计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2018.20180438
55(10):2233-2243,2018

基于区块链和同态加密的电子健康记录隐私保护方案

徐文玉[1,2] 吴 磊[1,2,3] 阎允雪[1,2]
[1](山东师范大学信息科学与工程学院 济南 250358)
[2](山东省分布式计算机软件新技术重点实验室 济南 250014)
[3](山东省软件工程重点实验室 济南 250101)
(xu__wy@163.com)

Privacy-Preserving Scheme of Electronic Health Records Based on Blockchain and Homomorphic Encryption

Xu Wenyu[1,2], Wu Lei[1,2,3], and Yan Yunxue[1,2]
[1](*School of Information Science and Engineering, Shandong Normal University, Jinan* 250358)
[2](*Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology, Jinan* 250014)
[3](*Shandong Provincial Key Laboratory for Software Engineering, Jinan* 250101)

Abstract The privacy protection of electronic health records (EHR) has become an issue which attracts more and more attention in public. Blockchain is a technology that has emerged with the spread of digital cryptocurrency such as Bitcoin and has features of "decentralization" and "unmodifiable". Existing electronic health record management systems ignore the security problems of patients' interaction with other roles while focusing on protecting the user's privacy data, especially there is no such an appropriate solution to problem nowadays that insurance can view patients' sensitive data and invade privacy. This paper proposes a scheme based on blockchain for solving the above three problems. In combination with homomorphic encryption and smart contract technology based on Ethereum, we implement the feature that the insurance company can judge whether to handle the claim requests, although it has no way to obtain the plaintext of EHR and the ID. So there is no sensitive data of the patient which will be leaked to unauthorized users during interaction, thus the privacy protection of users' data is strengthened. This thesis focuses on analyzing the interaction process of different roles under different application requirements based on the premise of patients' privacy and carries out security analysis and performance evaluation.

Key words blockchain; electronic health records (EHR); privacy protection; smart contract; automatic claim; homomorphic encryption

摘 要 电子健康记录(electronic health records, EHR)的隐私保护成为现代人越来越关注的问题.区块链是随着比特币等数字加密货币普及而兴起的技术,具有"去中心化"和"不可篡改"等特点.现有的电子

收稿日期:2018-06-12;**修回日期**:2018-08-11
基金项目:国家自然科学基金项目(61602287,61672330);山东省重点研发计划项目(2018GGX101037);山东省科技重大创新工程项目(2018CXGC0702)
This work was supported by the National Natural Science Foundation of China (61602287, 61672330), the Key Research & Development Program of Shandong Province (2018GGX101037), and the Major Scientific and Technological Innovation Project of Shandong Province (2018CXGC0702).
通信作者:吴磊(wulei@sdnu.edu.cn)

健康记录管理系统往往注重保护用户隐私数据而忽略了患者与其他角色交互时存在的安全问题,尤其是并未针对理赔过程中保险公司可查看患者敏感数据侵犯患者隐私这一问题提出明确的解决方案.故提出了一个基于区块链的可同时解决以上3个问题的方案,并将同态加密和以太坊的智能合约技术相结合,实现了保险公司在无法获取用户EHR明文和理赔对象ID的情况下仍能判断是否理赔的功能,交互过程中不向非授权用户泄露患者的任何敏感数据,加强了对用户数据的隐私保护.分析了在保护患者隐私的前提下不同角色在不同应用需求下的交互过程,并对该方案进行安全性分析和性能评估.

关键词 区块链;电子健康记录;隐私保护;智能合约;自动理赔;同态加密

中图法分类号 TP393

电子健康记录(electronic health records, EHR)早在20世纪60年代末就开始投入使用,它记录了个人疾病发生、发展、治疗转归的过程,具有很高的医疗价值.正因如此,医疗行业一直是信息窃取的主要目标.医疗数据的爆炸式增长及互联网的迅速发展,致使医院内网的数据逐渐走向公网,然而医疗行业的信息安全保护起步较晚,医疗数据泄露的事故层出不穷.2017年我国发生重大医疗信息泄露事件,导致7亿条个人信息被倒卖.除此之外,医院对EHR具有绝对的控制权,可能会存在恶意删除、修改或内部人员泄露数据的情况.

传统理赔中,保险公司有权查看患者的明文电子健康记录,并核对数据库患者的投保信息判断是否应该理赔,这实际上侵犯了患者的隐私.区块链是一种新兴的技术,本质上是一个授权的账本,具有"去中心化"和"不可篡改"等特点,能够使用户核实和记录所有交易,且存储在链上的信息不可更改.以太坊被称作区块链2.0的技术,其上拥有智能合约,供区块链上所有节点执行.如何保护患者数据隐私,加强患者对数据的控制权及解决保险公司安全理赔对EHR建设具有重要的意义.

1 相关工作

文献[1]阐述了基于云计算的EHR管理系统要求达到的安全性要求,并提出了一些建议.文中提到提供者必须保护电子健康记录的安全性和隐私性,实施必要的安全机制以保证患者信息在云中的安全.文献[2]将属性基加密与可搜索加密查询技术相结合构建了安全的、隐私保护的EHR系统,使患者和提供商之间以灵活、动态、可扩展的方式共享数据.Xhafa等人提出云计算中具有隐私意识的基于属性的个人健康记录共享系统[3],该方案是一种秘钥策略的多机构的属性基加密,通过隐藏访问策略保护用户的隐私.文献[4]中,电子健康记录由患者拥有数据所有权并将其存储至云中,保证了患者拥有数据的所有权,但忽略了其他角色和患者共享医疗方面存在的问题.此外,此方案无法解决患者不拥有数据所有权(如心理治疗记录)的特殊情况.随着云计算中保护患者EHR隐私的技术逐渐成熟和区块链技术的不断发展,许多研究学者开始把研究方向转向基于区块链的EHR隐私保护,大量分析区块链可用于医疗行业的文章出现.文献[5-8]分析了区块链应用于医疗行业的可行性,并分别提出了医疗共享模型.其中文献[5]提出的基于区块链的医疗数据共享模型解决了各医疗机构间校验、保存和同步医疗数据的难题.该模型使用的是股份授权证明机制(delegated proof of stake, DPoS),较于现有解决方案具有一定优势,但存在不具备预测模块、机器学习算法的能力,在数据存储能力等方面具有一定的缺点.Brodersen等人指出区块链可以促进医疗信息共享,并且探讨了如何将当前的医疗IT投资与分布式账本技术(distributed ledger technology, DLT)相结合[6],此外文中提出了3种可以解决长期困扰医疗行业难题的区块链结构,区块链技术与医疗行业的结合有利于患者、医疗机构和医疗保健参与者等各个参与方.文献[7]介绍了当前提出的基于区块链的EHR系统,指出区块链具有应用于医疗保健行业的潜力,但在实际应用中仍然存在实施障碍还需进一步研究.文献[8]提到区块链用于医疗数据系统是可行的,但存在重大障碍和挑战,如不存在适用于医疗保健的区块链设施和节点用户不愿意用自己的算力维护医疗区块链.文章就抵抗中间人攻击的问题提出了3个不同的转移方案.文献[9-11]对在区块链上以何种方式存储电子健康记录和如何抵抗区块链上的中间人攻击提出了一些解决方案.其中文献[9]利用区块链技术提出了一个去中心化的系统MedRec,该系统在区块链上存储电子健康

记录索引的 Hash 值,以此来为患者提供完整的日志,该系统还使用了一系列智能合约跟踪某些状态的转换,提高系统的可操作性和对患者的数据隐私保护. Linn 和 Koo 等人将医疗记录存储医院的数据库中,把区块链集成为一个访问控制层,存储了指向电子健康记录 Hash 值的索引,提高了系统的可拓展性[10]. 文献[10]中还提出了一个用于身份认证的生物认证系统,用于保护访问控制,抵抗中间人攻击. Halamka 等人[11]提出将数据存储在权限区块链中,对存储在区块链上的 EHR 和权限 Hash 值进行加密,以此来提高数据的安全性. 文献[12]实施了一个只有提供者、政府和患者参与的权限区块链,利用区块链及相关技术将提供者、患者和政府监管机构之间复杂的协议转化成计算机代码(智能合约),保证了互操作性和透明性. 为抵抗中间人攻击,文中提到可以将虚假记录和区块链上的有效记录混合. 为了达到患者拥有数据所有权和 EHR 控制权的目的,Dagher 等人提出了一个基于区块链的隐私保护框架 Ancile[13],在节点注册阶段就为其提供明确的权限级别,文献[13]中设计了多个智能合约对区块链上的节点进行权限控制,扩大了患者对电子健康记录的控制权.

从上述分析来看,基于区块链的 EHR 隐私保护是重点研究课题,其中涉及到数据如何在区块链上存储、如何保护区块链上的数据隐私、如何保护访问控制、如何扩大患者的控制权以及如何抵抗区块链上的攻击,如中间人攻击. 国内学者大都在研究基于云存储的电子健康记录隐私保护,对基于区块链的患者敏感数据保护领域的研究处于刚刚起步状态. 相对于国外来说,有较多学者在研究此课题,设计出了一些基于区块链的 EHR 系统,这些系统不仅具有理论指导意义,更具有一定的应用可能性,更有部分系统已经应用于现实生活中,但仍存在一些问题处于不断改进的状态. 现如今国内外还未有文章涉及到基于区块链的保险公司理赔的具体应用场景,本文就改进传统理赔方式提出了解决方案. 传统理赔时,保险公司具有查看患者明文病历的权限,这实际上侵犯了患者的隐私. 本文基于区块链,对 Ancile[12] 进行了扩展,以 Ancile 为基础添加了保险公司的角色,将同态加密、智能合约和两方安全计算技术相结合,实现了保险公司在不能获取患者明文 EHR 和用户 ID 的前提下仍能正确理赔的功能,对改进传统理赔方式具有重要理论意义.

2 区块链技术背景知识

2.1 区块链概述及特征

区块链是从比特币底层技术衍生出来的新技术体系,最早的定义来自于中本聪在 2009 年发表的论文[13]. 区块链是一种按照时间顺序将数据区块以链条的方式组合而成的特定数据结构,并以密码学方式保证的不可篡改、不可伪造的去中心化公共总账[14]. 账本是由不同区块构成的,一个区块包含区块头和区块体 2 部分,前一区块的 Hash 值、挖矿的难度及本区块中一定数量的交易数据的默克尔树、时间戳等参数构成了区块头,区块头的结构如图 1 所示,区块和区块之所以形成链是因为区块中存储了前一个区块的 Hash 值.

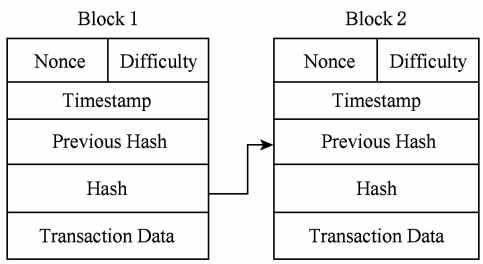

Fig. 1 The structure of the block in block chain

图 1 区块链中区块的结构

区块链由所有参与节点共同维护,每一个节点可以通过定期与邻居节点交换信息使全局账本保持同步. 区块链的特点:

1) 开放共识. 任何用户都可以参与到区块链中,参与其中的节点都可以获得一份完整的账本;任何用户都可以通过公开的接口查询区块链上的数据.

2) 去中心化. 区块链是由参与节点组成的 P2P 网络,不存在中心化的硬件或管理机构,任何节点在网络中都是平等的.

3) 不可篡改性. 区块链是一个公开的账本,区块通过节点的验证后会被永久存储. 区块中含有上一个区块的 Hash 值,如果修改其中某一个区块,在这之后的所有区块都要重新计算,除非能控制系统中超过 51%的节点,否则单个节点对数据库的修改是无效的.

2.2 共识机制

区块链技术采用共识机制保证所有合法节点维持的全局账本是一样的. 常见的共识机制有工作量

证明(proof of work, PoW)、权益证明机制(proof of stake, PoS)、股份授权证明机制(DPoS)等. 比特币中使用的共识机制是工作量证明,若节点可以通过计算获得一个满足规则的随机数,就可以拥有此区块的记账权,其他节点验证通过后存储此区块. 这种方法完全去中心化,但是挖矿会造成大量的资源浪费,且达成共识的周期较长. PoS 权益证明根据每个节点所占代币的比例与时间等比例的降低挖矿难度,由此可以减少达成共识的时间,可以代替 PoW 机制. DPoS 股份授权机制类似于董事会投票,全部的节点投票选举出一定数量的节点代表,由他们来确认区块.

本文中使用的共识机制是"QuorumChain"[15]共识,是一种基于投票的共识机制. QuorumChain 规定一定数量的节点具有投票的能力,获得大部分选票(高于阈值数量)的块将会被添加到区块链. 该算法在创建块之前设置了一个随机时间量,可以减少多个矿工节点同时创建相同区块的情况发生.

2.2 以太坊及 gas[16]

以太坊(Ethereum)是一个基于区块链的去中化应用平台,被称作区块链 2.0 的技术,本质就是保存数字交易永久记录的公共数据库. 以太坊上的程序称为智能合约,是代码和数据的集合,通过以太坊虚拟机来进行处理.

比特币是区块链 1.0 的技术,开创了数字货币的先河,但比特币协议的扩展性不足,协议中使用的是基于堆栈的脚本语言,不能构建更高级的应用. 由此出现了图灵完备的以太坊,解决了比特币扩展性不足的问题.

以太坊在区块链的运行环境被称作以太坊虚拟机(ethereum virtual machine, EVM),每个区块链上的节点都会验证区块中包含的每个交易并在 EVM 中运行交易所触发的代码. 网络中的全节点(full node)会进行相同的计算,存储相同的值,这使得合约的执行都会需要大量的消耗,以太坊中用 gas 为单位来表示消耗.

在以太坊中,每个调用智能合约的交易都有一个 gas 消耗的上限,称之为 gas limit. 如果一个交易不仅需要原始计算而且还需要触发别的合约,当该交易需要使用的总 gas 数量小于 gas limit 时,这个交易才会被处理. 每笔交易的交易费与 gas used 和 gas price 相关,发送、存储和执行计算等每项操作在 EVM 中都被设置了固定的 gas 消耗值,称之为 gas used,单位 gas 的价格(用以太币计算)称之为 gas price,二者的乘积是每笔交易的交易费. 发起交易的用户余额必须大于执行交易的 gas 成本,否则交易终止. 若在交易过程中发现用户余额不足,消耗的 gas 不会退还.

2.3 智能合约

"智能合约"(smart contract)这个术语至少可以追溯到 1995 年,文献[17]将智能合约定义为一套以数字形式定义的承诺(promises),包括合约参与方可以在上面执行这些承诺的协议". 智能合约是一段可以写入区块链的代码,可以被所有参与节点运行. 智能合约不仅仅是一个可以自动执行的计算机程序更像是一个可信的第三方,可以临时保管资产并且严格按照事先约定好的规则执行操作. 智能合约的优点是可以多方合作,消除潜在的人为错误和腐败风险,对于每个参与用户来说都是公开透明的,目前主流的开发平台是以太坊.

2.4 区块链的隐私保护

隐私即为数据所有者不愿意被披露的敏感信息,包括敏感数据以及数据所表征的特性. 个人隐私指的是任何可以确认特定个人或与可确认的个人相关,但个人不愿暴露的信息. 隐私保护技术主要是考虑 2 个方面:1)如何保证数据应用过程中不泄露隐私;2)如何更有利于数据的应用[18],一般而言隐私分为个人隐私和共同隐私 2 类[19].

区块链为了维护各节点的数据同步必须要公开部分信息,如交易内容,但为了保护用户隐私,要求不能泄露用户的敏感数据. 通过分析区块链技术的特点,区块链中的隐私分为 2 类[20]:身份隐私和交易隐私. 身份隐私是指用户身份信息和区块链地址之间的关联关系,区块链的地址是由公钥加密算法生成,与个人信息无关,一个用户可以生成多个地址. 因此通过地址很难推断出用户信息,相比较于传统的账号来说具有较好的匿名性. 交易隐私指的是区块链中的交易记录和交易记录背后的知识,由于在区块链中交易是公开的,而交易中往往存在一些私人交易,比如金钱交易等,所以应该对交易记录进行一定的保护.

区块链的特性决定了区块链对保护用户隐私具有天然优势,首先区块链网络是一种 P2P 网络,很难实现网络窃听;其次区块链支持匿名交易,即以太坊的地址是由用户公钥经过一定算法生成,与个人信息无关,具有一定的匿名性. 但也存在不足之处,如区块链中任何节点都可查看交易,攻击者可以通过查看交易逐渐分析出地址所对应的用户信息.

3 预备知识

3.1 相关定义

定义 1. 不基于第三方的公平的安全两方数据匹配. Alice 拥有数据块 M,Bob 拥有数据 M_1,M_2,…,M_n,在不向对方泄露 M 和 M_1,M_2,…,M_n 也不借助第三方的情况下,查看 M 是否与 Bob 的某一块数据匹配.

定义 2. 半诚实参与模型[2]. 就是各参与者都完全按照协议的规定执行协议的每一步,并且在协议执行过程中不会中途退出协议,也不会恶意输入或提供虚假数据,但它们可能会通过试图分析和利用协议执行过程中自己所得到的信息来推断协议之外的其他信息.

3.2 同态加密

同态加密最初是由 Rivest 等人于 1978 年提出的,是一种允许直接对密文进行操作的加密变换技术[21].

如果 2 个密文 $E(x)$和 $E(y)$通过运算计算出 $E(x+y)$,即 $F(E(x),E(y))=E(x+y)$,F 代表任意运算,则 E 是加法同态算法.

如果 $E(x)$和 $E(y)$可以通过运算计算出 $E(x\times y)$,即 $F(E(x),E(y))=E(x\times y)$,则 E 是乘法同态.

如果 $E(x)$和 y 通过运算计算出 $E(x\times y)$,即 $F(E(x),E(y))=E(x\times y)$,则 E 是混合乘法同态.

如果一个加密方案同时满足加法同态和乘法同态称之为全同态加密.

3.3 Paillier 加密体制[22]

3.3.1 Paillier 基于的困难问题

Paillier 加密体制是基于判定合数剩余类问题的加法同态加密密码体制,于 1999 年由学者 Paillier 提出. 判定合数剩余类问题是指使 $N=p\times q$,其中 p 和 q 都是大素数,任意给定 $y\in Z_{N^2}^*$,使 $z=y^N \bmod N^2$,判定 z 是模 N^2 的 N 次剩余还是非剩余是困难的,基于合数剩余类问题是公认的数学难题[23]. 依据

$$\lambda=\lambda(n)=lcm(p-1,q-1),$$

设:

$$S_n=\{u<n^2 \mid u=1 \bmod n\},\forall u\in S_n,$$

$$L(u)=\frac{u-1}{n},$$

则困难问题的解为

$$(L(c^\lambda \bmod n^2)/L(g^\lambda \bmod n^2)) \bmod n.$$

3.3.2 Paillier 公钥密码体制

密钥生成阶段,$n=pq$,其中 p,q 为 2 个大素数,选择 $g\in Z_{n^2}^*$,使得 $\gcd(L(g^\lambda \bmod n^2),n)=1$,公钥为$(n,g)$,私钥为$(p,q)$.

加密阶段,存在明文 $m\in Z_n$,且 $m<n$. 选择一个随机数 $r<n$,则密文为

$$c=g^m\times r^n \bmod n^2.$$

解密阶段,密文 $c<n^2$,明文为

$$m=\frac{L(c^\lambda \bmod n^2)}{L(g^\lambda \bmod n^2)} \bmod n.$$

3.3.3 Paillier 加密的同态性质

根据 3.2 节介绍,$F(E(x),E(y))=E(x+y)$(F 代表任意运算),则 E 是加法同态算法. 假设有明文 m_1,m_2,使用 Paillier 加密方案对 2 个明文进行加密,得到:

$$E(m_1)=g^{m_1}\times r_1^n \bmod n^2,$$

$$E(m_2)=g^{m_2}\times r_2^n \bmod n^2,$$

则:

$$E(m_1)\times E(m_2)=g^{m_1+m_2}(r_1r_2)^n \bmod n^2=E(m_1+m_2),$$

此表达式满足加法同态的性质,因此 Paillier 加密体制是加法同态.

3.4 代理重加密

代理重加密解决了用户在共享数据时的不便,在减轻用户端负担的同时强化了数据的可靠性和安全性. 代理重加密实质上是一种用于密文之间的密钥转换机制. 在代理重加密过程中,各参与方均得不到任何明文信息. 简单来说,代理重加密就是授权者通过代理产生针对被授权者的转换密钥,代理者利用该密钥就能将由授权者公钥加密的密文转化为用被授权者的公钥所加密的密文.

3.5 分布式 ElGamal 代理重加密[24]

ElGamal 代理重加密方案基于 ElGamal 加密,取 2 个大素数 p 和 q,且 $p=2q+1$. G_p(q 阶)是 $Z_p^*=\{i \mid 1\leqslant i\leqslant p-1\}$的循环子群,$q$ 是 G_p 的生成元. 取 $k\in Z_q^*$ 为 ElGamal 加密方案的私钥,则 $K=(p,q,g,y)$为公钥,其中 $y=g^k \bmod p$. 存在明文 $m\in G_p$,密文则为(g^r,my^r),r 是随机数($r\in Z_q^*$),记为 $E(m)=(a,b)$. 解密时计算 $\frac{b}{a^k}=\frac{m(g^k)^r}{(g^r)^k}=m$. 假设授权者的公钥为 A,被授权者的公钥为 B,基于 ElGamal 的代理重加密方案如下:

1) C_j 向各服务器 i 发送初始化消息启动协议;

2) 从 C_j 收到消息的服务器 i 生成随机数 p,

计算($E_A(p_i)$,$E_B(p_i)$),并将其发送给 C_j;

3) C_j 计算 $E_A(p)=\times_{i\in I}E_A(p_i)$ 和 $E_A(p)=\times_{i\in I}E_B(p_i)$,将其发送给授权者;

4) 授权者计算 $E_A(mp)=E_A(m)\times E_A(p)$,解密 $E_A(mp)$,得到 mp,并将其发送给 C_j;

5) C_j 计算 $E_B(m)=(mp)\times E_B(p)^{-1}$.

4 本文方案

4.1 概 述

本文方案涉及到 7 个智能合约,患者可以通过智能合约查看和更改其他用户对该记录的权限,这实现了患者对隐私数据的访问控制;电子健康记录可以在不同节点之间进行安全转移,实现了对隐私数据的保护;除此之外,本文方案改变了传统理赔方式,将智能合约和同态加密技术相结合,在隐藏理赔对象 ID 和该理赔对象电子健康记录的情况下,实现保险公司自动理赔的功能. 方案中各参与方与区块链的交互如图 2 所示,患者的 EHR 存储在医院的数据库中,医院将对称加密密钥和患者部分信息存储至区块链中,患者可通过与区块链上的智能合约交互更改 EHR 的访问权限和获取加密密钥以解密 EHR. 保险公司可通过区块链获取患者的信息,投票节点使用上文提到的 QuorumChain 算法决定新节点是否具有合法性. 本节将会详细介绍文中涉及到的智能合约,尤其是自动理赔合约的生成和结构以及如何实现添加节点、授予权限以及自动理赔功能.

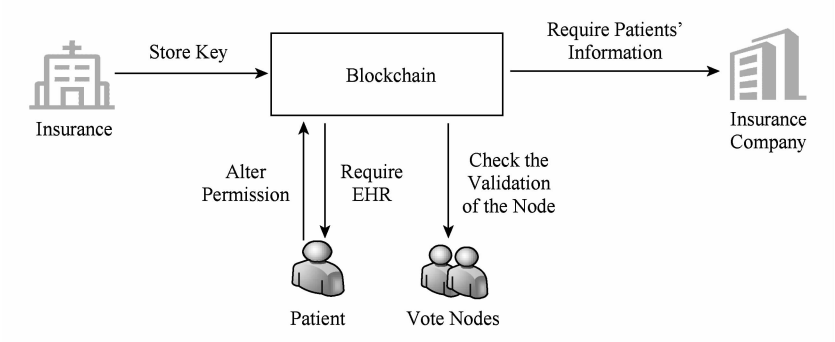

Fig. 2 Interactions between roles in the system

图 2 系统中角色之间的交互

4.2 智能合约

本文方案中的 7 个智能合约分别为共识合约、分类合约、服务历史合约、所有权合约、权限合约、代理重加密合约以及自动理赔合约,本文将详细介绍各个智能合约的结构功能.

智能合约中可以存储一定的数据,共识合约内存储了具有投票权力的节点以太坊地址,合约依靠共识算法—Quorum 算法[15]进行挖矿,Quorum 是一个在私有区块链上使用投票共识机制的算法. 当用户节点注册时,投票节点验证注册节点是否有效,若该注册节点获得大部分投票节点的选票则可以加入到系统中,该合约有效地减少了恶意节点蓄意注册的情况,也可以避免重复注册的情况发生. 除此之外,共识合约还可以对节点重写,即可以移除存在的恶意节点.

本文方案中共有 3 种类型的节点,分别是患者、医院及保险公司,共识合约不仅用于验证节点是否合法,还可以对合法的注册节点进行分类,将分类结果和节点的以太坊地址存储于分类合约中. 分类合约存储了所有节点的分类信息,当新节点注册时可先查询分类合约中是否已经存储该节点信息,由此可以简化注册过程.

为加强患者的访问控制及增加数据混淆程度,使用了 3 个层层递进的合约,即服务历史合约、所有权合约以及权限合约. 每个医院与患者节点均拥有历史服务合约,用于存储与本节点有联系的节点信息以及所有权合约的地址. 所有权合约则存储了 EHR 拥有者和 EHR 的相关信息以及权限合约的地址,该合约的主要功能就是为了追踪医院存储的数据. 患者可通过合约中存储的数据库信息查看自己的数据是否存储在合法的位置,通过查看查询链和 EHR 的 Hash 值建立数据完整性. 权限合约则存储了节点的权限信息,根据不同类型的节点定义了相应的权限. 本文方案划分了 4 种级别的访问权限:1)读权限. 指的是具有读取 EHR 的权限,一般情况下医院具有此类型的权限;2)转移权. 指的是某节点具有转移 EHR 读权限的权限,一般情况下医院具有级别的权限,如医院 A 可以向医院 B 转移某患者的 EHR;

3)所有权权限.指的是该节点具有 EHR 的控制权,一般情况下患者具有此权限,特殊情况下如患者患有心理疾病,则医院具有所有权权限;4)盲权限.指的是某节点只能查看 EHR 密文,如代理重加密节点、保险公司节点以及患有心理疾病的患者节点只有盲权限,只能查看 EHR 密文和哪些节点可访问 EHR.

在转移记录的过程中,为保证 EHR 不被泄露,使用了代理重加密技术,本文中使用了 3.5 节中介绍的基于 ElGamal 的代理重加密算法.在区块链中使用代理重加密需要一组代理重加密节点,每当生成一组代理重加密节点就会创建代理重加密合约,用于存储代理节点的信息以及 p 值密文对和盲化的明文信息.代理重加密过程如下:代理重加密合约首先获取主密钥并向各代理节点发送接受者的公钥,假设有 i 个代理节点,每个代理节点都生成一个随机的 p 值,记作 p_i,分别用主密钥和公钥对 p_i 值加密生成 p_i 密文对,并将其发送给代理重加密合约.合约利用 ElGamal 的同态乘法性质将加密的 p_i 整合成 p,返回给代理,代理从中解出盲化消息 mp 的值,再将盲化的消息发送给合约,合约利用同态乘计算接收者的新密钥.

自动理赔合约主要用于存储患者的部分信息和实现自动理赔功能.图 3 是自动理赔合约的存储结构,在此合约中,存储了该患者的以太坊地址、一组以数组形式存储的计算结果 $A[n]$,以及由对称加密密钥加密的 EHR 和患者病历中具体疾病的哈希值.该合约在患者初次就诊时由医院创建,功能类似于传统理赔中的理赔申请单,合约具体的创建过程以及理赔过程可参考 4.5 节.保险公司通过访问理赔合约中的数组 $A[n]$,使用私钥解密获取计算结果,将是否满足理赔条件的结果返回给合约,自动理赔合约根据输入判断是否理赔,若符合条件则自动向患者以太坊地址进行转账.

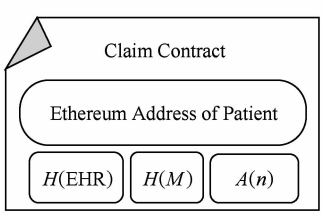

Fig. 3 The structure of the claim contract

图 3 自动理赔合约的结构

4.3 添加保险公司

添加保险公司节点的过程如图 4 所示(假设保险公司在被添加前已经有以太坊的地址和唯一的标识符),图 4 中数字代表添添保险公司的顺序.保险公司首先向患者发送添加请求,患者向分类合约发送该保险公司的以太坊地址,分类合约检索该节点是否已经存在于数据库中,若不存在,将请求的地址和类型发送给共识合约,投票节点验证收到的公共标识符是否符合请求的分类,投票完成后将结果返回给分类合约,分类合约确认授权,将该保险公司以太坊地址和分类结果存储至合约数据库中,添加保险公司节点的操作完成;若该保险公司以太坊地址已经存储于分类合约中但存储的分类与要求的分类不一致,则再次对此节点进行投票验证,将投票结果存储至合约中;若请求的分类与存储的分类一致则不再进行下一步操作.

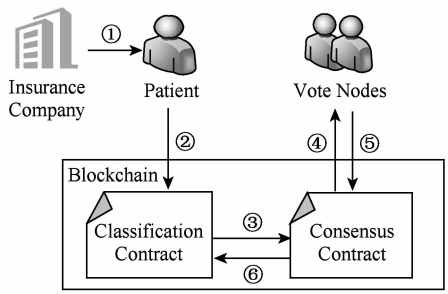

Fig. 4 Adding an insurance company node

图 4 添加保险公司节点

4.4 授予保险公司权限

在本文中,为保护患者敏感数据的隐私,保险公司不可读取明文电子健康记录,只具有读取电子健康记录密文的权限,即盲权限.根据 4.2 节介绍,一般情况下患者具有电子健康记录的拥有权,任何角色添加权限或更改权限都需获得患者同意.授予保险公司盲权限的过程如图 5 所示,保险公司首先向患者发起授权请求,患者向服务历史合约获取所有权合约的地址,再向所有权合约请求权限合约的地址,

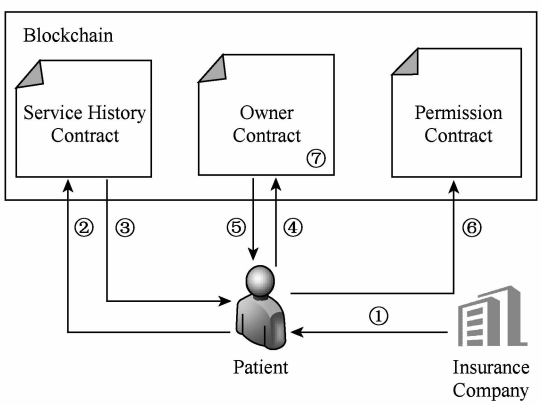

Fig. 5 The process of adding permission for insurance company

图 5 授予保险公司权限

得到该电子健康记录的权限合约地址后向其发送更改保险公司权限请求,权限合约(步骤⑦代表在合约内进行的操作)检索是否已经存储该节点的信息,若不存在该节点信息则直接将该保险公司节点的以太坊地址和请求的权限添加进合约;若合约中已经存储了该节点的信息,则比较存储的节点权限是否与请求权限相同,相同的话不做处理,不同的话更新权限合约.另外权限更改的发起者也可以是医院,当医院请求更改保险公司权限时,医院通过访问自己的服务历史合约找到权限合约的地址,权限合约检索是否已经存储节点的以太坊地址及权限,较前一种情况不同的是,在权限合约确认更改保险公司权限时,要先询问患者是否同意更改.此情况并没有越过患者,患者仍拥有电子健康记录的所有权.

4.5 自动理赔

为改进传统理赔方式中保险公司可查看患者明文数据的弊端,加强对患者隐私数据的保护,本文提出了一个基于区块链的可自动理赔的方案.在该方案中,每个患者都拥有一个由医院创建的理赔合约,该合约相当于传统理赔中的理赔申请单,智能合约一旦发布,就无法进行更改,所以保证了合约中的信息具有不可篡改性,若想调用智能合约,只需知道合约地址即可.理赔过程中保险公司主要与自动理赔合约、医院进行交互.假设在理赔之前保险公司和医院已经约定好使用的同态加密算法 E_{pk}(本文中使用的同态加密算法是 Paillier 加密),公私秘钥对为 (P_k,S_k),私钥由保险公司保存.自动理赔合约创建过程如图 6 所示,具体方案如下:

步骤 1.患者就诊,医生将该患者的 EHR 存储至医院数据库.

步骤 2.医院向保险公司请求该患者的保险信息;

步骤 3.保险公司响应请求,将该患者投保的病种分别加密(形式为 $E(M_1),E(M_2),\cdots,E(M_n)$)发送给医院;若无此患者的理赔信息,发送通知提示医院.

步骤 4.医院创建理赔合约,计算由对称加密密钥加密的电子健康记录和患者电子健康记录中所记录的疾病 M 的 Hash 值;计算 $E(M_1)\times E(M)^{-1}$, $E(M_2)\times E(M)^{-1},\cdots,E(M_n)\times E(M)^{-1}$,并将计算结果乱序放入数组 $A[n]$中,若该患者无保险信息则将数组数据域置空.将患者的以太坊地址、2 个 Hash 值以及数组 $A[n]$存储至自动理赔合约.

此时,自动理赔合约创建完成.当用户发起理赔请求时,较之传统理赔不同的是,患者不直接与保险公司交互,而是先向医院发起请求,医院收到请求向保险公司发送自动理赔合约的以太坊地址.保险公司访问理赔合约,获取数组 $A[n]$中的计算结果,使用私钥 S_k 进行解密.若解密的结果为 0,证明 2 个明文相同,可进行理赔;否则证明 2 个数据不相等,不需要理赔.由于计算结果是乱序存储在数组 $A[n]$的,保险公司无法辨别究竟是哪个明文与密文匹配,无法获取患者的任何保险信息,保证了患者的隐私安全.保险公司将计算结果及自己的以太坊地址发送给自动理赔合约,若收到的结果为 0,合约执行保险公司向患者以太坊地址转账的操作;若收到的结果不为 0,则不进行操作.

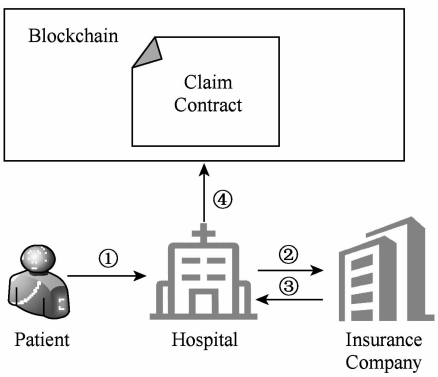

Fig. 6 The process of creating Claim Contract

图 6 创建自动理赔合约过程

据 2.4 节所述,以太坊地址是由用户公钥生成,通过以太坊地址不能推测出用户的身份信息,由此可见,当保险公司访问理赔合约时,并不能通过以太坊地址判断患者的身份,即通过智能合约实现了隐藏患者身份信息的功能;其次在理赔过程中,保险公司除解密外,还可以向医院请求加密的 EHR 和 M 以便与自动理赔合约中的 Hash 值比较,验证数据完整性.

5 分 析

5.1 正确性分析

定理 1. 当保险公司从智能合约中获取数组 $A[n]$后,可用私钥对该数组内的数据解密,以验证 2 个数据是否匹配.

证明.

$$E(M)\times E(M_1)=$$
$$(g^m\times r^{n_1})\times(g^{m_1}\times r^{n_2})^{-1}\bmod n^2=$$
$$(g^{m-m_1}\times g^{n_1-n_2})\bmod n^2=$$
$$E(M-M_1),$$

所以如果 M_1 和 M 相等,则 $D(E(M)\times E(M_1)^{-1})=D(M-M_1)=0$. 证毕.

定理 2. 在转移记录时,代理可通过计算盲化消息和用医院公钥 B 加密的 p 逆值得到由公钥 B 加密的密文.

证明. 用 $E(m,r)$ 表示计算 $E(m)$ 时用到的 r 值,$\varepsilon(m)$ 表示 m 所有可能的密文集合$\{E(m,r)|(r\in Z_q^*)\}$. 假设存在 3 个明文 m,m_1,m_2,相应的密文为 $E(m_1)=(a_1,b_1)$,$E(m_2)=(a_2,b_2)$, $E(m)=(a,b)$. 定义 3 种运算:

$$E(m_1)^{-1}\equiv(a^{-1},b^{-1}),$$

$$m'\times E(m)\equiv(a,m'b),$$

$$E(m_1)\times E(m_2)\equiv(a_1a_2,b_1b_2).$$

并且保持以下 3 种特性:

ElGamal 的逆. $E(m_1)^{-1}\in\varepsilon(m^{-1})$;

ElGamal 的并置. $m'\times E(m,r)=E(m'm,r)$;

ElGamal 的乘法. 如果 $r_1+r_2\in Z_q^*$,那么:

$$E(m_1,r_1)\times E(m_2,r_2)\in\varepsilon(m_1m_2).$$

由代理重加密算法步骤 5 可知,$E_B(m)=(mp)\times E_B(p)^{-1}$,验证过程如下:

$$E_B(m)=(mp)\times E_B(p)^{-1}=$$
$$(mp)\times E_B(p^{-1},-r)=$$
$$E_B(mpp^{-1},-r)=$$
$$E_B(m-r)\in\varepsilon_B(m).$$

由此可知通过计算 $mp\times(E_B(p,r))^{-1}$,可得到用公钥 B 加密后的密文 $E_B(m)$. 证毕.

5.2 安全分析

定理 3. 保险公司从数组 $A[n]$中成功推断出患者疾病的概率为 $1/n$,其中 n 是患者保险信息所包含的疾病数量.

证明. 为了验证患者疾病 M 与保险信息中的疾病是否相符合,医院将计算结果乱序的存储在数组 $A[n]$中. 理赔时,保险公司用私钥解密访问智能合约获取的数组 $A[n]$. 若保险公司想获得额外的信息,可以通过计算 $D(E(M)\times E(M_1)^{-1}/E(M_1)^{-1})$ 获得 M,但由于医院将计算结果乱序存入 $A[n]$,每个数据的可能性是相等的,所以如果存在 n 个数据,猜测正确的概率为 $1/n$,由此数据 M 是安全的.

证毕.

定理 4. 医院不能推断出患者所投的保险信息,即不能从 $E(M_1),E(M_2),\cdots,E(M_n)$中获取到关于 $M_1,M_2,\cdots,M_n$ 的任何一条信息.

证明. Paillier 加密方案是语义安全当且仅当区分 n^2 的 n 次剩余与非 n 次剩余是困难的.

充分性证明. 假设 M_0,M_1 是 2 个已知的消息,C^* 是其中一个(设为 M_β)的密文,即 $C^*\leftarrow g^{M_\beta}\times r^n \bmod n^2$,$C^*\,g^{-M_\beta}\equiv r^n \bmod n^2$ 是 n^2 的 n 次剩余,而 $C^*\,g^{-M_{1-\beta}}\equiv g^{M_\beta-M_{1-\beta}}\times r^n \bmod n^2$ 是 n 次非剩余. 因此敌手能够区分 C^* 对应哪个消息,可以区分 n 次剩余和 n 次非剩余,这与判定合数剩余类假设是困难的矛盾.

必要性证明. 假设有 2 个消息 M_0,M_1,同上 C^* 是 2 个消息中的其中一个密文,$C^*\,g^{-M_\beta}\equiv r^n \bmod n^2$ 是 n^2 的 n 次剩余,而 $C^*\,g^{-M_{1-\beta}}\equiv g^{M_\beta-M_{1-\beta}}\times r^n \bmod n$ 是 n 次非剩余,由于判定合数剩余类假设是困难的,所以无法区别 $C^*\,g^{-M_\beta}$ 和 $C^*\,g^{-M_{1-\beta}}$,因此是语义安全的. 由此证明医院无法从获得的消息中获得更多的知识. 证毕.

5.3 隐私保护分析

保险公司在理赔过程中只能获取到患者的以太坊地址,难以推断出患者的身份信息,这在一定程度上保护了患者的身份隐私,也可以结合签名技术(如环签名)保护交易发起方的隐私;所有放在区块链上的信息都使用了加密技术进行加密,降低了未授权的节点访问敏感信息的可能性;文中使用服务历史合约、所有权合约和权限合约来分离信息,每个合约存储一定的信息且权限合约是由所有权合约创建,权限合约由服务历史合约创建,若想更改权限必须层层访问,增加了数据混淆的程度;任何节点在注册时共识合约和分类合约都会验证该节点的合法性,保证了所有参与到区块链的节点都是合法的;假设存在恶意节点试图访问患者的 EHR,访问 EHR 之前必须经过权限合约的核查,若该节点的确有读取明文的权限才可获取到解密 EHR 的密钥,如此可以防止非法或恶意节点访问患者隐私数据;最后在所有交互过程中,除医院和授权者外其余各参与方均无法获取到任何明文消息. 在转移电子健康记录时,使用了代理重加密技术,允许 EHR 在密文状态下被传输,并分离了对称密钥,进一步提高了安全性;理赔过程中,医院收到的仅是患者保险信息的密文,在加密方案安全的情况下医院无法对其解密;保险公司只能获取密文状态的 EHR 和疾病,无法获取相关解密密钥,保证了患者敏感数据的安全. 虽然以太坊地址和个人信息无关,使得攻击者很难从以太坊地址推断出个人 ID,在一定程度上保护了用户隐私,但通过分析特定节点访问医院或保险公司的频率可能会推断出此节点的信息,未来工作中可以继续在这方面做出努力.

5.4 性能分析

如2.2节所述,gas是以太坊计算成本的度量单位,在分析性能时主要考虑gas的消耗.文献[13]中添加电子健康记录的成本高于文献[9]中MedRec系统添加电子健康记录的成本,在对比时主要比较区块链之外和区块链上的操作.在比较区块链之外的操作时,Ancile比MedRec多一个使用解密密钥解密查询链的操作,所以Ancile在区块链之外的成本比MedRec略高一些,但就实施模型来说性能成本还是很低的.在对比区块链上的操作时,由于Ancile需要发送内部交易连接不同合约,使用合约创建其他合约,所以gas成本会达到gas limit的一半,成本高于MedRec.

本文中创建自动理赔合约时,在区块链外的操作主要有保险公司发送患者保险保险信息、医院使用同态加密算法进行计算以及计算散列值,性能成本较低.区块链上的操作主要是医院将数据存储至合约中,保险公司从合约中检索数据,检索和存储单个数据的gas成本取决于存储/读取数据的大小,所以gas成本小于gas limit的一半.由于文中涉及了多个智能合约,且医院为每个患者都生成一个自动理赔合约,这意味着将存在大量的智能合约,导致节点需要大量的计算能力和时间来执行每个合约,效率低下.下一步工作中可以在不降低安全性的前提下适当减少智能合约的数量提高效率.总体来说,本文方案在性能成本和效率方面稍有不足,但实现了更多功能,如自动理赔功能等,尤其是在保护患者隐私方面,增加了数据混淆程度和攻击难度.

6 总 结

由于云环境下患者没有对数据的控制权且在传统理赔过程中保险公司具有查看患者隐私数据的弊端,因此实现基于区块链的数据隐私保护,增强患者的控制权以及在隐私保护的前提下实现正确理赔具有重要意义.考虑到区块链具有不可篡改性和去中心化等特点,以及智能合约使以太坊更具扩展性,本文提出了基于区块链的EHR隐私保护方案.本方案将智能合约及同态加密技术相结合,实现了患者拥有数据所有权及在患者匿名前提下自动理赔的功能.通过对安全和性能方面分析得知该方案不会在交互过程中泄露任何敏感数据,但由于智能合约数量较多,性能方面还有待提高.

如何在保证隐私数据安全的前提下提高效率,减少智能合约的数量及如何在恶意模式下实现安全交互是下一步研究的主要工作.

参 考 文 献

[1] Rodrigues J J P C, De La Torre I, Fernández G, et al. Analysis of the security and privacy requirements of cloud-based electronic health records systems [J]. Journal of Medical Internet Research, 2013, 15(8): 418-426

[2] Narayan S, Gagné M, Safavi-Naini R. Privacy preserving EHR system using attribute-based infrastructure [C] //Proc of the 2010 ACM Workshop on Cloud Computing Security. New York: ACM, 2010: 47-52

[3] Xhafa F, Feng Jianglang, Zhang Yinghui, et al. Privacy-aware attribute-based PHR sharing with user accountability in cloud computing [J]. The Journal of Super Computing, 2015, 71(5): 1607-1619

[4] Samarin A. Electronic health records implementation with blockchain, bpm, ecm, and platform [EB/OL]. (2016-07-01) [2018-06-04]. https://improving-bpm systems. blogspot. com/2016/07/electronic-health-records-ehr. html

[5] Xue Tengfei, Fu Qunchao, Wang Cong, et al. A medical data sharing model via blockchain [J]. Acta Automatica Sinica, 2017, 43(9): 1555-1562 (in Chinese)
(薛腾飞, 傅群超, 王枞, 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)

[6] Brodersen C, Kalis B, Leong C, et al. Blockchain: Securing a new health interoperability experience [EB/OL]. [2018-06-11]. http://www. truevaluemetrics. org/DBpdfs/Technology/Blockchain/2-49-accenture_onc_blockchain_challenge_response_august8_final. pdf

[7] Angraal S, Krumholz H M, Schulz W L. Blockchain technology: Applications in health care [J]. Circulation: Cardiovascular Quality and Outcomes, 2017, 10(9): 91-94

[8] Ivan D. Moving toward a blockchain-based method for the secure storage of patient records [EB/OL]. [2018-06-04]. https://www. healthit. gov/sites/default/files/9-16-drew _ ivan _ 20160804_blockchain_for_healthcare_final. pdf

[9] Ekblaw A, Azaria A, Halamka J D, et al. A case study for blockchain in healthcare: "MedRec" prototype for electronic health records and medical research data [C] //Proc of IEEE Open & Big Data Conf. Piscataway, NJ: IEEE, 2016: 25-30

[10] Linn L A, Koo M B. Blockchain for health data and its potential useb in health it and health care related research [EB/OL]. [2018-06-04]. https://www. healthit. gov/sites/default/files/11-74-ablockchainforhealthcare. pdf

[11] Halamka J D, Ekblaw A. The potential for blockchain to transform electronic health records [EB/OL]. [2018-06-04]. http://www. medtronic. com/content/dam/medtronic-com/global/Corporate/Initiatives/harvard-business-review/downloads/insight-center/potential-for-blockchain_paper_hbr. pdf

[12] Culver K. Blockchain technologies: A whitepaper discussing how the claims process can be improved [EB/OL]. [2018-06-04]. https://www.hhealthi.gov/sites/default/files/3-47-whitepaperblockchainforclaims_v10.pdf

[13] Dagher G G, Mohler J, Milojkovic M, et al. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology [J]. Sustainable Cities and Society, 2018, 39(4): 283-297

[14] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system [EB/OL]. [2018-06-11]. http://www.bitcoin.org/bitcoin.pdf

[15] Pilkington M. Research handbook on digital transformations [M]. Northampton: Edward Elgar Publishing, 2016: 225-264

[16] Lobban T. Quorum chain consensus [EB/OL]. (2017-12-01) [2018-06-04]. https://github.com/jpmorganchase/quorum/wiki/QuoQuorumCh-Consensus

[17] Buterin V. A next-generation smart contract and decentralized application platform [EB/OL]. (2013-11) [2018-06-04]. https://github.com/ethereum/wiki/wiki/White-Paper

[18] Zhou Yonggeng, Li Feng, Tao Yufei, et al. Privacy preservation in database applications: A sruvey [J]. Chinese Journal of Computers, 2009, 32(5): 847-861 (in Chinese)
(周水庚, 李丰, 陶宇飞, 等. 面向数据库应用的隐私保护研究综述[J]. 计算机学报, 2009, 32(5): 847-861)

[19] Clifton C, Kantarcioglu M, Vaidya J. Defining privacy for data mining [C] //Proc of National Science Foundation Workshop on Next Generation Data Mining. Menlo Park, CA: AAAI, 2002: 126-133

[20] Zhu Liehuang, Gao Feng, Shen Meng, et al. Survey on privacy preserving techniques for blockchain technology [J]. Journal of Computer Research and Development, 2017, 54(10): 2170-2186 (in Chinese)
(祝烈煌, 高峰, 沈蒙, 等. 区块链隐私保护研究综述[J]. 计算机研究与发展, 2017, 54(10): 2170-2186)

[21] Rivest R L, Shamir A, Adleman L. A method for obtaining digital signatures and public-key cryptosystems [J]. Communications of the ACM, 1978, 21(2): 120-126

[22] Paillier P. Public-Key cryptosystems based on composite degree residuosity classes [J]. Eurocrypt, 1999, 547(1): 223-238

[23] López-Alt A, Tromer E, Vaikuntanathan V. On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption [C] //Proc of the 44th ACM Symp on Theory of Computing. New York: ACM, 2012: 1219-1234

[24] Zhou L, Marsh M A, Schneider F B, et al. Distributed blinding for distributed ElGamal re-encryption [C] //Proc of IEEE Int Conf on Distributed Computing Systems. Alamitos, CA: IEEE Computer Society, 2005: 824-824

Xu Wenyu, born in 1995. Currently master candidate of information science and engineering of Shandong Normal University. Her main research interests include block chain and data security and privacy.

Wu Lei, born in 1980. PhD. Currently associate professor at Shandong Normal University. His main research interests include cryptology and cloud computing security.

Yan Yunxue, born in 1994. Currently master candidate of information science and engineering of Shandong Normal University. Her main research interests include data security and privacy, Lattice cryptography.