使用跨链技术来解决。

区块链跨链技术可以将不同的区块链连接起来，进行跨链数据共享。区块链跨链技术中较为突出的是 Polkadot 中继链技术，它保障了数据的透明性、不可篡改性、实时性以及系统的稳定性。在医疗领域，引入区块链及其跨链技术有以下优点：

（1）可以溯源。区块链能记录对数据的所有操作，能有效地追溯任何数据的创建、调用和删除。

（2）不可篡改。区块链具有不可篡改的性质，能有效保证电子健康记录数据不被医生或云服务提供商篡改。

（3）信息互通。各医院建立自己的区块链，并通过中继链连接在一起，数据可以通过中继链在不同医院的区块链间传输。

（4）保障隐私。区块链的匿名性，确保了电子健康记录数据的隐私得到保护。

3.2 需求分析

基于前文提出的患者在医院 A 就诊需要转院到医院 B 时的场景，结合对区块链用于电子健康记录的分析，抽象出了图 3-1 的系统架构。该系统架构基于 Polkadot 中继链，构建了基于区块链的电子健康记录隐私保护系统，可应用于各个医院、诊所、医疗机构的信息系统构建，安全存储电子健康记录，有效保障患者隐私，提升不同医院间的信息共享互通性。

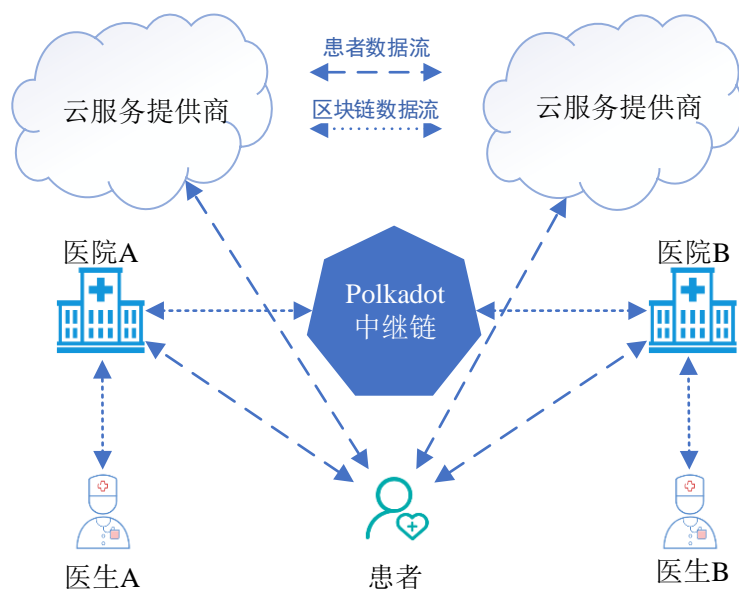


图 3-1 系统架构

3.2.1 角色分析

系统中的角色包括：医生、患者、医院、云服务提供商、Polkadot 中继链，各角色的主要功能及作用如下：

医生：是电子健康记录数据的创建者、查看者和调用者，即医生持有患者电子健康记录的控制权。医生可能会出于牟利的目的而泄露患者电子健康记录，甚至当出现医疗事故时，医生可能会为了逃避责任而恶意修改电子健康记录，因此本系统应当对医生的行为做出限制。医生对电子健康记录的每次操作都应当存储在区块链中，因区块链中的数据不篡改，就可以按时间顺序追溯出电子健康记录的所有操作，有效保护患者的隐私。

患者：是电子健康记录数据的所有者，却不具备对电子健康记录的控制权，无法创建、修改、删除自己的电子健康记录，但是电子健康记录中包含大量的患者隐私数据，无论是法律要求还是出于隐私保护的目，患者都应当具备对自己电子健康记录操作的权利。系统应当允许用户对自己的电子健康记录进行操作，包括删除操作。

医院：为患者提供诊疗服务的场所，医生通过医院的终端设备对患者的电子健康记录进行操作。医院的终端设备应当作为区块链的节点加入区块链，以构建基于区块链的电子健康记录隐私保护系统。

云服务提供商：是电子健康记录数据真实存储的地方。医院难以维持大量存储空间的设备，所以医院数据的存储需要依靠诚实可信的云服务提供商，如腾讯、阿里、亚马逊等。

Polkadot 中继链：负责在不同医院的区块链间跨链传输数据。不同医院使用不同的区块链系统，当需要在两个医院间跨链传输时，就需要引入 Polkadot 中继链。所有需要跨链的医院区块链都接入 Polkadot，就能在 Polkadot 中继链网络中实现数据跨链传输。

3.2.2 功能分析

结合角色分析，将对系统的主要功能进行介绍。主要功能包括两个部分：区块链链上功能和对电子健康记录操作的功能。区块链链上功能包括：账户、交易、共识、构造区块，链上存储，区块上链，区块查询等区块链应当具备的基本功能；电子健康记录操作功能包括：创建电子健康记录、查看电子健康记录、修改电子健康记录、删除电子健康记录和电子健康记录跨链传输。有别于传统的电子健康记录管理系统，本文提出的方案及系统中，医生对电子健康记录的所有操作都是在区块链上进行的，能够按时间顺序记录医生对电子健康记录的流程，同时将区

区块链接入 Polkadot 中继链实现不同医院区块链的数据跨链流转。本系统方案的主要功能及特点如下：

(1) 基于区块链构建的电子健康记录隐私保护系统应当具备区块链的基本功能。区块链作为数字货币的交易平台，应当具备账户、交易、共识等功能；作为数据存储平台，应到具备构造区块、数据存储、数据上链等功能。

(2) 区块链上电子健康记录创建功能。电子健康记录需要具有一定的专业知识的人才能进行操作，因此对于电子健康记录的创建操作应当交给医生而不是患者，患者只需在就诊时向医生提供必要的个人信息即可。而电子健康记录一般需要较大的存储空间，医生创建电子健康记录是基于区块链创建的，区块链中不宜存储较为复杂或者需要存储空间较大的数据，所以在区块链中应当只保存对电子健康记录的索引，完整的电子健康记录存储于云服务提供商中。具体流程是：医生需要持有区块链的账号，并将电子健康记录索引信息构造成区块准备上链，待区块上链成功后将完整的电子健康记录信息发送到云服务提供商中，云服务提供商对数据进行检测后并完整的存储电子健康记录。

(3) 区块链上电子健康记录修改功能。当患者第二次就诊于同一家医院时，医生可能会对电子健康记录做出修改，需要将修改操作记录在区块链中并将修改后的电子健康记录完整的保存在云服务提供商中。具体流程是：医生通过患者提供的信息生成电子健康记录索引，在区块链中找到上次操作电子健康记录所保存的区块，读取出区块中的相关信息，将相关信息发送到云服务提供商，云服务提供商验证信息后将完整的电子健康记录发送给医生，医生修改电子健康记录后按照创建功能的流程再次将修改后的电子健康记录发送给云服务提供商。

(4) 区块链上电子健康记录跨链传输功能。患者首先在医院 A 就诊后转诊至医院 B 时，可以选择在医院 B 重新创建电子健康记录，也可以选择跨链从医院 A 获取电子健康记录的相关信息以获取完整的电子健康记录。如果在医院 B 重新创建电子健康记录，为了和医院 A 创建的电子健康记录区分，就需要在电子健康记录中加入创建者字段，用以区别不同医院创建的患者的电子健康记录；如果选择从医院 A 获取电子健康记录，就需要使用 Polkadot 中继链技术，通过中继链在两个医院的区块链之间跨链传输数据。

(5) 删除电子健康记录功能。患者和医生都可以删除电子健康记录。患者可以通过系统查看属于自己的电子健康记录，并且可以对电子健康记录进行删除，将删除信息发送给云服务提供商，由云服务提供商确保对电子健康记录的删除操作成功。医生对电子健康记录的删除需要在区块链中记录，并将删除操作发送到云服务提供商，云服务提供商删除电子健康记录数据。

3.2.3 跨链流程分析

跨链流程如图 3-2 所示。

- (1) 患者在医院 A 就诊。
- (2) 患者向医院 A 的医生 A 提供自己的个人信息，并进行诊断。
- (3) 医生 A 诊断后为患者创建电子健康记录，并存储到医院 A 的区块链中。
- (4) 患者转诊到医院 B。
- (5) 患者向医院 B 的医生 B 提供自己的个人信息，并进行诊断。
- (6) 医生 B 在医院 B 的区块链中检索是否有患者的电子健康记录信息，若没有则向医院 A 发起跨链获取电子健康记录的请求。
- (7) 医院 B 通过区块链经由 Polkadot 中继链向医院 A 的区块链发起跨链交易，以获取患者的电子健康记录。
- (8) 医院 B 的跨链交易请求被 Polkadot 中继链共识通过后，医院 A 向医院 B 发送电子健康记录相关信息。

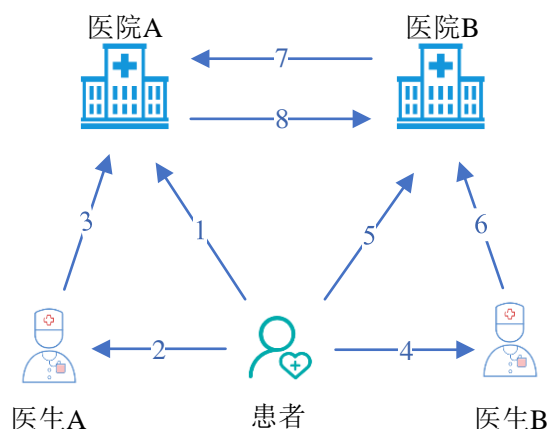


图 3-2 跨链流程

3.3 本章小结

本章总结了电子健康记录传统存储方式存在的不足与缺陷，接着分析了区块链用于电子健康记录隐私保护的价值，以及现有基于区块链的电子健康记录隐私保护方案中的不足，分析了 Polkadot 中继链跨链技术用于电子健康记录隐私保护的可行性。最后对基于 Polkadot 的方案的需求进行了分析，包括角色分析、功能分析、流程分析。

第四章 基于区块链的电子健康记录隐私保护方案

电子健康记录存储着大量的患者隐私数据和医疗科研数据，传统电子健康记录系统不论是应用层还是存储层都存在隐私信息泄露的风险，患者的隐私可能遭到护士、医生以及云服务提供商的泄露，同时传统电子健康记录系统也为医生逃避责任提供了便利。各医院间还存在信息独立不互通的问题，这不利于患者转院就诊，也不利于医疗研究。因此，改变传统电子健康记录系统，保护电子健康记录数据隐私的同时达到医疗机构间信息互通具有重要研究意义。

依据前一章的分析，为了利用区块链解决电子健康记录隐私保护问题，本章提出了基于区块链的电子健康记录隐私保护方案，简称 CEPS，该方案提出了将区块链作为服务基础，解决了区块链应用于电子健康记录隐私保护时的数据不可修改删除的问题，以及数据量过大不利于区块链存储的问题。本方案在基于区块链的电子健康记录隐私保护方案的基础上还引入了 Polkadot 中继链跨链技术，提高了整个方案的隐私保护能力以及信息处理能力。

4.1 方案概述

如图 4-1 所示，本方案模型总共包括应用层、区块链层、数据层等三个部分。应用层可以对区块链层的医疗区块链进行操作，也可以在数据层中获取电子健康记录数据，区块链层可以通过 Polkadot 中继链跨链数据传输。

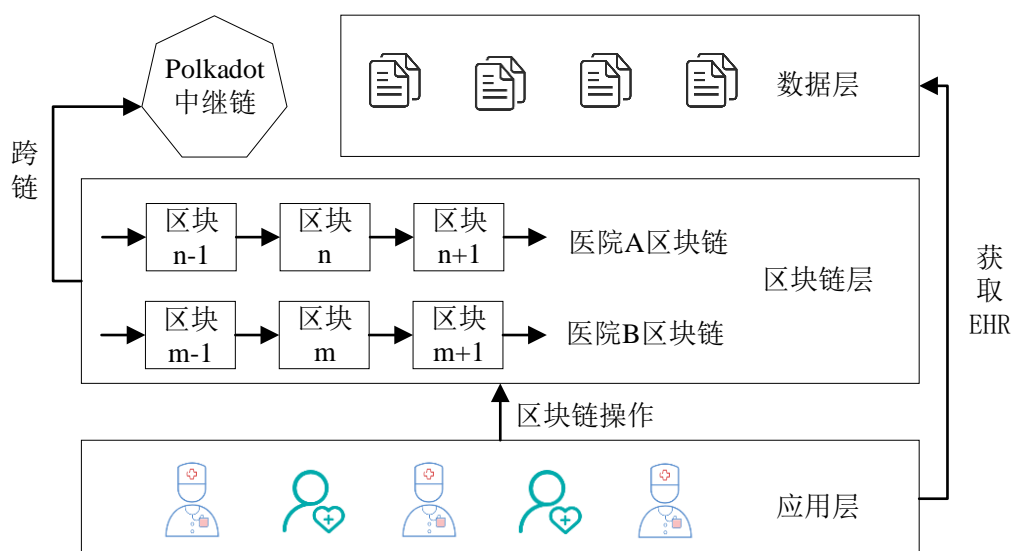


图 4-1 基于区块链的电子健康记录隐私保护方案模型

每层的详细功能如下：

应用层：应用层中的包括数据所有者、数据创建者和数据使用者。在该模型中数据指电子健康记录数据，数据所有者为患者，数据创建者和使用者为医生。用户可以使用区块链来记录电子健康记录的创建和使用的全过程，可以通过区块链追溯电子健康记录的流程，可以从数据层获取完整的电子健康记录数据。应用层通过 web 页面来提供服务，本方案采用 Node.JS 和 React 构建应用层。

区块链层：区块链作为电子健康记录隐私保护方案的核心，提供了强大的隐私保护能力。电子健康记录并不完整存放于区块链中，而是将电子健康记录的索引信息存放在区块链中，当需要对电子健康记录进行任何操作时，都需要先在区块链中构造区块记录本次操作，并在达成共识后上链获取电子健康记录索引，依据索引在云存储中获取完整的电子健康记录信息。不同的医疗机构持有不同的区块链。本方案采用 Substrate 区块链开发框架构建区块链用于医院存储电子健康记录，有效保护电子健康记录隐私。区块链层中还包含了 Polkadot 中继链。Polkadot 中继链用于区块链层中不同的区块链交换信息。医院 A 的区块链中包含的电子健康记录信息的索引可以通过 Polkadot 中继链发起跨链交易，转移到医院 B 的区块链中。

数据层：数据层中存储完整的电子健康记录信息以及电子健康记录信息的索引，数据层可以选取可信的云服务提供商，在本方案中用 MongoDB 数据库代替。

自此，已可以较好的实现电子健康记录在不同的区块链网络中进行共享。然而，还有两个问题需要考虑：

（1）区块链上存储的数据不可删除，但是美国、欧洲、中国等国家均已颁布了电子健康记录数据隐私保护条例，规定电子健康记录应属于患者，患者有权要求医院不保留患者的电子健康记录。医院不能长期保存患者数据，当患者请求删除其电子健康记录数据时，医院应该能够满足这些请求。

（2）电子健康记录数据可能包含图片、视频和其他信息，这些信息可能占用大量存储空间。如果数据完全存储在区块链上，会导致区块变大，构建区块的时间变长，降低区块链网络的性能。

针对这两个问题，本方案从电子健康记录中选取重要字段，通过哈希函数构建电子健康记录的唯一索引，并将电子健康记录中的隐私数据进行脱敏处理，然后存储于医院的云存储服务提供商中。这样，不仅可以大大减少存储于区块链网络中的数据体量，而且当患者要求删除其电子健康记录时，只需要在云服务提供商中删除对应患者的电子健康记录索引即可，方便有效。

4.2 基于 Substrate 的电子健康记录隐私保护方案

在本节中，将详细讨论基于 Substrate 构建的医疗区块链的隐私保护方案，其中包括方案概述，流程概述，功能概述等。

4.2.1 方案概述

基于电子健康记录数据的特点，研究了现行的多种基于区块链的电子健康记录隐私保护的方案，提出了基于 Substrate 构建区块链的电子健康记录链隐私保护方案，如图 4-2。方案分为两部分，区块链和电子健康记录。

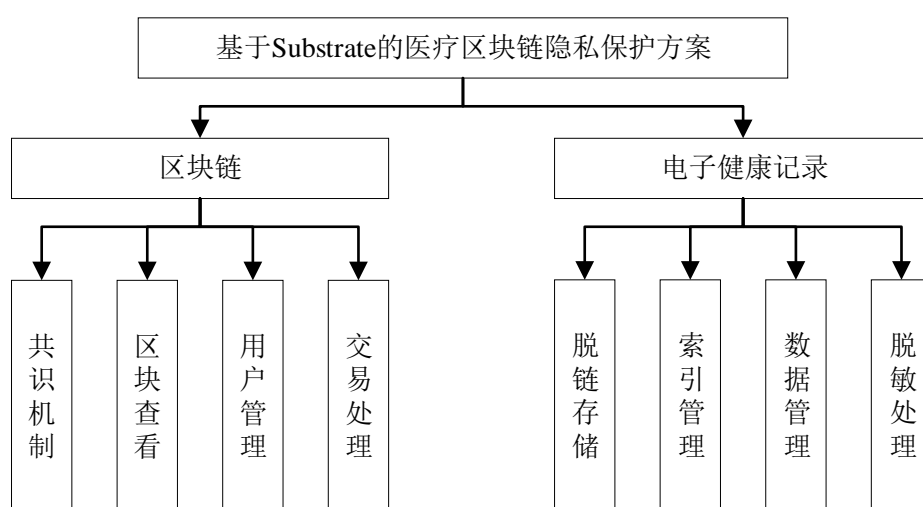


图 4-2 基于 Substrate 的医疗区块链隐私保护方案

本方案实现了区块链的主要功能，这些功能包括共识机制、区块内容查看、区块链用户管理、区块链交易处理等，为电子健康记录隐私保护提供了区块链基础。

对于电子健康记录的存储，本方案采用将完整的电子健康记录存储在数据库中，在区块链中存储电子健康记录的索引信息的方式，医生通过区块链上存储的索引信息就能在数据中获取完整的电子健康记录。这种方式就需要设计区块链的区块结构。一条区块链由多个包含着前一个区块哈希值的区块连接而成，每个区块的区块体都包含了若干交易信息，这些交易信息就是区块链中实际保存的数据。区块链就像是一个分布式数据库，区块链上的每一个区块就像是数据库中的一张表，每个交易就是表中的一条记录。

基于 Substrate 构建的区块链的区块主要由两部分构成区块头和区块体，每个区块的具体构成如图 4-3 所示。区块头表示的是需要进行数字签名的部分。区块头

中包含上一个区块的哈希值，区块生成者的公钥，由交易内容生成的 Merkle 树根哈希值和生成区块的时间戳。区块头体的内容包括区块生成者对于区块头的数字签名，交易的个数，和保存在此区块中所有的交易。数字签名是为了保证区块内容不被篡改，并确保区块生成者在生成恶意区块后无法抵赖。另外，区块中仅保存交易的 ID，即仅保存指向某个交易的索引，而不保存交易本身，这样便可使每个区块容量降低，便于同步与备份。区块、交易物理上都保存在数据库中，在逻辑上以区块链的形式来存储。在交易设计存储上，实际上只是在正常存储于数据库的数据上添加交易 ID，交易类型、时间戳、公钥、数字签名等交易字段信息，将所要存储的信息作为交易单内容，形成逻辑上的交易单，其物理存储上与一般数据存储并无太大区别。

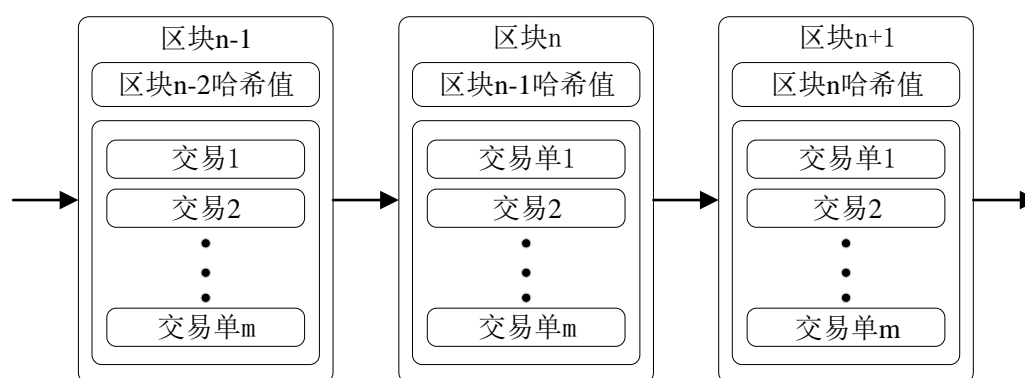


图 4-3 区块链中区块构成

交易类型表示这个交易的类型，如电子健康记录的创建、删除、查询和修改，以指示验证器集群进行相应的操作。之所以在增删查改的操作时使用交易而不是直接访问某一拥有区块链的节点，有两个原因。首先某一节点可能进行恶意操作，违规暴露患者的信息，或是篡改患者的信息等。另一方面是为了将操作者的操作记录在区块链中，操作者在进行操作时需要使用自己的私钥对交易进行数字签名，使得操作者对于自己进行过的操作无法抵赖。交易内容是该交易中所存储的内容，即电子健康记录索引。时间戳表示该交易单生成的时间，公钥为交易生成者的公钥，交易 ID 为对交易类型，交易内容，时间戳和公钥进行哈希运算后生成的哈希值。哈希算法和编码算法可以选择 SHA-256 哈希算法或 BASE64 编码算法其可靠性在各种区块链系统中得到了检验。数字签名是交易生成者对交易 ID 的签名，防止交易被篡改。每个交易中的内容如图 4-4 所示。



图 4-4 区块中交易内容

对于电子健康记录索引的链上存储，主要流程如下：

(1) 数据通过“哈希函数”传递，将源（假设无限量的比特）转换为精确 256 位（哈希）的字符串。这些位由大约 80 位数字表示，并且对于数据源是完全唯一的。如果源数据完全相同并再次通过散列函数，则它将生成完全相同的散列。如果它被改变了，那么两个哈希将完全不同。可以比较这两个结果以验证源数据是否被篡改，从而消除了欺诈行为。这些散列函数只能以一种方式工作，这意味着从散列中获取原始源数据在计算上是不可行的。

(2) 然后将散列存储在网络中的每个节点上而不是原始数据上，占用的空间比以前少得多。每个节点都可以通过将数据放入散列函数并将结果与网络中提供给它们的散列进行比较来验证数据。

(3) 当用户希望下载数据时，可以从网络提供的散列中找到数据源。一旦通过散列（时间戳等）提供的信息找到用户/数据库，他们可以假设地建立“对等”网络。一些数据存在这种传输的服务，例如 IPFS，但需要另一个程序才能与区块链一起运行。

(4) 一旦数据被传输，双方和所请求的系统上的任何其他节点都可以验证该数据的真实性，提供区块链技术的公证，而不需要花费不必要的时间，金钱和存储空间。

4.2.2 方案流程

基于 Substrate 医疗区块链隐私保护方案的具体步骤如下：

(1) 用户选择对电子健康记录的操作类型（共定义了四种操作类型，创建、修改、查看和删除），并对电子健康记录做相应的操作。

(2) 对电子健康记录做完相应操作后, 将电子健康记录的索引信息和相应操作信息, 封装打包构造造成区块, 并准备上链。

(3) 区块链验证当前用户身份信息, 验证通过后用户发起一次上链交易申请, 等待区块链对此次交易达成共识。

(4) 如果此次操作未能在区块链网络中达成共识, 网络认为此次操作不合法, 将回滚之前的所有操作。

(5) 如果此次操作在区块链网络中达成了共识, 在区块中添加当前时间戳和前一区块的哈希值, 并将构造的区块追加到区块链中。

(6) 无论达成共识与否, 都将此次操作的结果构造造成 Substrate 的 Event 结构体, 并将构造好的 Event 结构体发送至客户端, 通知客户端此次操作结果。

基于上述步骤, 本方案的流程图如图 4-5 所示。

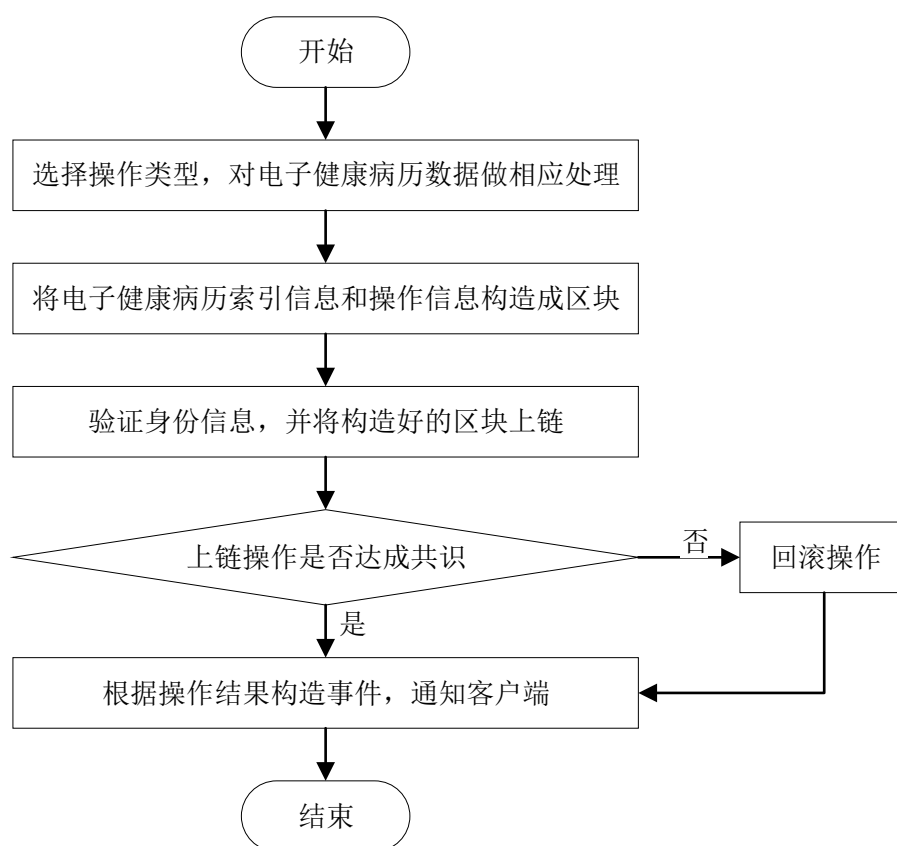


图 4-5 基于 Substrate 医疗区块链隐私保护方案的流程

4.2.3 电子健康记录脱敏设计

电子健康记录数据不仅包含患者隐私数据, 而且需要较大的存储空间, 用区块链直接存储患者的电子健康记录信息, 一方面受限制于一个区块的大小固定,

一个较大的电子健康记录数据可能会被分割成几个部分存储于不同的区块中，不利于医疗数据存储；另一方面，由于区块链是一个完全公开的账本，参与医疗区块链的患者、护士、医生等都平等的拥有区块链账本记录，不利于保护患者隐私。为了合理存储电子健康记录数据，并有效保护患者隐私，本文采用数据脱敏技术^[45]来处理电子健康记录中的敏感数据，将电子健康记录数据脱敏处理，即使获取到个人数据，也无法被准确识别身份，同时又没有将数据完全加密使得数据具备一定的可用性。

通常根据不同数据特征选择不同的脱敏算法，脱敏算法通常包括屏蔽、变形、替换、随机、泛化、格式保留加密和强加密等。由于医疗场景下设计的数据类型较多，各种数据用途不同，所以需制定合理的数据脱敏策略，并正确识别敏感数据字段，结合多种数据脱敏算法才能达到医疗数据脱敏化的目的。

在医疗场景中，电子健康记录数据中通常包含姓名、性别、年龄、手机号、身份证号、家庭地址、时间、病情、病史、用药史、医嘱等较为隐私的数据。综合分析上述各种字段数据的类型和特征，对这些字段的数据定义了一下几种脱敏策略，以有效的保护患者隐私：

(1) 部分屏蔽：屏蔽是将敏感数据的部分内容用不暴露数据原始信息的其他字符进行替换，例如#、&、*等特殊符号，一般只是用一种字符去替换一条数据的部分内容，这样既能保证敏感数据部分保持公开，又能合理的保护患者隐私。电话号码是一个极度需要保护的字段，保留电话号码的前三位和后四位，中间四位用“*”屏蔽。姓名能直接定位到具体的患者，但是对于医学而言又不具备较大的统计学意义，所以对于姓名留姓氏，用“*”屏蔽名。

(2) 身份证号：国家质量技术监督局于 1999 年 7 月 1 日实施的 GB11643-1999《公民身份号码》中明确规定了 18 位身份证标准，依顺序依次为六位数字地址码，八位数字出生日期码，三位数字顺序码和一位数字校验码。六位数字地址码能体现出患者的所在地，对于疾病发生的地域性有一定的统计效力，需要保留；八位数字出生日期码能反映患者的完整年龄信息，需要屏蔽；三位数字顺序码和一位数字校验码不能直接定位到患者信息，但是三位数字顺序码又能体现患者性别，有统计学意义，不屏蔽，例如 510108*****2411。

(3) 年龄泛化：在医疗场景下年龄有重要意义，有一定的统计价值，例如某类病症在哪个年龄阶层较为高发，某个科室哪个年龄阶层就诊较多等，这些统计都能有利于医生、医院做相关调查研究。据泛化就是将数据集从较低的概念层抽象到较高的概念层的过程，用较高层次的概念来代替较低层次的概念。对于年龄这种数值型的字段来说，将其用一个合理的范围来代替即可，例如患者 24 岁，则

用 21-25 岁来表示。以 5 位间距对年龄进行泛化的操作如下：定义人类最大年龄为 150 岁，并以 5 岁为间距划分为 30 个分段，对患者年龄采用折半查找的方法找到对应的年龄段，其决策树如图 4-6 所示。

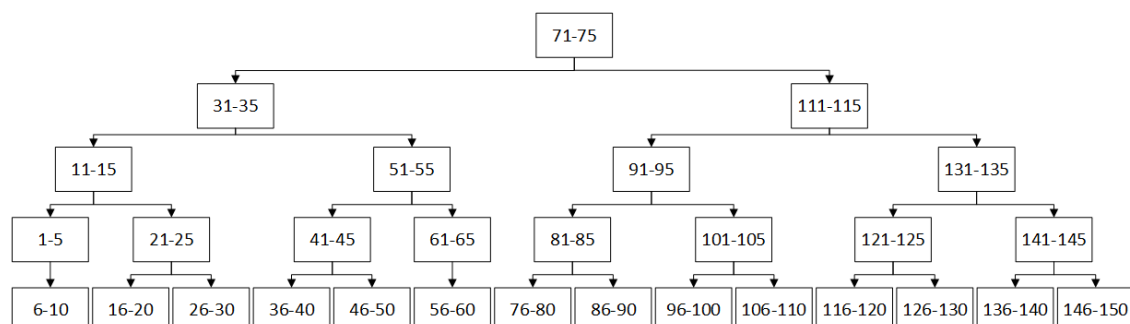


图 4-6 年龄泛化决策树

4.2.4 电子健康记录索引

不同于数据库技术中的索引，这里的索引是指能唯一标识一组数据的一个特殊值。这个特殊值应当具有唯一性，在一个数据集里能通过一个索引唯一的找到一个数据元，应当是一一对应的映射关系。这个特殊值还应当便于生成，不能过于复杂，基于它指代的数据集就能生成一个唯一的索引值。哈希函数能将不同长度的数据生产固定长度的二进制串，且这个二进制串不具备可逆性，即不能通过二进制串找到原数据。所以哈希函数非常适合生成索引。

电子健康记录的数据包含较多字段，且类型复杂，为了能使生成的索引值唯一，应当合理的选择字段通过哈希函数生成索引值。电子健康记录中通常都包含身份证号字段和电话号字段。身份证号作为公民的标识具有唯一性，电话号也因个人使用具有唯一性，所以将身份证号和电话号拼接，并通过哈希函数生成的索引值一定具有唯一性，且不可逆，不会暴露出患者的身份证号和电话号。

两种最常使用的哈希算法是 SHA256 和 MD5^[46]。而 MD5 早在 2005 年已经被中国密码学家王小云攻破^[47]，所以本方案选取 SHA256 作为生成索引的哈希函数。

SHA256 的流程如下：

(1) 初始化常量：SHA256 算法中用到了 8 个哈希初值以及 64 个哈希常量其中，SHA256 算法的 8 个哈希初值如表 4-1 所示。这些初值是对自然数中前 8 个质数（2,3,5,7,11,13,17,19）的平方根的小数部分取前 32bit 而来。

(2) 预处理：SHA256 算法中的预处理就是在需要 Hash 的消息后面补充需要的信息，使整个消息满足指定的结构。信息的预处理分为两个步骤：附加填充比特和附加长度

表 4-1 SHA256 算法哈希初值

h0	0x6a09e667
h1	0xbb67ae85
h2	0x3c6ef372
h3	0xa54ff53a
h4	0x510e527f
h5	0x9b05688c
h6	0x1f83d9ab
h7	0x5be0cd19

(3) 附加填充比特：在报文末尾进行填充，使报文长度在对 512 取模以后的余数是 448。填充是这样进行的：先补第一个比特为 1，然后都补 0，直到长度满足对 512 取模后余数是 448。需要注意的是，信息必须进行填充，也就是说，即使长度已经满足对 512 取模后余数是 448，补位也必须要进行，这时要填充 512 个比特。因此，填充是至少补一位，最多补 512 位。

(4) 附加长度值：附加长度值就是将原始数据（第一步填充前的消息）的长度信息补到已经进行了填充操作的消息后面。SHA256 用一个 64 位的数据来表示原始消息的长度。因此，通过 SHA256 计算的消息长度必须要小于 2^{64} ，当然绝大多数情况这足够大了。

(5) 计算摘要：将消息分解成 512-bit 大小的块，假设消息 M 可以被分解为 n 个块，于是整个算法需要做的就是完成 n 次迭代，n 次迭代的结果就是最终的哈希值，即 256bit 的数字摘要。一个 256-bit 的摘要的初始值 H_0 ，经过第一个数据块进行运算，得到 H_1 ，即完成了第一次迭代 H_1 经过第二个数据块得到 H_2 ，……，依次处理，最后得到 H_n ， H_n 即为最终的 256-bit 消息摘要。

4.3 基于 Polkadot 的电子健康记录跨链方案

在本节中，将详细讨论基于 Polkadot 的电子健康记录跨链方案，其中包括方案概述，Polkadot 角色，功能概述等。

4.3.1 基于 Polkadot 的医疗区块链隐私保护方案流程

基于 Polkadot 的电子健康记录跨链方案如图 4-7 所示，总共包含中继链、二级中继链、平行链、异构区块链和平行链桥五个部分组成。

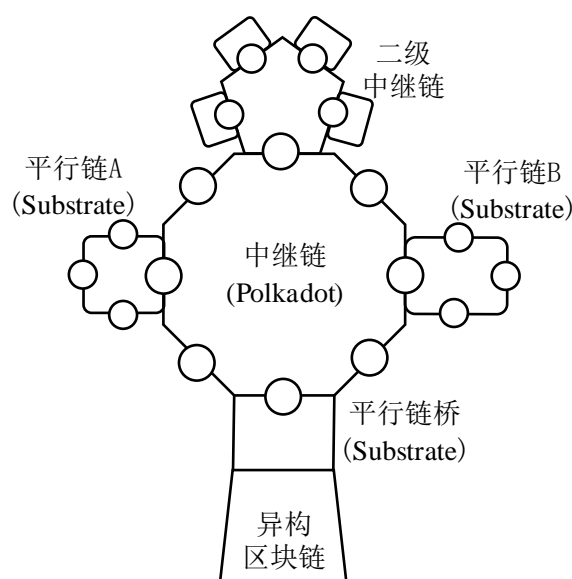


图 4-7 跨链架构

(1) 中继链：中继链是跨链技术的主要发展分支之一，其中最为突出的技术之一是 Polkadot，基于前一节的方案构建的区块链可以作为平行链接入 Polkadot 中继链网络，并且通过 Polkadot 中继链网络，不同的平行链之间可以跨区块链交易和数据交换等。

(2) 二级中继链：一个团体可以维护自己的中继链网络，并通过这个中继链网络跨链传输数据，同时这个中继链网络还能接入另一个中继链网络，这就形成了两级中继架构。

(3) 平行链：接入中继链网络的区块链称为平行链，基于 Substrate 构建的区块链与 Polkadot 架构一致，可以直接接入 Polkadot，而其他不同架构的区块链需要通过其他方式接入 Polkadot。

(4) 异构区块链：与 Polkadot 架构不同的区块链称为异构区块链，例如比特币网络、以太坊等，这些异构区块链网络也可作为平行链接入 Polkadot。

(5) 平行链桥：可以基于 Substrate 构建平行链桥用于桥接中继链网络和异构区块链网络，使得异构区块链也作为平行链网络接入 Polkadot。

4.3.2 Polkadot 角色

在区块链网络中，由于节点的性能、作用的不同，将节点分为了不同的角色，在 Polkadot 中，角色构成如图 4-8 所示。

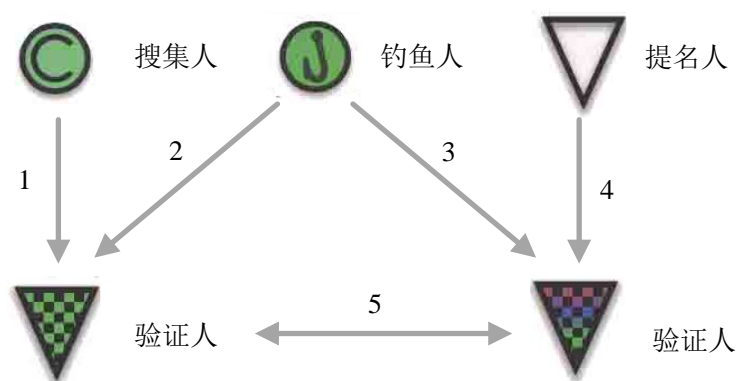


图 4-8 Polkadot 角色

Polkadot 将中继链网络中的节点分为四个角色。收集人为特定的平行链维护一个“全节点”，这意味着它们保留了所有必要的信息，以便能够编写新的区块并执行交易。这些收集人从平行链的用户那里收集并验证交易，然后将经过验证的交易发送给中继链的参与者。这些参与者称为验证人。验证人负责验证和广播来自整理器的块。为此，对于接受的每个块，验证人必须保证足够数量的 DOT (Polkadot 中继链代币)。为了确保验证者的行为准确，引入了一个称为“钓鱼人”的角色。如果钓鱼人证明了验证人的错误行为，他们将获得高额奖励。此外，验证人需要得到提名人的批准，提名人还需要承诺其代币来提名验证人。

4.3.3 Polkadot 跨链

结合系统设计和医疗区块链设计以及 Polkadot 的特性，设计实现了医疗区块链跨链模型如图 4-9 所示。电子健康记录跨链系统由三部分组成：医院 A 及其平行链 A、医院 B 及其平行链 B、中继链。医院 A 和 B 的平行链负责保存电子健康记录信息索引，中继链则负责在不同的平行链间跨链交易电子健康记录索引。

如图 4-9 所示，跨链流程可以分为以下七个步骤：

(1) 患者访问医院 A。如果患者首次访问医院 A，患者将在医院 A 区块链上创建一个账号，并初始化存储在区块链中的信息。

(2) 患者由医院 A 的医生诊断。患者向医院 A 的医生提供其自己的区块链账号。医院 A 的医生在区块链上查询账户患者的电子健康记录。如果区块链上没有电子健康记录，则从指示医院 B 查询其记录。然后，发起医院 A 到医院 B 跨区块链交易。

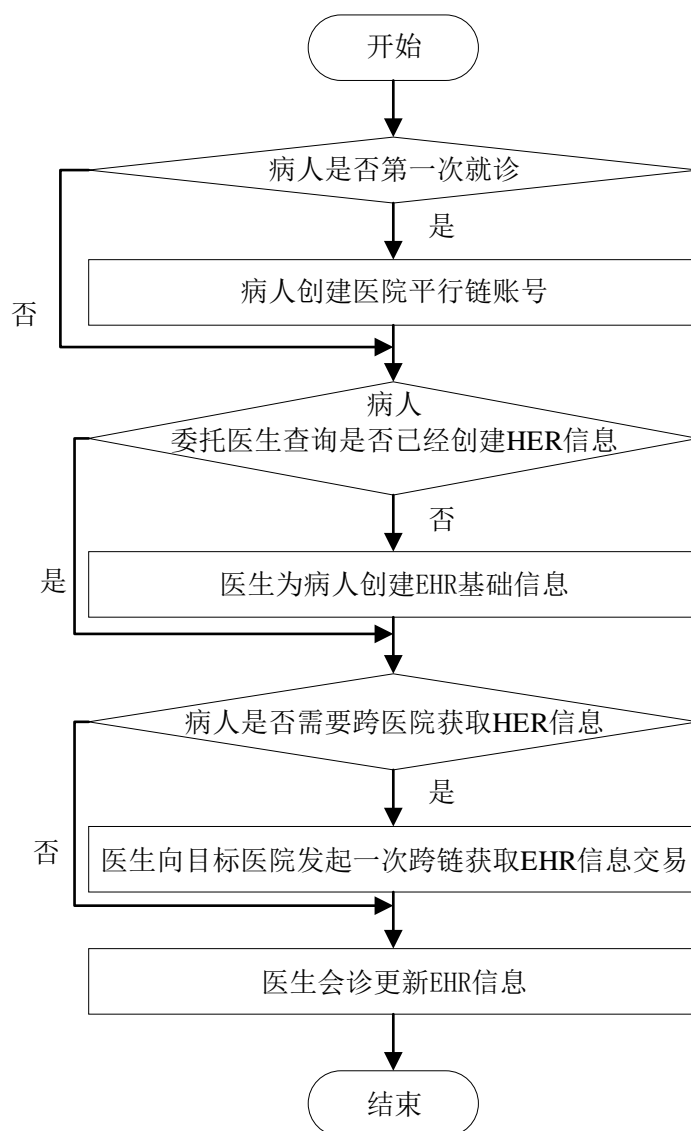


图 4-9 跨链设计流程图

(3) 医院 A 的医生作为跨链构造区块信息，并将验证信息发送给医院 A 区块链的验证者节点。如果验证者接受此跨区块链交易，它将更新状态，并等待 Polkadot 中继链操作。

(4) Polkadot 中继链获取跨链交易状态的 Merkle 树，并将交易信息传输到医院 B 区块链的跨链交易状态 Merkle 树，以完成医院 A 向医院 B 的跨链操作。

(5) 患者收到医院 A 跨链消息后，发起医院 B 到医院 A 的跨区块链交易。

(6) 医院 B 区块链其中一个节点作为 Polkadot 的收集人节点获得来自患者的跨链交易请求，过程与步骤 3 相同。

(7) Polkadot 中继链获得院 B 区块链的跨链交易状态 Merkle 树，并将交易信

息（区块信息）转移到院 A 区块链的跨链交易状态 Merkle 树，以完成医院 B 的跨区块链交易。医院 B 不再有权操作患者的电子健康记录数据，医院 A 的医生可以获取患者的电子健康记录私有数据。

过程的形式化表示如下，其中表 4-2 描述了相关符号的定义。

表 4-2 Polkadot 流程符号表示

符号	描述
P	患者
H _A	医院 A
H _B	医院 B
D _{HA}	医院 A 的医生
D _{HB}	医院 B 的医生
Egr _{HA}	医院 A 的区块链出口
Egr _{HB}	医院 B 的区块链出口
Igr _{HA}	医院 A 的区块链入口
Igr _{HB}	医院 B 的区块链入口

(1) P-> H_A: P 访问 H_A。如果 P 首次访问 H_A, P 将使用社交唯一标识符在 H_A 私有区块链上创建一个帐户 ID (P), 并初始化加密并存储在区块链中的隐私信息。

(2) P-> D_{HA}: P 由 D_{HA} 诊断。P 向 D_{HA} 提供其自己的区块链帐户。D_{HA} 在区块链上查询帐户 ID (P) 的 EHR 并将其解密。如果区块链上没有 EHR, 则 P 从指定的医院 H_B 查询其记录。然后 P 启动 H_A→H_B 跨区块链交易。

(3) D_{HA} -> H_A: 作为 Polkadot 的整理者, D_{HA} 从 P 收集跨区块链交易请求并开始构建包含信息块的块: (b.header, b.ext, b.proof, b.receipt, b.egress) , 并将验证信息证明=(b.header, b.ext, b.proof, b.egress) 发送给 H_A 私有区块链的验证者。 如果验证者接受此跨区块链交易, 它将更新 Egr_{HA}, 并等待 Polkadot 中继链操作。

(4) H_A -> H_B: Polkadot 中继链获取 Egr_{HA} 的 Merkle 树, 并将交易信息（区块信息）传输到 Igr_{HB} 的 Merkle 树, 以完成 H_A→H_B 的交叉区块链。

(5) P-> H_B: P 收到 H_A 跨区块链消息后, 通过其在 H_B 私有区块链上的帐户发起 H_B→H_A 跨区块链交易。

(6) D_{HB} -> H_B: H_B 私有区块链的整理者捕获来自 P 的跨区块链交易请求, 过程与步骤 3 相同。H_B 通过 Polkadot 锁定 P EHR 的区块。Egr_{HB} 将被更新。

(7) $H_B \rightarrow H_A$: Polkadot 中继链获得 E_{grHB} 的 Merkle 树, 并将交易信息 (区块信息) 传输到 I_{grHA} 的 Merkle 树, 以完成 $H_B \rightarrow H_A$ 的交叉区块链。 H_B 不再有权限操作 P 的 EHR 数据, D_{HA} 可以获取 P 的 EHR 私有数据。

4.3.4 算法设计

方案总体算法设计如下:

Input:

The ID of the Patient, IDP ;

The operation of EHR on CEPS, $Oper$;

The index of EHR, $Index$.

Output:

The result of $Oper$;

- 1: $Oper == \text{Creation}$, Doctor create the EHR diagnostic information E , and get the $Index$ of E by function $index(E)$, then send E and $Index$ to CSP;
- 2: $Oper == \text{T ransfer}$, get $Index$ from other hospital, and get E form CSP;
- 3: $Oper == \text{Deletion}$, patient delete the E on CSP by using $Index$;
- 4: return ($E \parallel Index \parallel \text{Operation result (true} \parallel \text{false)}$).

当患者到医院 A 就诊时, 他/她首先在 CEPS 中创建医院私有区块链帐户。医生在诊断患者后, 创建电子健康记录, 并将电子健康记录的索引信息存储在医院私有区块链上。医生可以将完整的电子健康记录信息 (包括索引信息和诊断信息) 发送给云服务提供商。

当患者转移到医院 B 后, 如果医院 B 要获取患者的 EHR 信息, 则需要获取存储在医院 A 私有区块链中的电子健康记录索引信息。患者首先在 CEPS 中为医院 B 私有区块链创建一个区块链账户, 然后向医院 A 发起跨区块链请求。医院 B 锁定存储在区块中的患者电子健康记录的索引信息, 并发起跨区块链请求。医院 A 将电子健康记录的索引信息发送到医院 B 私有区块链。由于索引已被锁定, 因此医院 A 将不再能够对该索引信息进行操作。在获得电子健康记录的索引信息后, 医院 B 根据索引信息, 向云服务提供商获取完整的电子健康记录。

当患者要删除电子健康记录的诊断信息, 则通过 CEPS 将请求发送到保存电子健康记录的云服务提供商。收到请求后, 云服务提供商删除相应患者的电子健康记录索引信息。索引信息存储在医院私有区块链中, 但是电子健康记录仅存储在云服务提供商中。从云服务提供商中删除诊断信息后, 医院将无法再从索引信息中获取患者的电子健康记录。当医院希望通过跨区块链获取患者的电子健康记录时, 云服务提供商会通知他们该患者的数据已被删除。

4.4 安全性分析

4.4.1 51%攻击

区块链也不是完全安全的，它面临着许多持续的攻击。其中，51%的攻击是损坏最严重的攻击。当单个节点比其他网络节点具有更多的计算资源时，可能会发生攻击。区块链通过共识算法来防御 51%攻击，例如比特币的 POW 算法和以太坊的 POS 算法。Polkadot 中继链上使用的共识算法基于 Tendermint 和 Honey Badger BFT 的异步拜占庭容错算法。Polkadot 中的共识参与者是验证者。假设少于 1/3 的验证者是不可信任的，那么共识算法保证了安全性永远不会受到损害。这意味着 Polkadot 上的验证人（数量大于 2/3）将永远不会在同一高度提交冲突的块。因此，使用异步拜占庭容错算法的区块链将永远不会分叉。只有掌握整个网络的 2/3 个节点，才能执行 51%的攻击。

4.4.2 不诚实节点

（1）验证人节点。验证人被用来提名 Polkadot 中继链上的区块并获得奖励，因此验证人出于自身利益可能是不诚实的。验证人由提名人提名，而 Polkadot 使用提名权益证明（NPoS）方案来实施验证人的选举。传统 PoS 容易产生节点寡头的主要原因在于选择节点的方式上。NPoS 放弃了简单的以投票数排序来确定超级节点的方式。它是采用一种相对比较均等的原则选出验证节点，即基于所有提名人投票情况，按照票数均等原则，选出若干验证节点。验证人严格依赖于他们抵押的 Polkadot 代币。验证人的不当行为会受到惩罚，例如减少奖励。如果故意破坏网络的完整性，则验证人将失去部分甚至全部代币。NPoS 非常适合跨区块链应用程序，因为它允许所有持有代币的网络参与者持续参与，使验证人的数量不受限制，从而使网络中所有的操作都变得高效。

（2）收集人节点。收集人的不诚实行为将导致验证人抵押的代币减少。然后，验证人将拒绝帮助收集人在区块链中写入 EHR 信息，这样收集人即使进行了一次不诚实行为也没有下一次机会在进行不诚实行为。因此，收集人应始终避免不诚实的行为才能在 Polkadot 网络中有效运行。

4.4.3 恶意共谋

Polkadot 中继链的共识机制不会导致分歧，我们给出以下证明。Polkadot 中继链网络的节点数量假设见公式 4-1

$$\begin{cases} x_1, x_2, \dots, x_l & \frac{l}{s} < \frac{1}{3} \\ y_1, y_2, \dots, y_m & \frac{m}{s} < \frac{1}{3} \\ z_1, z_2, \dots, z_n & \frac{n}{s} < \frac{1}{3} \end{cases} \quad (4-1)$$

Polkadot 中继链网络中总共有 s 个验证人节点。 x_i 表示当前已提交的验证人的节点，所有节点均为诚实节点，其数量为 l ； y_i 代表 m 个不诚实节点； z_i 表示 n 个其他不确定的节点，即 $s = l + m + n$ 。

l 个节点已提交，这意味着 n' 个节点已向当前块提交了一个预提交票证，且 $n' < n$ 。需要证明的是，在提交了 l 个节点之后，其余未提交到当前块的 $m + n$ 个节点将不会在另一个块上达成共识。也就是说，我们需要证明：

$$\frac{m + n - n'}{s} < \frac{2}{3} \quad (4-2)$$

利用反证法，首先假设：

$$\frac{m + n - n'}{s} > \frac{2}{3} \quad (4-3)$$

在式(4-3)的情况下，可能会在当前共识过程中造成 Polkadot 中继链分叉，然后有式(4-4)的推断：

$$\frac{l + m + n - n'}{s} > \frac{2}{3} + \frac{l}{s} \Rightarrow l - \frac{n'}{s} > \frac{2}{3} + \frac{l}{s} \Rightarrow \frac{l}{s} + \frac{n'}{s} < \frac{1}{3} \quad (4-4)$$

上面我们提到，因为 l 个节点已经提交给当前块，所以有式(4-5)。

$$\frac{l + m + n'}{s} > \frac{2}{3} \quad (4-5)$$

而 $m/s < 1/3$ ，则 $(l + n')/s$ 必须大于 $1/3$ ，从而得出式(4-6)的结论。

$$\frac{m + n - n'}{s} < \frac{1}{3} \quad (4-6)$$

式(4-6)与式(4-3)矛盾，假设不成立，也就是说，不诚实的节点无法在另一个

块上达成共识，因此也不会导致 Polkadot 网络分叉。

4.5 花费分析

在 CEPS 中，存储电子健康记录的货币成本是由 Polkadot 中继链的跨链交易引起的。具体来说，一旦医生为患者生成了电子健康记录，就需要创建新的交易记录电子病历并保护其免受非法修改。在跨区块链获取电子健康记录时，需要支付一定数量的 Polkadot 代币。由于 Polkadot 交易成本与以太坊相同，且以太坊平均花费为 8 美分，因此 CEPS 的成本是可以接受的。此外，Polkadot 平台将为患者提供一种自我协商机制的功能，该机制可以调节货币费用以具有更理想的条件。

4.6 本章小结

本章我们提出了一个基于区块链的电子健康记录隐私保护方案，通过使用 Substrate 区块链开发框架构建医疗区块链对电子健康记录进行存储，确定了医疗区块链工作流程，以及对电子健康记录中的隐私数据进行脱敏处理，然后引入 Polkadot 中继链，将基于 Substrate 构建的医疗区块链接入 Polkadot，以实现电子健康记录的跨链传输，并对 Polkadot 中的角色进行了描述，对跨链流程进行了描述。

第五章 系统关键功能实现与测试

前文对基于区块链的电子健康记录隐私保护方案进行了分析和研究，本章中将对该方案的应用系统进行阐述，实现关键功能并对其进行测试。

5.1 整体架构

基于区块链的电子健康记录隐私保护方案的系统分层结构如图 5-1 所示，该系统采用三层结构设计。第一层为应用层，负责与用户交互，包括提供图形界面的前端页面和处理数据的后端程序。第二层为区块链层，由 Polkadot 中继链和 Substrate 区块构成，为电子健康记录提供存储和传输的功能。第三层是数据层，负责处理电子健康记录数据，并传输给区块链层存储。

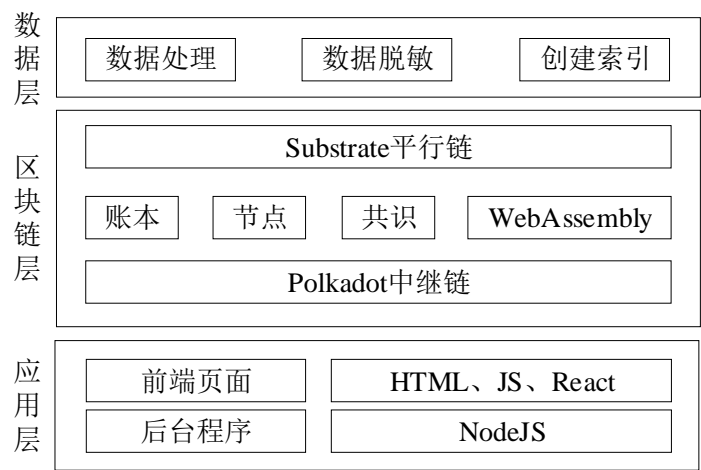


图 5-1 系统分层结构

应用层包括电子健康记录数据的可视化和医疗业务流程的可视化及其后端服务程序。用户通过应用层对电子健康记录进行操作，区块链的功能对用户来说是透明的，用户不需要知道区块链的工作原理。同时基于区块链开发的应用具有分布式的特点，又叫分布式应用，简称 DAPP，随着区块链的火热发展，DAPP 将会逐渐替代传统的基于中心化服务器的 APP。

区块链层中，本章设计的基于区块链的电子健康记录隐私保护系统是基于 Substrate 区块链开发框架构建的。基于 Substrate 区块链开发框架可以构建包含特定功能的区块链，本章基于 Substrate 区块链开发框架实现了医疗区块链的开发，采用 Rust 编程语言开发，除了实现区块链的基本功能外还实现了对电子健康记录

的隐私保护功能。然后将医疗区块链接入 Polkadot 中继链，实现电子健康记录的跨链传输，这种模式极大的提升了医疗区块链的可扩展性和实用性。

基于 Substrate 区块链开发框架构建的区块链中，当用户对电子健康记录进行任意操作时首先会向整个区块链网络广播本次操作，类似传统区块链网络中的发起交易，整个医疗区块链网络来对此次操作进行验证；然后验证通过，此次操作合法，将会打包区块，该区块包含电子健康记录的索引值和对应操作的枚举值，并将该区块上链；最后整个医疗区块链网络的其他节点验证该区块后同步到本地存储，这样一次对电子健康记录的操作就存储在区块链中。由于区块链是分布式的、按时间顺序存储的链式结构，所以区块链中的数据难以被篡改并且可以按时间顺序追溯数据，使得电子健康记录的隐私得到了保护。

数据层主要功能有两个：接收应用层用户发送的数据并对数据进行处理随后发送至区块链层；接收区块链层发送来的数据，做相应处理后发送至应用层。数据层包含两个部分：后端程序数据层和区块链数据层。后端程序数据层负责接收应用层用户发送的数据，对数据进行优化，包括输入异常检测、字符串格式化、创建时间戳、创建索引等；区块链数据层负责接收区块链层向应用层发送的数据，对数据进行处理，包括隐私数据脱敏、时间戳转换等。

5.2 关键功能实现

本小节将实现前文提出的基于区块链的电子健康记录隐私保护方案的关键功能，包括区块链基本功能，电子健康记录操作，数据脱敏，跨链功能以及其他相关功能。

5.2.1 开发环境

系统的开发环境包括 Substrate 区块链开发环境、Web 客户端开发环境和 Polkadot 跨链开发环境。本系统开发使用的编程语言及相关开发工具如表 5-1 所示。

所有的开发工作都是在 Ubuntu 18.04 操作系统下完成的，Substrate 区块链开发环境使用 Rust 语言，Web 客户端开发环境中前端使用 JavaScript 及 React 完成，后端使用 Node.JS 完成，Polkadot 跨链开发环境使用 Rust 语言，代码编辑器使用的是 Visual Studio Code，并用 MongoDB 数据库替代方案中的云服务提供商。

表 5-1 编程语言及相关开发工具

名称	版本	描述
Rust	1.42.0-nightly	Substrate 开发语言
rustup	1.21.1	Rust 语言版本管理工具
cargo	1.42.0-nightly	Rust 语言包管理工具
React	16.12.0	构建用户界面的 JavaScript 库
Node.js	12.14.1	JavaScript 运行环境
yarn	1.21.1	Node.JS 依赖管理工具
npm	6.13.4	Node.JS 包管理工具
Substrate	2.0.0	区块链开发框架
Visual Studio Code	1.41.1	代码编辑器

5.2.2 基于 Substrate 的区块链实现

区块链的主要功能包括节点信息显示，区块高度显示，区块链账号管理，交易等。

5.2.2.1 开发框架搭建

(1) Substrate 提供了一键部署的安装脚本，通过执行命令：

```
curl https://getsubstrate.io -sSf | bash -s -- --fast
```

可以快速安装 Substrate 所需要的必要组件，这些组件包括：Cmake、pkg-config、OpenSSL、Git、Rust。

(2) Substrate 源码获取

Substrate 2.0.0 的源码由 github 托管，可以通过 git 指令将 Substrate 源码拷贝到本地，指令如下：

```
Git clone https://github.com/paritytech/substrate.git
```

(3) Substrate 在 github 中托管了 Substrate 开发者可以使用的最小的组件，其中 <https://github.com/substrate-developer-hub/substrate-node-template> 是区块链网络开发的最小组件，本模型就是基于这个组件开发的。可以通过以下命令将代码拷贝到本地：

```
Git clone https://github.com/substrate-developer-hub/substrate-node-template.git
```

(4) 一个完整的 Substrate 模板结构分为五个部分，其代码结构如下所示。

```

use support::{decl_module, decl_storage, decl_event, dispatch::Result};

pub trait Trait: system::Trait {}

decl_storage! {}

decl_module! {}

decl_event!();

```

第一部分为 `use support` 导入模块：类似于 java 预言的 `import`，将需要使用的到的包导入到本项目中。

第二部分为 `pub trait Trait: system::Trait` 配置接口：所有的 `Runtime` 类型和常量都在这里。如果 `Pallets` 依赖于其他特定的 `Pallets`，那么它们的配置特征应该被添加到继承的特征列表中。如果需要获取账户余额的接口，需要将配置改为 `pub trait Trait: balances::Trait`，其中 `Balance` 是 `Substrate` 提供的表示账户余额的模块。

第三部分为 `decl_storage! {}` 宏定义：在这里面声明本项目所需要存储在区块链中的数据结构。允许使用类型安全的 `Substrate` 存储数据库，因此可以在区块之间保存数据。

第四部分为 `decl_module! {}` 可调用函数声明及实现：这部分声明本模型所使用的功能函数以及功能函数的逻辑实现。这部分定义了公开的可调用函数，并在整个区块链执行过程中编辑了此函数所采取的操作的实现过程。

第五部分为 `decl_event!()` 事件声明：事件是一种简单的方法，用来报告客户端用户、数据应用程序或区块链用户希望获取但很难发现的特定条件和情况。客户端通过监听区块中的事件来更新链下的存储状态或与用户交互。

5.2.2.2 节点信息

节点信息是区块链的节点展示。展示的内容包括当前运行的区块链版本号，还有构造区块和区块上链的功能。

```

impl client_api::Core<Block> for Runtime {
    // 当前运行的区块链版本
    fn version() -> RuntimeVersion { VERSION }

    // 区块上链
    fn execute_block(block: Block) {
        Executive::execute_block(block)
    }

    // 初始化区块
    fn initialize_block(header: <Block as BlockT>::Header) {
        Executive::initialize_block(header)
    }
}

```

5.2.2.3 区块信息

区块链的一个区块应该包含区块头和存储内容两个部分，区块的结构体代码如下：

```
pub struct Block<Header, Extrinsic: MaybeSerialize> {
    /// 区块头
    pub header: Header,
    /// 区块交易信息
    pub extrinsics: Vec<Extrinsic>,
}
```

构造区块的代码如下：

```
impl block_builder_api::BlockBuilder<Block> for Runtime {
    // 提交交易
    fn apply_extrinsic(extrinsic: <Block as BlockT>::Extrinsic) -> ApplyResult {
        Executive::apply_extrinsic(extrinsic)}
    // 区块构造完成，返回区块头
    fn finalize_block() -> <Block as BlockT>::Header {
        Executive::finalize_block()}
    // 将外部交易信息构造成加入区块链的 Extrinsic
    fn inherent_extrinsics(data: InherentData) -> Vec<<Block as BlockT>::Extrinsic> {
        data.create_extrinsics()}
    // 检查外部交易
    fn check_inherents(block: Block, data: InherentData) -> CheckInherentsResult {
        data.check_extrinsics(&block)}
    // 构造区块头哈希值
    fn random_seed() -> <Block as BlockT>::Hash {
        RandomnessCollectiveFlip::random_seed()}
}
```

代码中包含了构造区块的完整过程，包括提交交易内容，检查交易内容，将交易内容构造成存储于区块中的结构体，构造区块头的哈希值，完成区块的构造。

5.2.2.4 交易

通过调用 Substrate 的 Exectutive 模块中的 validate_transaction 函数来实现交易。

```
impl client_api::TaggedTransactionQueue<Block> for Runtime {
    fn validate_transaction(tx: <Block as BlockT>::Extrinsic) -> TransactionValidity {
        Executive::validate_transaction(tx)
    }
}
```

Substrate 将 Extrinsic 分为了两种：UncheckedExtrinsic 和 CheckedExtrinsic。当一个交易的状态从 UncheckedExtrinsic 转变到 CheckedExtrinsic 就可以构造区块并上链了。

5.2.2.5 共识

Substrate 支持自定义的、可插拔的共识算法，可以随时更换，并提供了三种基础共识算法：Aura、BABE 和 PoW。默认使用 Aura 算法，通过以下代码使用 Aura：

```
impl aura_primitives::AuraApi<Block, AuraId> for Runtime {
    fn slot_duration() -> u64 {
        Aura::slot_duration()
    }
    fn authorities() -> Vec<AuraId> {
        Aura::authorities() }
}
```

5.2.3 电子健康记录操作实现

医疗区块链上的关键功能是对电子健康记录的各种操作，定义了电子健康记录的四种基本操作，用以标记电子健康记录的上次操作状态，对电子健康记录的操作就是一个状态机，这个状态机包括创建、删除、修改、查看共四种状态，并在这四个状态间不断转换。

这四种状态定义如下所示。

```
// EHR 状态
enum EHRstatus {
    Establish(String),    // 创建
    Deletion(String),     // 删除
    Modification(String), // 修改
    Examination(String),  // 查看}
```

区块链中电子健康记录结构主要体信息如表 5-2 所示。

表 5-2 电子健康记录结构主要体信息

变量名	数据类型	含义
EHR_index	Hash	索引
EHR_name	String	姓名
EHR_IDCard	String	身份证号
EHR_sex	char	性别
EHR_age	U32	年龄
EHR_phone	String	电话
EHR_address	String	地址
EHR_diagnosises	Vec	诊断信息
EHR_medicine	Vec	药物信息
EHR_status	Enum	电子健康记录状态
EHR_abstract	String	电子健康记录摘要
EHR_random	String	随机数

5.2.3.1 电子健康记录创建

电子健康记录创建代码如下所示：

```
fn create_claim(origin, proof: Vec<u8>) {  
    // 身份验证.  
    let sender = ensure_signed(origin)?;  
    // 验证信息  
    ensure!(!Proofs::::exists(&proof), "This proof has already been claimed.");  
    // 调用'system'中的 pallet 获取当前区块高度  
    let current_block = <system::Module<T>>::block_number();  
    // 保存  
    Profs::::insert(&proof, (sender.clone(), current_block));  
    Self::deposit_event(RawEvent::ClaimCreated(sender, proof));  
}
```

电子健康记录创建流程如图 5-2 所示，具体流程如下：

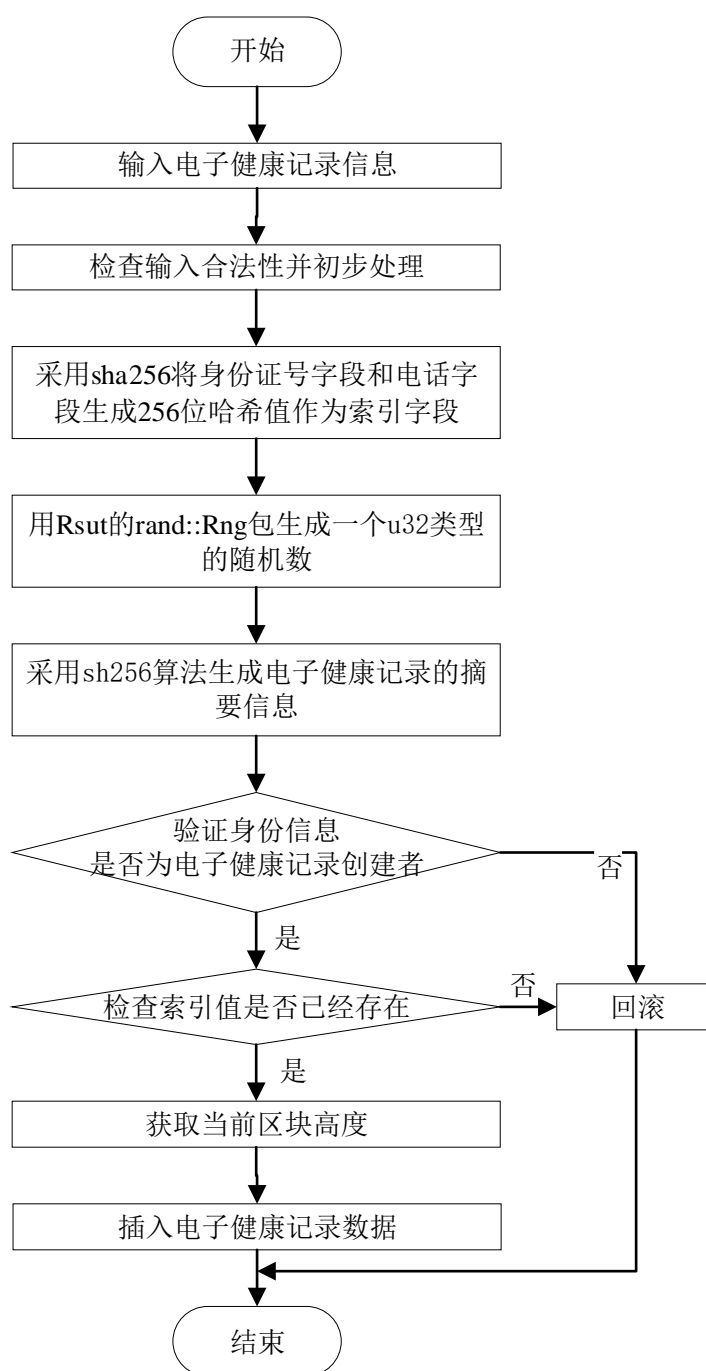


图 5-2 电子健康记录创建操作流程

(1) 录入电子健康记录数据，前端对录入的数据进行合理性检测，通过合理性检测后传输到区块链基本模组 `template.rs` 中，用 `EHRInfo` 结构体接收。

(2) 拼接身份证号字段和电话字段，并用 `SHA256` 算法将拼接后的字符串生成 256 位的哈希值，并更新结构体 `EHRInfo` 中的 `index` 字段。

(3) 更新 `EHRStatus` 字段为 `Establish(timestamp)`,并存储在数据库中，这里数

数据库模仿云存储环境，即电子健康记录数据完整存放的位置。

(4) 通过身份验证函数 `pub fn ensure_signed<OuterOrigin, AccountId>(o: OuterOrigin) -> Result<AccountId, Error>` 验证当前操作者身份，`OuterOrigin` 是当前操作的身份信息，`AccountId` 是当前操作者的区块链账号，`Result` 是 `Sustrate` 区块定义的标准返回结构，如果验证失败则回滚之前的操作。

(5) 检测当前 `EHRInfo` 的 `index` 字段是否在链上已经存在，如果存在则回滚之前所有操作。

(6) 获取当前区块高度，构造区块，在区块中包含以下数据：索引值和创建操作值。

(7) 数据上链后，构造一个 `Substrate Runtime Event` 并向客户端发送，通知客户端电子健康记录创建操作上链成功。`Substrate` 的 `Runtime` 模块要想向外部实体（例如用户，区块链浏览器或者 `dApps`）通知 `Runtime` 内部的变化时，需要通过发送 `Event` 的方式。`Substrate Front-end` 通过接口对 `Event` 进行监听，使得客户端可以获得 `Runtime` 的内部变化并展示。

5.2.3.2 点击健康记录查看

电子健康记录查看流程如图 5-3 所示，具体流程如下：

(1) 输入身份证号和电话号，并拼接这两个字符串，使用 `Hash256` 算法生成索引哈希值。

(2) 在区块链中检索索引值是否存在于某一区块中，如果不存在向客户端返回结果并结束。

(3) 如果索引值在区块链中存在，在数据库中检索索引值是否存在，如果不存在向客户端返回结果并结束。

(4) 如果在数据库中检索到索引字段，提取出对应的电子健康记录的完整数据，识别出电子健康记录中的隐私字段，并对各字段按照脱敏策略做脱敏处理，准备向客户端发送。

(5) 向客户端发送脱敏后的电子健康记录数据，并在前端页面展示，医生不会看到患者的隐私数据。

(6) 获取当前区块链中的区块高度，并构造区块，包含电子健康记录的索引值和电子健康记录查询操作值。

(7) 数据上链后，构造一个时间 `Event` 并向客户端发送，通知客户端电子健康记录查看操作上链成功。

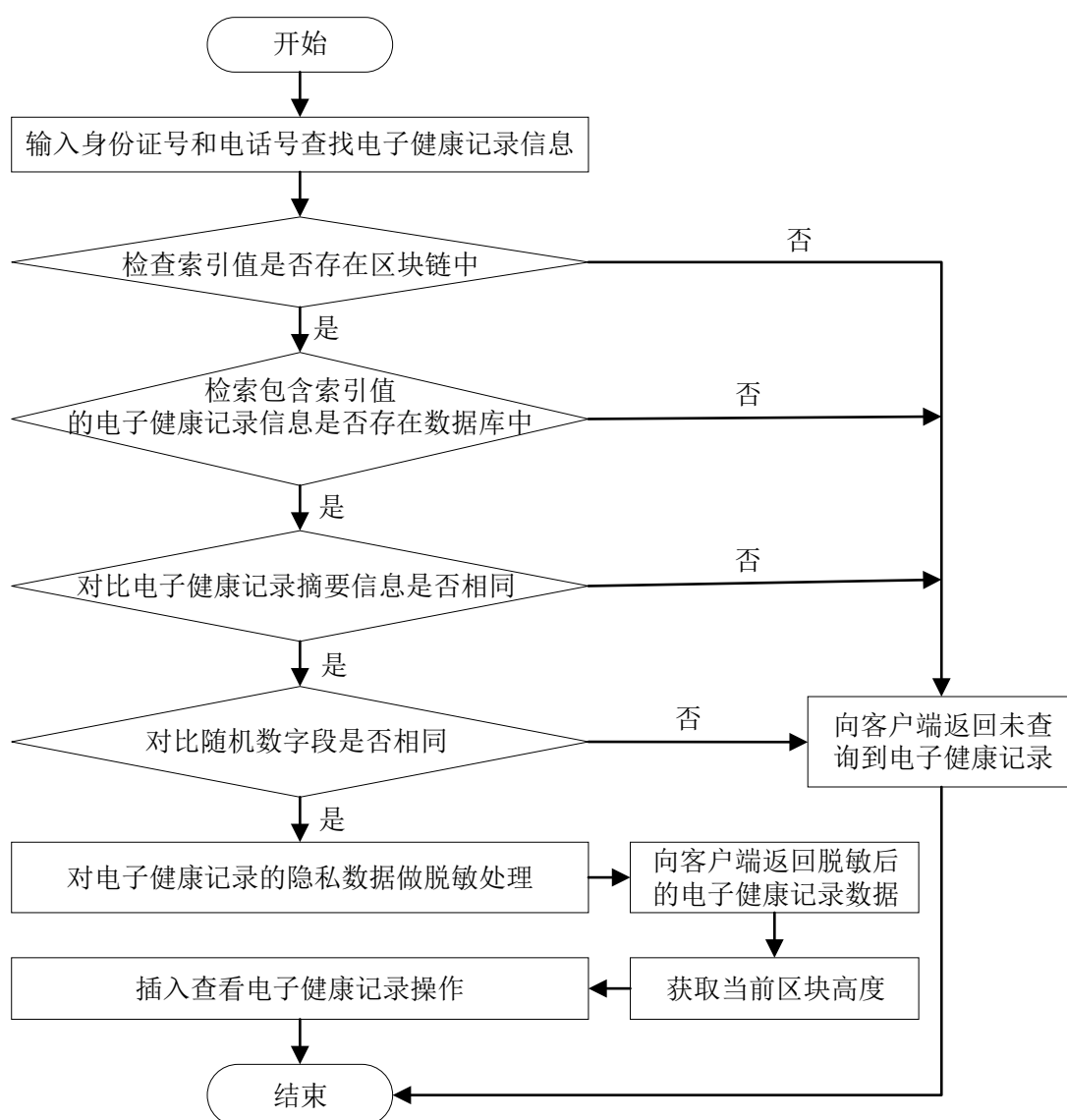


图 5-3 电子健康记录查看操作流程

5.2.3.3 电子健康记录删除

电子健康记录删除流程如图 5-4 所示，具体流程如下：

- (1) 输入身份证号和电话号，并拼接这两个字符串，使用 Hash256 算法生成索引哈希值。
- (2) 在区块链中检索索引值是否存在，如果不存在向客户端返回结果并结束。
- (3) 如果索引值在区块链中存在，在数据库中检索索引值是否存在，如果不存在向客户端返回结果并结束。
- (4) 如果在数据库中检索到索引字段，提取出对应的电子健康记录数据，识别出电子健康记录中的隐私字段，并对各字段按照脱敏策略做脱敏处理，删除索

引字段后保存进数据库，覆盖原来的电子健康记录。

(5) 向客户端发送删除结果。

(6) 获取当前区块高度，构造区块，包含索引值和删除操作值。

(7) 数据上链后，构造一个时间 Event 并向客户端发送，通知客户端电子健康记录查看删除上链成功。

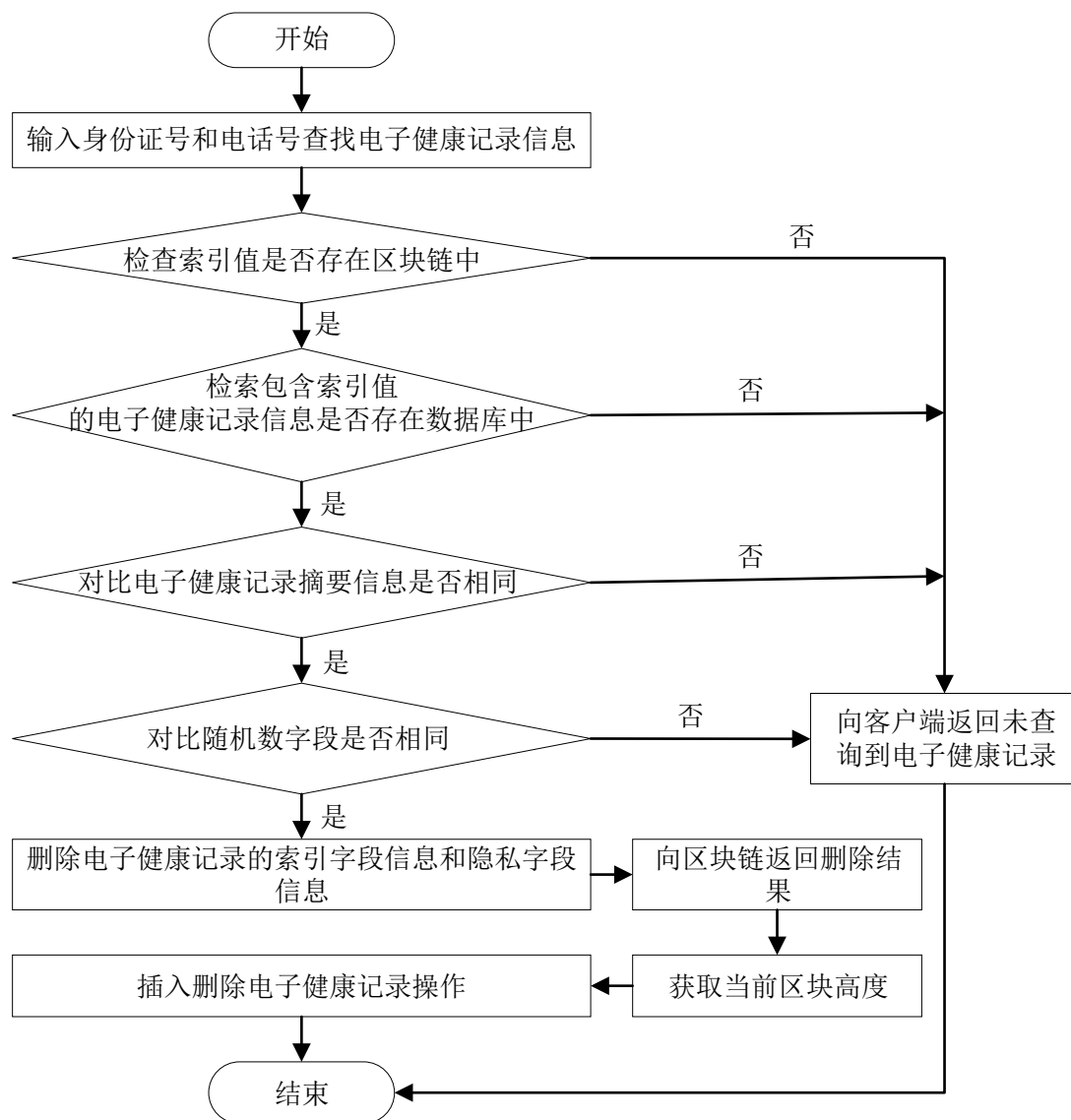


图 5-4 电子健康记录删除操作流程

5.2.3.4 电子健康记录修改

电子健康记录的删除流程是电子健康记录查看和创建的流程拼接，具体流程如下：

(1) 输入身份证号和电话号，并拼接这两个字符串，使用 Hash256 算法生成索引哈希值。

(2) 在区块链中检索索引值是否存在，如果不存在向客户端返回结果并结束。

(3) 如果索引值在区块链中存在，在数据库中检索索引值是否存在，如果不存在向客户端返回结果并结束。

(4) 如果在数据库中检索到索引字段，提取出对应的电子健康记录数据，识别出电子健康记录中的隐私字段，并对各字段按照脱敏策略做脱敏处理，删除索引字段后保存进数据库，覆盖原来的电子健康记录。

(5) 向客户端发送删除结果。

(6) 获取当前区块高度，构造区块，包含索引值和删除操作值。

(7) 数据上链后，构造一个时间 Event 并向客户端发送，通知客户端电子健康记录查看删除上链成功。

5.2.4 数据脱敏实现

电子健康记录中包含大量患者的个人隐私数据，包括姓名、电话、住址、年龄、性别、身份证号等，为了不暴露患者的隐私，需要将这些数据做脱敏处理，不让其他人看到完整的患者隐私数据但是又能使脱敏后的数据有一定的查看或统计作用。脱敏流程如图 5-5 所示。

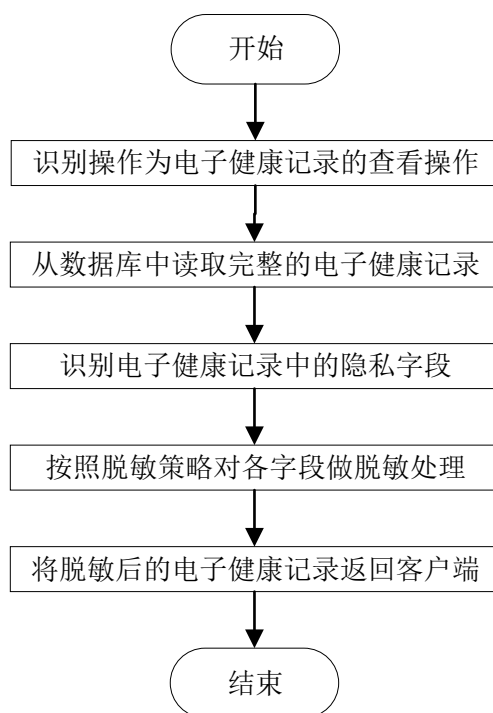


图 5-5 电子健康记录隐私数据脱敏流程图

电子健康记录个字段脱敏的函数如下所示：

```
// 姓名脱敏
pub fn name_desensitization(surname: String, givenname: String) -> String

// 电话脱敏
pub fn phone_desensitization(phone: String) - String

// 身份证脱敏
pub fn idcard_desensitization(idcard: String) - String

// 地址脱敏
pub fn address_desensitization(address: String) - String

// 年龄脱敏
pub fn age_desensitization(age: u32) - String
```

其中身份证脱敏的代码如下所示，首先将身份证按照第四章的描述，将身份证号划分为三个区间，然后对三个区间做相应的脱敏操作，最后将三个区间拼接成脱敏后的字符串并返回。

```
pub fn idcard_desensitization(idcard: String) - String{
    let idcard_des = "".to_string();
    let len = idcard.len();
    let idcardchars = idcard.chars();
    for i in 0..6{idcard_des.push(phonechars[i]);}
    for i in 6..14{idcard_des.push('*');}
    for i in 14..18{idcard_des.push(phonechars[i]);}
    idcard_des}
```

5.2.5 跨链实现

基于 Substrate 构建的医疗区块链要接入 Polkadot 实现跨链，引入 polkadot 下的 parachain module，代码如下：

```
extern crate polkadot_parachain as parachain;
extern crate polkadot_primitives as primitives;
extern crate polkadot_collator as collator;
...
use parachain::codec::{Encode, Decode};
use primitives::parachain::{HeadData, BlockData, Id as ParaId, Message};
use collator::{InvalidHead, ParachainContext, VersionInfo};
```

为了将 Substrate 区块链中的节点作为 Polkadot 的收集人节点，需要实现如下

代码:

```
fn produce_candidate<I: IntoIterator<Item=(ParaId, Message)>>>(
    &self,
    last_head: HeadData,
    _ingress: I,) -> Result<(BlockData, HeadData), InvalidHead>{
    let next_body = AdderBody {state,add,};
    let next_head = ::adder::execute(adder_head.hash(), adder_head,
    &next_body).expect("good execution params; qed");
    let encoded_head = HeadData(next_head.encode());
    let encoded_body = BlockData(next_body.encode());
    Ok((encoded_body, encoded_head))}
```

函数 `produce_candidate` 用于配置作为 Polkadot 收集人的节点的平行链上下文环境, 主要流程是从本地读取到创世区块的区块头和区块体并输出。

Polkadot 中继链跨链实现是基于跨链消息传递协议 (Cross-chain Message Passing, XCMP) 的, 其主要实现流程如下:

(1) 假设有两条平行链 A 和 B, 且都维护有跨链进口和出口队列, 分别称为 Ingress 和 Egress, 出口队列负责发送跨链消息, 进口队列负责接收跨链消息。

(2) 平行链 A 向平行链 B 跨链, 首先平行链 A 的收集人节点会将这个新的跨链消息以及目标平行链和一个时间戳放入其出站消息队列中。

(3) 侧链 B 的收集人节点定期对所有其他收集人节点进行 ping 操作, 以请求新消息 (按目标平行链字段过滤)。当平行链 B 的收集人进行下一次 ping 操作时, 将会在平行链 A 的收集人上看到此新消息, 并将其添加到其自己的进站队列中, 以处理到下一个区块中。

(4) 平行链 A 的验证人还将读取出站队列并获取消息。平行链 B 的验证人将执行相同的操作。这样, 他们便可以验证消息传输是否发生。

(5) 当平行链 B 的收集人正在构建其区块链中的下一个区块时, 它将处理进站队列中的新消息以及可能已找到或者接收到的任何其他消息。

(6) 在消息处理和区块构建期间, 平行链 B 将会完成跨链数据的处理。

(7) 收集人现在将此区块交给验证程序, 验证程序将验证此消息是否已处理。如果消息已处理且该块的所有其他信息均有效, 则验证人会将针对平行链 B 的该块包加入到中继链中。

5.3 测试

5.3.1 测试环境

为了模拟实际的医院跨链场景，一共需要四个实体：Polkadot 中继链，医院 A 区块链，医院 B 区块链和数据库。测试环境采用了两台 PC 一台操作系统为 ubuntu 18.04，用于搭建 Polkadot 中继链和 MongoDB 数据库，一台为 windows 10 并搭载了 VMware 虚拟机，VMware 安装了两个 ubuntu 18.04 虚拟机，用于搭建两个医院的区块链。具体测试环境内容如表 5-3 所示。

表 5-3 测试环境

实体	系统	IP 地址	描述
Polkadot 中继链	ubuntu 18.04	192.168.0.105	跨链传输
数据库	ubuntu 18.04	192.168.0.105	电子健康记录存储
医院 A 区块链	VMware ubuntu 18.04	192.168.0.112	医院 A 保存索引
医院 B 区块链	VMware ubuntu 18.04	192.168.0.113	医院 B 保存索引

5.3.2 功能测试

功能测试包括两个部分，区块链功能和电子健康记录管理功能。

区块链功能测试用例如表 5-4 所示，包括启动区块链、启动区块链浏览器、构造区块、链上交易、账户信息展示、区块内容展示、接入中继链。

表 5-4 区块链功能测试用例

测试功能	测试步骤	测试结果
启动区块链	命令行启动区块链，并能构造创世区块	成功
启动区块链浏览器	配置文件链接到区块链节点所在 IP 地址并启动区块链浏览器	成功
构造区块	区块链自动构造区块并上链	成功
链上交易	在两个账号之间交易代币	成功
账户信息展示	查看区块链账户地址信息	成功
区块内容展示	查看区块链中具体某个区块包含的内容	成功
接入中继链	将 Substrate 区块链接入 Polkadot 中继链	成功

电子健康记录管理功能测试用例如表 5-5 所示，包括电子健康记录创建、电子健康记录查看、电子健康记录修改、电子健康记录删除、电子健康记录跨链。

表 5-5 电子健康记录功能测试用例

测试功能	测试步骤	测试结果
EHR 创建	创建 EHR 并将索引存储在区块链汇总	成功
EHR 查看	根据区块链汇总存储的索引查看 EHR	成功
EHR 修改	根据区块链汇总存储的索引修改 EHR	成功
EHR 删除	根据区块链汇总存储的索引删除 EHR	成功
EHR 跨链	EHR 在两个区块链间跨链	成功

区块链启动, 结果如图 5-6 所示。图中表示了区块链正常启动, 并且正确的生成了创世区块的哈希值, 这个哈希值将会用于接入 Polkadot。

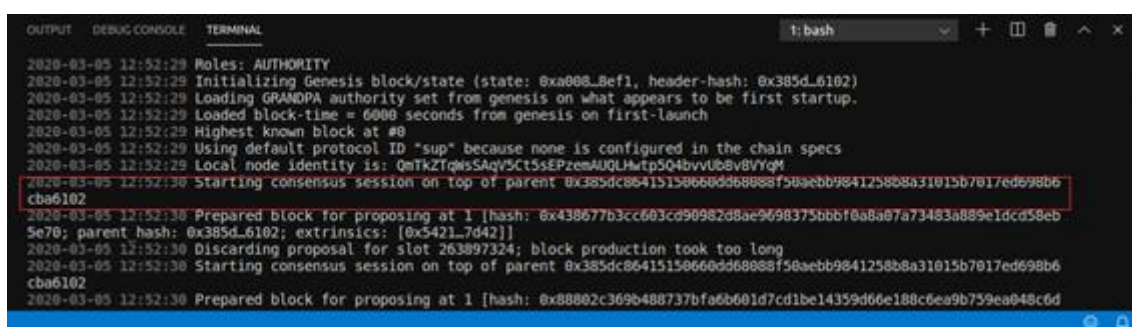


图 5-6 启动区块链构造创世区块

区块链功能如图 5-7 所示。图 5-7(a)中展示了节点信息、当前区块高度、下一个区块高度以及区块链账户信息。图 5-7(a)表示当前区块高度为 991, 表示目前区块链中已经有 991 个区块, 下一个区块为 992。Balances 为区块链上账户信息, 包括账户名, 账户地址, 以及账户余额。图 5-7(b)中当前区块高度为 1001, 与图 5-7(a)中的区块高度相差 10 个, 表明这段时间里区块链已经产生了 10 个区块并上链。

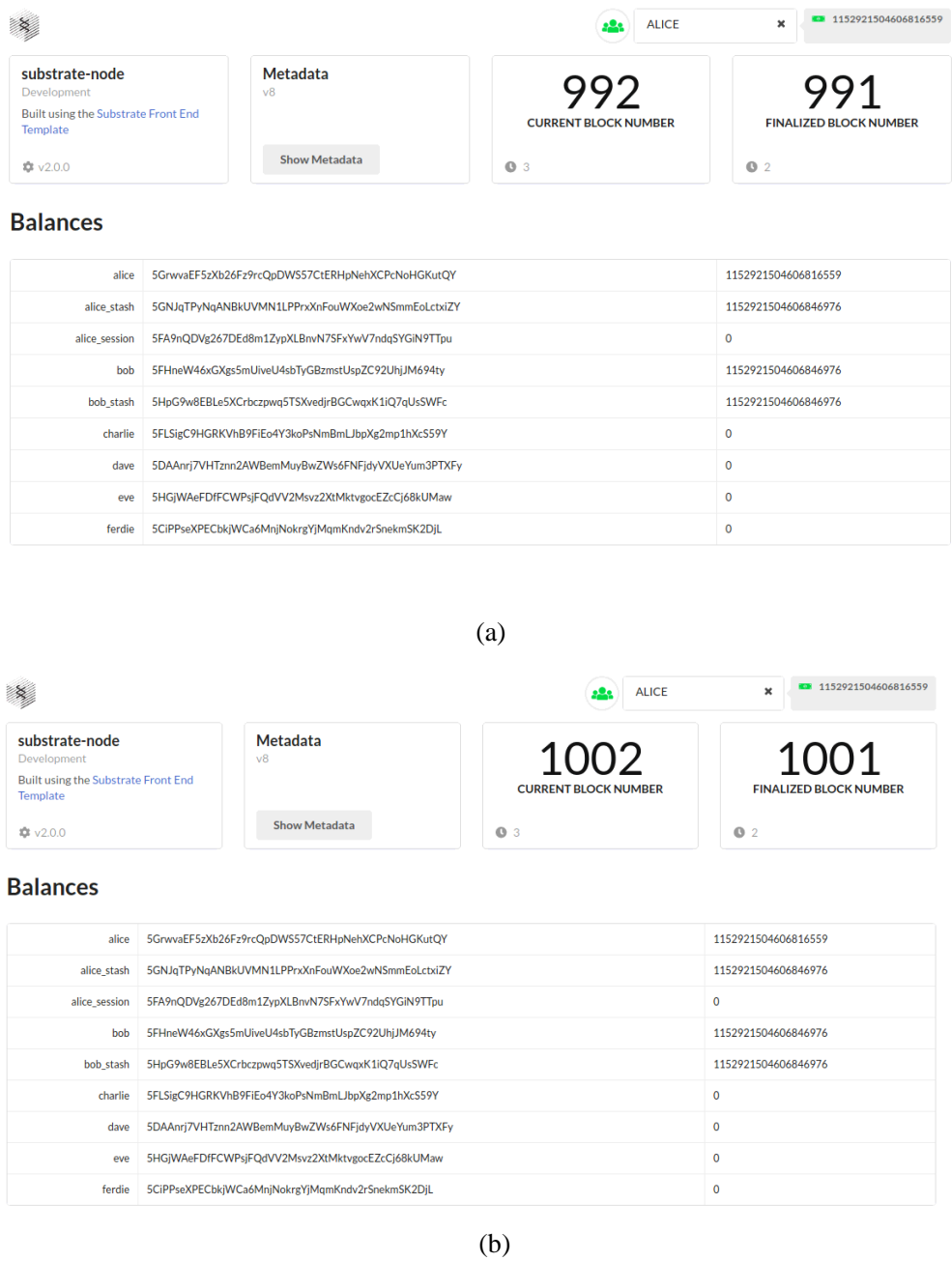


图 5-7 区块链功能

图 5-8 展示了当前区块链中所有账户的持有代币的金额,这些代币可以在不同的账户之间交易转移。

alice	5GrwvaEF5zXb26Fz9rcQpDWS57CtERHpNehXCPcNoHGKutQY	1152921504606816559
alice_stash	5GNJqTPyNqANBkUVMN1LPPrxXnFouWXoe2wNsmmEoLctxiZY	1152921504606846976
alice_session	5FA9nQDVg267DEd8m1ZypXLBnvN75FxywV7ndqSYGIN9TTpu	0
bob	5FHneW46xGXgs5mUiveU4sbTyGBzmstUspZC92UhjJM694ty	1152921504606846976
bob_stash	5HpG9w8EBLe5XCrbczpq5TSXvedjrBGCwqxK1iQ7qUsSWFc	1152921504606846976
charlie	5FLSigC9HGRKvHb9FiEo4Y3koPsNmBmLJbpXg2mp1hXcS59Y	0
dave	5DAAnrj7VHTznn2AWBemMuyBwZW6FNfjdyVXUeYum3PTXFy	0
eve	5HGjWAeFDfCWPsjFQdVv2Msvz2XtMktvgocEzcCj68kUMaw	0
ferdie	5CiPPseXPECbkjWCa6MnjNokrgYjMqmKndv2rSnekmSK2DJL	0

图 5-8 区块链账户信息

图 5-9 展示了地址为 5GrwvaEF5zXb26Fz9rcQpDWS57CtERHpNehXCPcNoHGKutQY 的用户 alice 向地址为 5FLSigC9HGRKvHb9FiEo4Y3koPsNmBmLJbpXg2mp1hXcS59Y 的用户 charlie 转移了 10000 个单位的代币。保存此次交易的区块的哈希值为: #0x2d97ab340bde3f9219f1036705811c2183b81ea33de58019c968f377f83e4d0a。

alice	5GrwvaEF5zXb26Fz9rcQpDWS57CtERHpNehXCPcNoHGKutQY	1152921504605806418
alice_stash	5GNJqTPyNqANBkUVMN1LPPrxXnFouWXoe2wNsmmEoLctxiZY	1152921504606846976
alice_session	5FA9nQDVg267DEd8m1ZypXLBnvN75FxywV7ndqSYGIN9TTpu	0
bob	5FHneW46xGXgs5mUiveU4sbTyGBzmstUspZC92UhjJM694ty	1152921504606846976
bob_stash	5HpG9w8EBLe5XCrbczpq5TSXvedjrBGCwqxK1iQ7qUsSWFc	1152921504606846976
charlie	5FLSigC9HGRKvHb9FiEo4Y3koPsNmBmLJbpXg2mp1hXcS59Y	10000
dave	5DAAnrj7VHTznn2AWBemMuyBwZW6FNfjdyVXUeYum3PTXFy	0
eve	5HGjWAeFDfCWPsjFQdVv2Msvz2XtMktvgocEzcCj68kUMaw	0
ferdie	5CiPPseXPECbkjWCa6MnjNokrgYjMqmKndv2rSnekmSK2DJL	0

Transfer

To	5FLSigC9HGRKvHb9FiEo4Y3koPsNmBmLJbpXg2mp1hXcS59Y
Amount	10000
<input type="button" value="Send"/>	

Upgrade Runtime

Wasm File	<input type="button" value="Choose File"/> No file chosen
<input type="button" value="Upgrade"/>	

图 5-9 账号交易

图 5-10 展示了电子健康记录的创建功能。图 5-10(a)中完善了电子健康记录信息后通过身份证号和电话号生成了索引值:

0x5a323fe8f67d05d24d8ca8fd023b79b3d397997c705f40ae8e7704a115b86fb7

该索引值未保存在区块链中可以作为当前电子健康记录的索引。图 5-10(b)中电子健康记录被创建成功并成功的将索引值保存在区块链中, 创建者为:

5GrwvaEF5zXb26Fz9rcQpDWS57CtERHpNehXCPcNoHGKutQY

区块高度为 2206, 即该索引值保存在第 2206 个区块中。

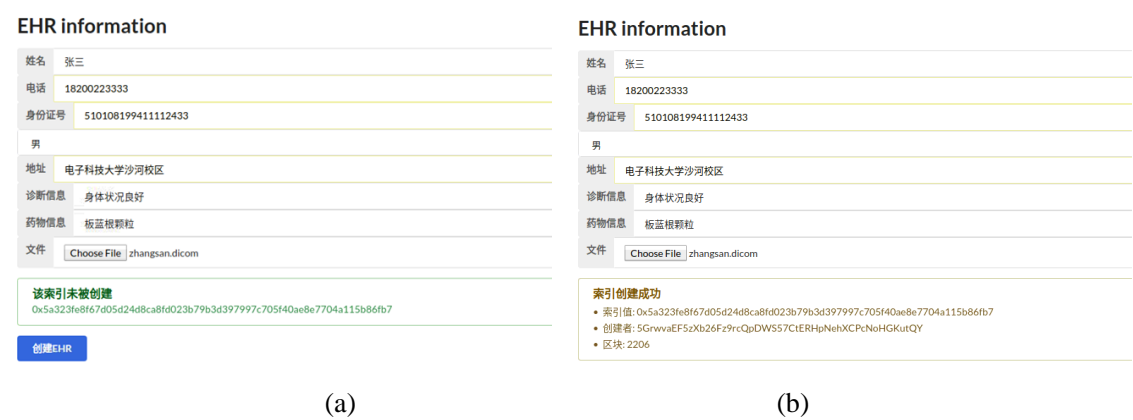


图 5-10 电子健康记录创建展示

图 5-11 展示了高度为 2206 区块的内容，该区块中正确的保存了创建的电子健康记录的索引值。

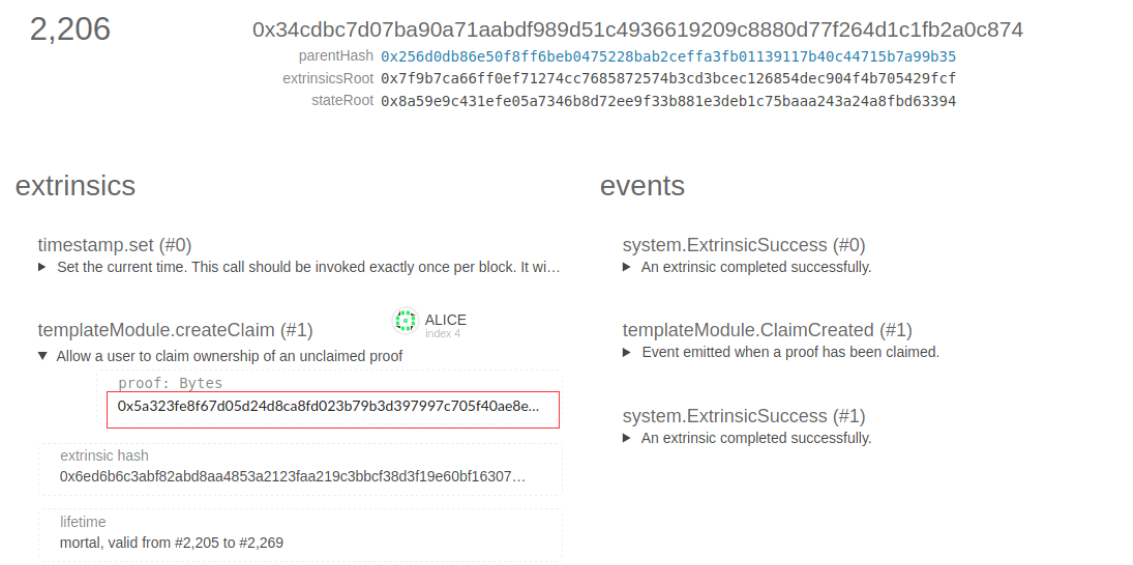


图 5-11 区块高度为 2206 的区块信息展示

电子健康记录信息查看和删除如图 5-12 所示，图 5-12(a)为查看界面，能完整显示电子健康记录信息，并提供修改、删除电子健康记录的功能。图 5-12(b)为删除界面，删除了区块链上的索引值与数据库中完整的电子健康记录索引信息的联系，不能通过该索引再找到完整的电子健康记录信息。

EHR information

修改EHR

EHR information

该索引删除成功

修改EHR

删除EHR

(a)
(b)

图 5-13 展示了基于 Substrate 构建的医疗区块链加入 Polkadot 的过程，本文测试中在 Polkadot 中继链中加入了两个医疗区块链实例用以模拟电子健康记录信息在两个区块链间传输的场景。图 5-13 展示了医疗区块链加入 Polkadot 的过程，这个过程中需要医疗区块链提供创世区块哈希值和编译为 wasm 版本的可执行文件，这样 Polkadot 才能在自己的网络中找到这个医疗区块链的 Collator 节点。图 5-14 中展示了加入 Polkadot 的 parachain 的数量，这里的 parachain 即指文本构建的医疗区块链。

[illegible]

图 5-13 区块链加入 Polkadot

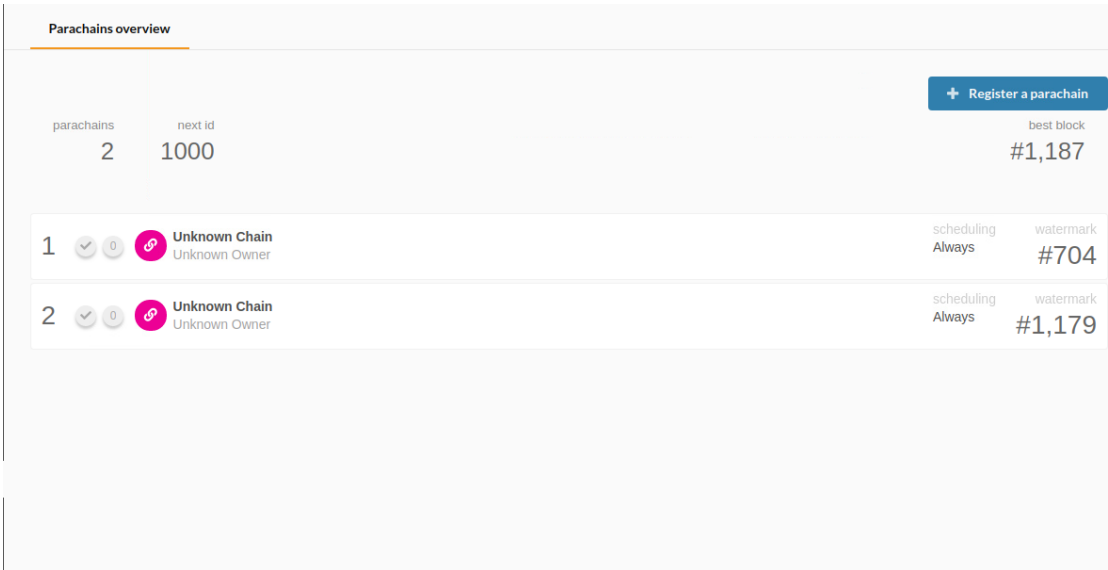


图 5-14 Polkadot 中平行链查看

5.3.3 性能测试

5.3.3.1 区块链性能测试

基于 Substrate 构建的区块链，从两个方面来提高区块链的性能。第一个方面是 Substrate 采用 Rust 语言开发：Rust 语言是一种系统语言，语法类似于 C++，执行速度也与 C++ 一样快，相较于其他主流区块链开发语言而言，Rust 有较快的执行速度，较高的安全性。

第二个方面是 Substrate 采用了 WebAssembly(简称 wasm)技术，将区块链代码分成了本地版本和 wasm 版本两个版本，wasm 版本代码的执行速度要比本地版本的执行速度慢。当一个区块链节点的本地版本和 wasm 版本代码一致时，运行本地版本代码，这样会具有较高的执行速度，而当 wasm 版本高于本地版本时，会执行 wasm 版本代码，并在执行过程中更新本地版本代码。图 5-15 展示了随着时间的推移，当区块链网络中运行 wasm 版本代码的节点所占比例不同时，区块链吞吐量的变化，区块链中包含 10 个节点。从图中可以看出，刚开始时，执行 wasm 版本代码节点的比例越高，每秒交易数量就越少，这是因为 wasm 版本代码的运行速度要地本地版本造成的，但随着时间的推移，当所有执行 wasm 版本代码的节点都将本地代码更新至与 wasm 版本代码一致时，整个区块链网络的所有节点都运行本地版本代码，区块链中每秒交易量就会趋于稳定。

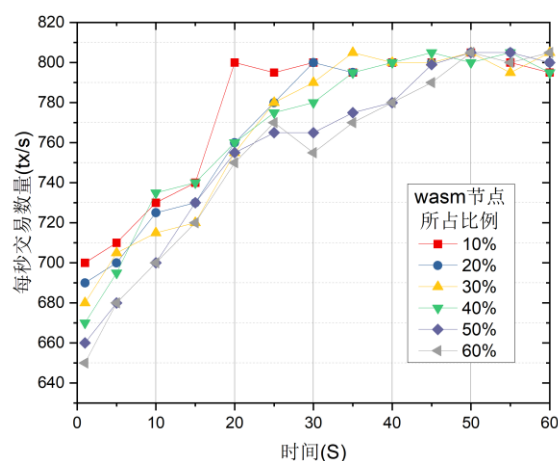


图 5-15 不同 wasm 节点所占比例对每秒交易量的影响

出块速度是评价区块链性能的重要指标，出块速度可以通过两个区块链上链的时间差来判断，这个时间差成为出块时间。出块时间和许多因素有关：区块大小、共识算法、网络环境、节点数量等。本系统中区块大小和共识算法都已经确定不易更改，网络环境为不可控因素，因此只考虑区块链网络中的节点数量对出块时间的影响。在表 5-6 中展示了区块链中包含 1 至 5 个节点数是，统计了一共构造 300 个区块的总时间，并计算了平均时间。

表 5-6 医疗区块链出块时间

全节点数	统计区块数	总时间(单位：秒)	平均出块时间(单位：秒)
1	300	1785	5.95
2	300	1803	6.01
3	300	1836	6.12
4	300	1842	6.14
5	300	1860	6.2

以太坊的平均出块时间是 15 秒，比特币的平均出块时间是 10 分钟，本文中提出的医疗区块链的平均出块时间是 6 秒，速度比常见的比特币网络快，基本能满足医疗场景下的数据保存响应时间的要求。

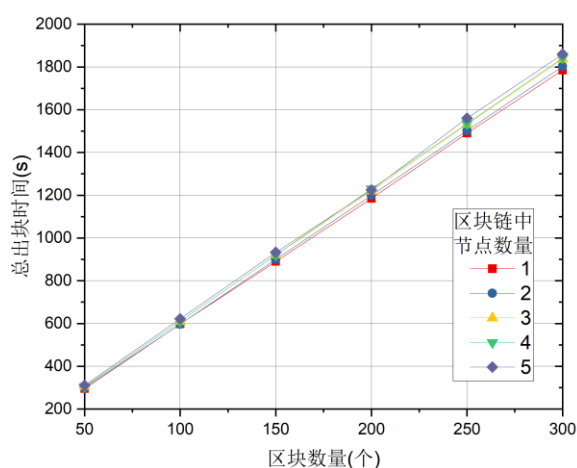


图 5-16 医疗区块链出块时间测试

图 5-16 中展示了当区块个数增加时，区块链网络中有不同的节点数时的总出块时间的变化图，当区块链网络中节点数固定时，总出块时间和区块个数基本成正比；当总出块个数固定时，总出块时间随着区块链网络中节点数的增加而增加，但是增加幅度不大。由此可以得出以下结论：

（1）出块速度不随区块链网络运行时间的增加而变化，相对固定，说明本文的区块链网络具有较强的稳定性。

（2）出块速度不随区块链网络中的节点数增加而大幅增加，增加幅度很小，符合医疗场景下的时间要求，说明本文的区块链网络具有较强的可扩展性。

5.3.3.2 计算开销

区块链中所有涉及密码学的函数都需要一定的计算开销，例如哈希函数，指数运算，对数运算，加解密运算等，在 Polkadot 中又涉及在 Polkadot 中发起一次交易和在 Polkadot 中进行一次跨链操作。这些计算都需要消耗一定的计算机算力，这是影响系统性能的主要原因。根据这些基本操作估算计算成本，图 5-17 展示了医院、患者、单条区块链下的医生、多条区块链下的医生和云服务提供商的在随着医院数量增加时计算开销的变化。

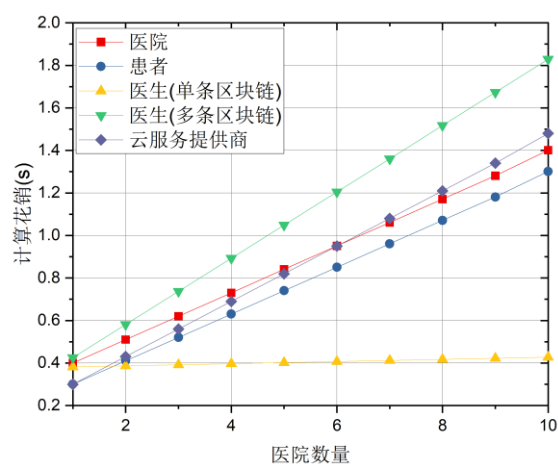


图 5-17 计算花销

5.4 本章小结

本章对第四章提出的基于区块链的电子健康记录隐私保护方案进行了实现和测试，实现了方案的关键功能，并对功能进行了测试，同时还对区块链的性能和跨链的性能进行了测试。

第六章 总结与展望

6.1 总结

自 2008 年中本聪提出比特币以来，其底层技术区块链得到了不断地发展，应用领域也从金融领域不断扩大，包括教育、交通、医疗等跟生活息息相关的领域。然而要在这些领域实施落地应用还需要很长一段时间的发展，目前我国无论是政府还是企业都在大力支持区块链的发展。在医疗领域中区块链就有很大的应用价值。电子健康记录是反应患者在一段时间内医疗信息的电子文档，以电子化方式管理个人的健康状态和医疗信息，涉及患者信息的采集、存储、传输、处理和利用，电子健康记录包含了患者大量的隐私信息。当前大多数医疗机构的电子健康记录的管理系统都是由第三方承包并由第三方存储电子健康记录信息，由于电子健康记录的所有者是患者，而管理者是医生（或医院），持有者是第三方，因此这种三权分离的场景将导致电子健康记录的存在隐私泄露的风险，患者在不知情的情况自身的隐私数据可能遭到修改或窃取。区块链本质是一个分布式的、去中心化的数据库，具有开放性、防篡改性、匿名性和可追溯性，这些性质十分契合电子健康记录隐私保护的场景。

本文研究了基于区块链的电子健康记录隐私保护的方法，具体内容包括：

（1）分析了电子健康记录的特点，以及区块链如何用于保护电子健康记录隐私，得出了可以用跨链技术来保护电子健康记录隐私的结论；并对电子健康记录隐私保护的需求以及设计的角色和功能进行了分析。

（2）提出了一种基于区块链跨链技术的电子健康记录隐私保护方案（CEPS），采用 Substrate 区块链开发框架构建了医疗场景下的区块链，并将区块链接入 Polkadot 中继链实现了基于跨链的电子健康记录隐私保护方案及系统。

（3）模拟实际医疗场景，进行了测试，构建了两条区块链并接入 Polkadot 中继链，验证了本文提出的方案对电子健康记录隐私保护的有效性。

6.2 展望

对于区块链的发展，从技术角度看，目前最为火热的发展方向是区块链跨链技术；从应用角度看，区块链正在进入各行各业。由于区块链技术还在发展，落地应用也都还在起步阶段，所以在医疗领域中基于区块链的研究也还在发展中。本文提出的基于区块链的电子健康记录隐私保护方案在一定程度上有效地解决了电子健康记录的隐私泄露问题，但是仍在一些方面存在不足需要进一步研究：

（1）功能完善。在实现方面为了开发方便，对电子健康记录的数据结构和医疗系统的功能进行了简化，为了实现完整的功能需要在后期的开发中继续完善和拓展。

（2）界面优化。在用户界面方面，区块链信息展示界面和用户操作界面没有分离，后期工作中应当分离成一个区块链浏览器和一个基于 Web 的客户端。

（3）性能提升。由于区块链的技术特点限制了区块链的性能，目前区块链能做到平均出块时间 6 秒左右，基本满足了大量的医疗场景，但是还应缩短出块时间以及提升高并发的能力。

致 谢

晃眼间今年已经是在电子科技大学的第七个年头，攻读硕士学位的三年时间就要过去了，论文工作也进入了尾声，回顾这三年学习和科研生活，不禁十分感慨。不论是在生活中，在学业上还是在科研中，学院老师、导师、家人、师兄们、同学、室友都给予了帮助、关心和指导。我要向这些给予过我帮助的人们表达我的感谢。

首先我要感谢我的导师曹晟老师，在科研工作中，曹老师总是秉持着一种严谨的态度，同时也用这种态度要求着我，使我在科研工作中稳步前行，在本论文的写作过程中，曹老师经常认真仔细的指导，每次都不厌其烦的指出论文中存在的不足，并给出意见，如果没有老师的指导，论文写作势必会遇到阻碍。再次感谢曹老师的辛勤付出。

同时我要感谢我的学长、同门、学弟、学妹和实验室的同窗们，在三年的研究生生活里，他们不论是在生活中还是在科研上，都给了我很大的帮助。

然后我要感谢我的父母，虽然他们在科研中能给予我的帮助很少，但他们在生活中给了我很大的支持，为我创造了良好的学习环境和氛围。感谢父母不求回报的付出，感谢父母默默无闻的支持。

最后，感谢各位参与论文评审的老师，感谢你们为论文提出的宝贵意见，谢谢！

参考文献

- [1] D. Blumenthal, M. Tavenner. The “meaningful use” regulation for electronic health records[J]. New England Journal of Medicine, 2010, 363(6): 501-504
- [2] 工信部. 2018 年中国区块链产业白皮书[EB/OL]. <http://www.miit.gov.cn/n1146290/n1146402/n1146445/c6180238/content.html>
- [3] S. Chenthar, K. Ahmed, H. Wang, et al. Security and privacy-preserving challenges of e-Health solutions in cloud computing[J]. IEEE access, 2019, 30(7): 74361-82
- [4] Q. Wang, D. Zhou, S. Yang, P. Li, et al. Privacy Preserving Computations over Healthcare Data[C]. In 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2019, 635-640
- [5] A. Sahi, D. Lai, Y. Li. Security and privacy preserving approaches in the eHealth clouds with disaster recovery plan. Computers in biology and medicine, 2016, 1(78): 1-8
- [6] 陈虹云, 王杰华, 胡兆鹏, 等. 面向医疗数据发布的动态更新隐私保护算法[J]. 计算机科学, 2019, 46(1): 206-11.
- [7] S Nakamoto. Bitcoin: A peer-to-peer electronic cash system[EB/OL]. <https://bitcoin.org/bitcoin.pdf>
- [8] Z. Zheng, S. Xie, H. Dai, et al. Blockchain challenges and opportunities: A survey[J]. International Journal of Web and Grid Services, 2018, 14(4): 352-375
- [9] M. Swan. Blockchain: Blueprint for a new economy[M]. Sebastopol: O'Reilly Media, Inc, 2015, 1-128
- [10] 袁勇, 王飞跃. 区块链技术发展现状与展望[J]. 数字通信世界, 2016, 042(004): 481-494
- [11] 韩璇, 袁勇, 王飞跃. 区块链安全问题: 研究现状与展望[J]. 自动化学报, 2019, 045(001): 206-225
- [12] 路爱同, 赵阔, 杨晶莹, 等. 区块链跨链技术研究[J]. 信息网络安全, 2019, 19(8): 83-90
- [13] 李芳, 李卓然, 赵赫. 区块链跨链技术进展研究[J]. 软件学报, 2019, 2019(06): 1649-1660
- [14] 魏昂. 一种改进的区块链跨链技术[J]. 网络空间安全, 2020, 10(6): 1-6
- [15] M. Borkowski, M. Sigwart, P. Frauenthaler, et al. DeXTT: Deterministic Cross-Blockchain Token Transfers[J]. IEEE Access, 2019, 7: 111030-111042
- [16] V. Buterin. Chain interoperability[J]. R3 Research Paper, 2016, 1(1): 1-25
- [17] J. Chow. BTC Relay[EB/OL]. <http://btcrelay.org>
- [18] J. Kwon. Cosmos Whitepaper[EB/OL]. <https://cosmos.network/resources/whitepaper>

- [19] T. Koens, E. Poll. Assessing interoperability solutions for distributed ledgers[J]. *Pervasive and Mobile Computing*, 2019, 1(59):1-29
- [20] G. Drosatos, E. Kaldoudi. Blockchain applications in the biomedical domain: a scoping review[J]. *Computational Structural Biotechnology Journal*, 2019, 17(1): 229-240
- [21] 薛腾飞, 傅群超, 王枫, 等. 基于区块链的医疗数据共享模型研究[J]. *自动化学报*, 2017, 43(9): 1555-1562
- [22] 徐文玉, 吴磊, 阎允雪. 基于区块链和同态加密的电子健康记录隐私保护方案[J]. *计算机研究与发展*, 2018, 55(10): 141-151
- [23] J. Liu, X. Li, L. Ye, et al. BPDS: A blockchain based privacy-preserving data sharing for electronic medical records[C]. *2018 IEEE Global Communications Conference (GLOBECOM)*, Dhabi, 2018, 1-6
- [24] H. Wang, Y. Song. Secure cloud-based EHR system using attribute-based cryptosystem and blockchain[J]. *Journal of Medical Systems*, 2018, 42(8): 152
- [25] S. Cao, G. Zhang, P. Liu, et al. Cloud-assisted secure eHealth systems for tamper-proofing EHR via blockchain[J]. *Information Sciences*, 2019, 485 (2019): 427-440
- [26] D. Ivan. Moving toward a blockchain-based method for the secure storage of patient records[C]. *ONC/NIST Use of Blockchain for Healthcare and Research Workshop*, Gaithersburg, 2016, 1-11
- [27] A. Azaria, A. Ekblaw, T. Vieira, et al. Medrec: Using blockchain for medical data access and permission management[C]. *2016 2nd International Conference on Open and Big Data (OBD)*, Vienna, 2016, 25-30
- [28] V. Ramani, T. Kumar, A. Bracken, et al. Secure and efficient data accessibility in blockchain based healthcare systems[C]. *2018 IEEE Global Communications Conference (GLOBECOM)*. 2018, 206-212
- [29] T. Oliveira, L. Reis, R. Carrano, et al. Towards a blockchain-based secure electronic medical record for healthcare applications[C]. *2019 IEEE International Conference on Communications (ICC)*. 2019, 1-6
- [30] C. Esposito, A. De Santis, G. Tortora, et al. Blockchain: A panacea for healthcare cloud-based data security and privacy?[J]. *IEEE Cloud Computing*, 2018, 5(1): 31-37
- [31] A. Act. Health insurance portability and accountability act of 1996[J]. *Public law*, 1996, 104(191): 1-14
- [32] M. Crosby, P. Pattanayak, S. Verma, et al. Blockchain technology: Beyond bitcoin[J]. *Applied Innovation*, 2016, 2(6-10): 71-85

- [33] X. Xu, I. Weber, M. Staples, et al. A taxonomy of blockchain-based systems for architecture design[C]. 2017 IEEE International Conference on Software Architecture (ICSA), Gothenburg, 2017, 243-252
- [34] P. Zimmerman. Blockchain structure and cryptocurrency prices[J]. Bank of England Working Paper, 2020, 855(1): 1-71
- [35] P. Szalachowski. Towards More Reliable Bitcoin Timestamps[C]. 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), Zug, 2018, 101-104
- [36] J. Coron, Y. Dodis, C. Malinaud, et al. Merkle-Damgård revisited: How to construct a hash function[C]. Annual International Cryptology Conference, California, 2005, 430-448
- [37] R. C. Merkle. A digital signature based on a conventional encryption function[C]. Conference on the Theory and Application of Cryptographic Techniques, Amsterdam, 1987, 369-378
- [38] S. Hejazi-Sepehr, R. Kitsis, A. Sharif. Transwarp Conduit: Interoperable Blockchain Application Framework[J]. arXiv Preprint arXiv:.03256, 2019, 1(1): 1-19
- [39] L. Deng, H. Chen, J. Zeng, et al. Research on cross-chain technology based on sidechain and hash-locking[C]. International Conference on Edge Computing, San Francisco, 2018, 144-151
- [40] A. Back, M. Corallo, L. Dashjr, et al. Enabling blockchain innovations with pegged sidechains[EB/OL]. <http://www.opensciencereview.com/papers>
- [41] S. Schulte, M. Sigwart, P. Frauenthaler, et al. Towards Blockchain Interoperability[C]. International Conference on Business Process Management, Vienna, 2019, 3-10
- [42] G. Wood. Polkadot: Vision for a heterogeneous multi-chain framework[J]. White Paper, 2016, 1(1): 1-21
- [43] M. Mettler. Blockchain technology in healthcare: The revolution starts here[C]. 2016 IEEE 18th International Conference on E-health Networking, Applications and Services (Healthcom), Munich, 2016, 1-3
- [44] 何波, 王桂胜. 基于区块链技术的医疗管理信息化应用分析[J]. 四川大学学报, 2018, 55(06): 93-98
- [45] D. E. Bakken, R. Rameswaran, D. M. Blough, et al. Data obfuscation: Anonymity and desensitization of usable data sets[J]. IEEE Security, 2004, 2(6): 34-41
- [46] D. Rachmawati, J. Tarigan, A. Ginting. A comparative study of Message Digest 5 (MD5) and SHA256 algorithm[C]. Journal of Physics: Conference Series, Indonesia, 2018, 978-984
- [47] X. Wang, H. Yu. How to break MD5 and other hash functions[C]. Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, 2005, 19-35

攻读硕士学位期间取得的成果

一、科研工作情况

- [1] 四川省人工智能重大专项, 人工智能应用区块链基础平台关键技术研究, 2019/6-2021/6, 在研、主研

二、论文

- [1] S. Cao, J. Wang, X. Du, et al. CEPS: a Cross-blockchain based electronic health records privacy-preserving scheme[C] IEEE International Conference on Communications (ICC2020), Dublin, 2020 (电子科技大学 A 类会议, 中国计算机学会 C 类会议)

三、申请专利情况

- [1] 曹晟, 蒋长红, 王靖, 等. 一种适用于分布式数据库的负载均衡方法、装置及服务器[P]. 中国发明专利, CN201810530688.2, 2018 年 11 月 13 日
- [2] 曹晟, 邹杰成, 王靖, 等. 一种数据库的安全审计系统、方法及服务器[P]. 中国发明专利, CN201810529452.7, 2018 年 11 月 6 日

四、获奖情况

- [1] 2017-2018 年, 获得研究生二等奖学金
- [2] 2018-2019 年, 获得研究生二等奖学金