基于区块链的链上数据安全共享体系研究

刘彦松[1]，夏琦[1]，李柱[1]，夏虎[1]，张小松[1]，高建彬[2]
1. 电子科技大学计算机科学与工程学院，四川 成都 611731；
2. 电子科技大学资源与环境学院，四川 成都 611731

摘要
针对人们在日益增长的数字化交互过程中越来越多地出现隐私直接或间接泄露的问题，主要研究基于区块链网络建立一套链上数据安全共享体系，基于密文策略的属性加密的访问控制算法以及同态加密算法实现链上数据的可靠共享，提出了一种链上数据共享架构，最后进行了仿真实验，并分析了实验结果。这项工作有效解决了恶意参与方利用区块链的交易透明性进行数据分析的问题，并保证了用户数据在共享流程中的隐私安全。

关键词
数据共享；区块链；基于密文策略的属性加密；同态加密；隐私安全

中图分类号：TP309.2 文献标识码：A doi: 10.11959/j.issn.2096-0271.2020046

Research on secure data sharing system based on blockchain

LIU Yansong[1], XIA Qi[1], LI Zhu[1], XIA Hu[1], ZHANG Xiaosong[1], GAO Jianbin[2]
1. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
2. School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract
In view of the increasing number of people in the digital interaction process of direct or indirect disclosure of privacy issues, the establishment of a set of data security sharing system based on blockchain network, the access control algorithm based on ciphertext-policy attribute-based encryption and the homomorphic encryption algorithm provides reliable chain data sharing were mainly researched. Data sharing on chain architecture for data sharing on the chain was proposed. Finally, simulation experiments were carried out, and the results of experimental data were analyzed. The work effectively solves the problem of malicious parties using the transaction transparency of blockchain for data analysis and ensures the privacy security of user data in the sharing process.

Key words
data sharing, blockchain, ciphertext-policy attribute-based encryption, homomorphic encryption, privacy security

1 引言

随着互联网时代的发展，用户对网络安全的需求越来越大，生活中无时无刻不在产生大量网络支付和社交活动等数据，而在不同的行业中，多领域、多企业的复杂交互式业务场景也会涉及多方数据的汇总，这些海量的异构数据不仅蕴含巨大价值，也隐含大量的个人隐私，继而产生数据权属与数据安全等多方信任问题。在医疗领域，以新型冠状病毒肺炎疫情为例，为了控制疫情的扩散以及精准把控疫情的进展，各地政府、高校及研究机构搭建了新型冠状病毒大数据交叉学科研究平台，对患者的数据进行分析和统计，在这个过程中，一旦一个环节出错就会导致数据泄露。

随着区块链技术越来越多地出现在人们的视野，其去中心化的特性颠覆性地解决了许多“信任”问题，为多企业、多单位共同参与项目提供了一致性的保障，为数据安全、数据增值、成果认定提供了平台支撑，促使各参与方更精于合作，更专注于研究工作本身，减少数据泄密带来的风险。

然而，区块链技术的交易公开性也有弊端，数据请求方的数据属性集合被封装在交易信息中，使用共识算法将其广播到区块链网络，使得所有节点公开可见。但数据属性集合极易被恶意参与方盗用，并且生成对应的用户密钥，从而窃取区块链上（以下简称“链上”）的交易数据。可见区块链架构本身并不一定是高可用的，维护区块链系统的可运维性不仅需要在技术上突破，更应从法律、行业实践及标准化层面加以约束[1]。

本文针对当前第三方传统中心化数据共享平台和机构上存在的数据管理、多方交互数据不可信以及当前区块链共享平台交易公开等关键问题，将医疗数据共享中患者体检报告中的各项指标数据作为本文的数据分析主体，研究智能合约（smart contract）、基于属性加密的访问控制算法、同态加密（homomorphic encryption）算法等，探索基于区块链技术的链上数据安全共享体系，设计基于智能合约及密文策略的属性加密的访问控制方案以及基于同态加密的数据处理模块，旨在为用户信息提供加密保护，合理管理共享数据，可靠保证用户的数据主权与数据完整性，最终建立一套完备的数据权属体系与数据安全保障规范。

2 研究现状

数据管理是指对各类数据进行采集、存储、分类、检索和传输等的过程。随着计算机以及网络技术的发展[2]，数据管理技术从最早的20世纪50年代的人工数据检索方式，到20世纪60年代的文件系统管理方式，最后在20世纪60年代后期演化为现在一直沿用的数据库系统管理方式。而针对数据应用场景与应用需求，数据管理又可被细分为科学数据管理、信息管理、内容管理[3]等。本文提到的基于区块链的链上数据安全共享体系更多地依赖于科学数据管理以及内容管理中的关键技术，以此保证数据的安全共享与有效利用。

美国是较早对开放数据进行科学管理与共享的国家[4]，通过颁布《信息自由法》和《版权法》等法案将政府作为数据开放共享的主体，其通过政府主动开放自身数据并吸引企业投资的方式来深化政府数据的创新应用，形成政府主导模式，同时也对数据进行监管，防止对其的二次泛滥利用。欧洲各国同样也针对政府和企业出台

了很多科学数据管理与共享的相关政策，旨在保障数据的质量与精度。

我国的科学数据管理与共享工作起步略晚，虽然已开展了关于地质调查、气象、海洋、水文、环境和地震等方面的数据监测，监测得来的数据量很大，但科学数据的有效利用率始终很低。直到2002年科学数据共享工程在科学技术部和相关管理部门的共同努力下正式启动，我国的科学数据管理与共享工作才有了较快发展，为科学研究以及大数据分析奠定了基础，但对科学数据进行安全有效的管理与共享依旧是需要持续研究的课题。

目前，我国在科学数据管理与共享利用方面取得了一些成绩，但科学数据共享观念淡薄，很多科研院所和政府管理职能部门的数据共享仍缺乏有效的政策法规保障。而大数据时代的到来使得传统关系型数据库也显得力不从心，NoSQL技术在一定程度上满足了大数据时代的数据管理需求[2]，但面对数据量的急剧增加以及数据敏感度的日益提升，数据的一致性、可用性以及安全性等方面仍限制了NoSQL的发展。如何在保证数据安全的基础上对数据进行安全共享，继而有效利用，是数据管理领域的一个重要挑战。基于区块链的数据共享技术为数据共享的创新开拓了新的思路。

2016年，Sun J J等人[5]提出了一种基于区块链的共享服务的概念模型，用于促进数据共享。然而，研究仅仅侧重于概念和模型，没有提出实际解决方案。Yue X等人[6]提出一个基于区块链的医疗数据架构，在不破坏隐私的前提下，帮助用户安全和轻松地掌握并且分享自己的医疗数据。他们提出一种基于目的的访问模型，实现了患者持有并掌握自己的医疗数据。虽然他们提到了安全多方计算的潜在前景，却没有提出具体的实现方案。

2017年，Zikratov I等人[7]系统地讨论了如何在去中心化环境下使用区块链存储、取回和共享文件。这是首次具体使用区块链来实现数据完整性的方案，讨论的主要内容包括定义交易信息、区块信息等具体的实现措施。此外，该方案也指出以区块链为底层的数据平台存在的无法抵御量子攻击的潜在威胁。同年，Shafagh H等人[8]首次提出一种基于区块链的支持细粒度访问控制和共享数据的方案。与传统方案不同，该方案赋予用户数据的完全持有权，而不是委托一个可信中心（云服务器）来操作。该方案的性能适中，主要针对物联网设备。显然，这不能满足基于区块链的数据平台对海量交易数据的处理需求。Xia Q等人[9-10]提出了一个云环境下的基于区块链的医疗数据共享框架，该框架充分解决了在云环境中存储的敏感数据相关访问控制问题，但普适性还稍有欠缺。

2018年，Zhang P等人[11]提出一种FHIRChain原型，用于向用户提供互动性更高的医疗诊断服务。该方案实现了用户身份识别、认证、安全数据交换等一系列要求。同年，Zhang A等人[12]提出一种基于安全和隐私保护的、基于区块链的个人医疗信息共享方案。通过应用基于公钥密码学的可搜索加密，该方案允许用户的医疗信息被安全和受控地访问及用于改善医疗诊断服务。李康等人[13]归纳并总结了基于零知识证明的隐私保护方案在区块链技术中的应用。祝烈煌等人[14]介绍了以混币机制为代表的区块链交易数据隐私保护方案。

2019 年，Muzammal M等人[15]将区块链和数据库结合，提出一种去中心化、分布式和可审计的数据系统。该方案是第一个基于区块链的防篡改的数据库，提供了对分布式数据的高效检索。同年，张超等人[16]设

计了一个基于实用拜占庭容错算法的联盟式医疗区块链系统，该系统能够防止数据被泄露和篡改。

在数据流通方面，闫树等人[17]在区块链改造授权存证环节、数据溯源和智能合约实现等研究领域梳理了区块链技术在数据流通中的应用。

3 相关背景技术

3.1 区块链技术

区块链的结构构想早在20世纪90年代就被提出，而到2008年区块链才真正进入了大众的视线。区块链技术作为一个由多方共同维护、去中心化的分布式账本技术，核心在于通过对等（peer to peer，P2P）网络协议、共识算法、非对称加密、哈希等关键技术解决数据传递与交换过程中的信任问题。区块链的链式结构是一种将数据区块按时间戳顺序相连，进行数据存储与验证的一种数据结构，是一种凭借共识算法对数据进行广播交易，基于密码学原理保证数据传输和访问的安全性；具备难以篡改性和难以伪造性的分布式账本技术。其可利用智能合约来编程和操作数据。

3.2 智能合约

智能合约是20世纪90年代由尼克·萨博（Nick Szabo）提出的理念。Nick Szabo将其描述为“一套以数字形式描述的承诺，以及合约参与方履行这些约定的协议”。智能合约是一种链上代码，是以信息化方式传播、验证或执行合同的计算机协议，解决了传统交易系统中需要第三方机构进行交易监管的问题。智能合约不仅是形式上的数字化合约，更是一组可以在一台计算机或计算机网络中按预置设定自动执行规则的契约。智能合约由预定义的多行代码和用于执行该代码的软件组成，软件中内置了合同条款和输出结果的内容。

3.3 基于密文策略的属性加密

在基于密文策略的属性加密（ciphertext-policy attribute-based encryption，CP-ABE）算法[18]中，密文对应于一个访问控制结构，密钥则对应于属性的集合，当访问用户的属性能够满足对应的访问控制结构时，解密才能成功。密钥通常是由用户根据自身条件和属性从属性机构中获取的，而加密者基于消息来设计访问控制结构，即由发送方规定访问密文的策略，将属性集合与访问资源关联，接收方可以根据自己的授权属性访问密文信息。CP-ABE算法流程主要由4个阶段构成。

- 初始设置：基于CP-ABE算法进行随机初始化，初始化数据包含隐藏的安全参数λ、系统公钥PK以及系统主密钥MK，由密钥分发中心进行密钥的初始化分发。

$$\text{Input}(\lambda)\rightarrow(\text{PK},\text{MK}) \quad (1)$$

- 加密阶段：加密阶段仍是一个随机算法，算法输入为系统公钥PK、待加密消息m和与访问策略相关联的访问控制结构A_{cp}，生成基于属性加密的密文Em。只有拥有访问策略的请求者才能解密密文Em。

$$\text{Encrypt}(\text{PK},m,A_{\text{cp}})\rightarrow\text{Em} \quad (2)$$

- 密钥生成：密钥生成阶段仍是一个随机算法，输入一组属性Y、系统主密钥MK、系统公钥PK，输出一个供数据请求方使用的解密密钥UK。

$$\text{KenGen}(\text{MK},\text{PK},Y)\rightarrow\text{UK} \quad (3)$$

- 解密阶段：基于访问结构A_{cp}加密的密文Em、对应属性组Y的解密密钥UK和系统公钥PK，若$Y\in A_{\text{cp}}$，则输出消息m。

$$\text{Decrypt}(\text{PK},\text{UK},\text{Em})\rightarrow m \quad (4)$$

3.4 同态加密技术

同态加密技术根据密文运算的次数与种类以及其时间顺序的发展阶段分为部分同态加密(partial homomorphic encryption, PHE)、类同态加密(somewhat homomorphic encryption, SHE)、全同态加密(fully homomorphic encryption, FHE)[19]。目前同态加密技术在云计算环境中的可信计算与基于密文的检索方面应用十分广泛,用户隐私数据在云端始终以密文形式进行存储,云服务商没有密钥则无法获取用户的真实明文数据。而同态加密技术则是用户在云环境中进行数据挖掘与分析计算的可靠助力与安全基础。本文将区块链技术与同态加密技术结合,通过引入Paillier加密算法[20]来实现共享数据的密文处理。

Paillier密码机制如下。

选取两个大素数p和q,计算$n=pq$,$\lambda(n)=\mathrm{lcm}(p-1,q-1)$,其中$\mathrm{lcm}(a,b)$表示求$a$和$b$的最小公倍数。随机选取参数$g$,$g\in Z_{n^2}^*$,且$n\mid \mathrm{ord}_{n^2}(g)$。令$S_n=\{u<n^2\mid u\equiv 1 \bmod n\}$,对于任意$u\in S_n$,定义函数$L(u)=\dfrac{u-1}{n}$,则生成公钥$(n,g)$,私钥$(p,q)$或$\lambda$。

加密算法:对于任意明文$m\in Z_n$,随机选取整数$r\in Z_n^*$,则加密后的密文为$C=E(m)=g^m\cdot r^n \bmod n^2$。

解密算法:利用私钥解密得到明文$m=\dfrac{L(c^\lambda \bmod n^2)}{L(g^\lambda \bmod n^2)}\bmod n$。

4 DSOC架构

本文基于区块链技术、基于密文策略的属性加密技术以及同态加密技术,针对当前医疗信息共享中的数据易泄露,以及通过分析用户相关数据之间的关联性来获取用户隐私等问题,提出链上数据共享(data sharing on chain, DSOC)架构,并以医疗领域中的数据结构为背景,阐述了DSOC架构在医疗数据共享中的应用流程。

4.1 流通数据分析

在DSOC架构中采用的数据案例为传统医疗系统中使用的患者体检报告。患者体检报告数据见**表1**(患者张三,病历号为1234567,年龄25岁,性别男)。

在传统医疗数据平台中,用户体检数据的各项指标皆为明文显示,在数据共享过程中极易泄露用户隐私。本文提出的DSOC架构对用户体检数据的各指标值进行加密,利用哈希算法中的MD5算法将{患者,年龄,性别}进行单项加密显示,将用户病历号作为ID。首先数据拥有者设置访问结构A_{cp},以患者体检报告数据为例,数据请求方的数据属性集合为{所属机构,所属部门,职务,信用评级},用户属性集合为{XX医院,心电科,主治医师,极好}。一旦数据属性集合与数据拥有者设置的访问结构不匹配,则驳回数据请求,此次数据共享失败;若匹配,则可以进行下一步数据分析。在数据分析中,本文引入加法同态加密算法,通过将结果值与最大正常参考值进行密文的加减法计算,对最终结果进行同态加密的范围型验证(验证是否超出正常值)。而数据请求方只会获得该指标是否超标的结果,并不清楚具体数值以及超出正常值的范围。

4.2 系统模块

DSOC架构由用户模块、密钥分发模块、访问控制策略模块以及数据分析模块构成。

用户模块由数据拥有者及数据访问者

构成。其中数据拥有者将可共享的数据发送至区块链网络，同时可对数据访问者设置访问策略。数据访问者则基于自身的数据属性集合来判定是否拥有访问数据的权限，取得数据访问权限后可对数据进行进一步的分析操作。

密钥分发模块主要由密钥分发中心进行控制。密钥分发中心在访问控制阶段和数据处理阶段都会参与密钥分发和数据加密过程。在访问控制阶段，因共享数据集庞大，考虑到加密效率问题，采用对称加密算法对共享数据属性集进行加密。为了保证较高的加解密效率，同时有效避免因暴力破解而导致的密钥泄露，本模块采用对称加密算法AES-192加密共享数据，该生成密钥由一对192位的随机数进行加法运算后执行哈希算法SHA192得出。

访问控制模块基于CP-ABE算法实现基于密文策略的属性加密。其中密钥分发中心结合安全参数生成系统公钥、系统主密钥以及用户的解密密钥。数据拥有者设置访问策略，并将公钥和访问策略嵌入密文中。数据访问者使用解密密钥来获取密文。

数据分析模块基于Paillier算法对数据进行加法同态加密，对最终结果进行范围型验证，得出结论，并将结论告知数据访问者。

4.3 总体架构与算法流程

DSOC架构如**图1**所示。

（1）初始阶段

为保证算法运行效率，采用对称加密算法来实现共享数据属性集的加密，继而生成对称密钥。数据拥有者随机生成192位的密钥对(k_1,k_2)，则对称密钥Ak=SHA192$(k_1 \oplus k_2)$。之后数据拥有者通过调用智能合约模块上传$(k_1$,DataSet)至区块链网络。其中DataSet为数据拥有者可进行共享的数据集合。

表1 患者体检报告数据

检查项目	结果值	单位	最大正常参考值
红细胞比容（HCT）	34%		45%
血红蛋白（HGB）	12	g/dL	15
平均血红蛋白浓度（MCHC）	31	g/dL	36
血小板（PLT）	207	K/uL	500
球蛋白（GLOB）	50	g/L	45

数据访问者调用相应智能合约查看链上数据集信息，根据数据描述选择自己需要的数据集合，并向数据拥有者发送包含目标数据属性集合标识的哈希值，数据拥有者收到数据请求后，对比收到的哈希值，并查找数据访问者想要访问的数据集合，开始制定访问策略A_{cp}。

（2）访问控制阶段

密钥分发中心基于CP-ABE算法进行初始化，初始化数据包含隐藏的安全参数、系统公钥以及系统主密钥，即$(\lambda$,PK,MK)。

密钥分发中心通过调用智能合约将PK存储于链上，而数据拥有者和数据访问者从链上获取PK。

数据拥有者获取PK后执行CP-ABE算法的加密操作，生成密文Ek_2，并通过智能合约将生成的密文Ek_2上传至区块链。

$$\text{Encrypt} \rightarrow Ek_2 \qquad (5)$$

数据访问者将目标属性集合$DataSet_{partial}$存储于链上，并利用智能合约触发密钥分发中心，进行下一步生成解密密钥的操作。

密钥分发中心从链上获取目标属性集合$DataSet_{partial}$，并基于CP-ABE算法执行密钥生成的功能函数来生成解密密钥UK。

$$\text{KenGen}(MK,PK,DataSet_{partial}) \rightarrow UK \qquad (6)$$

密钥分发中心将UK存放于链上。

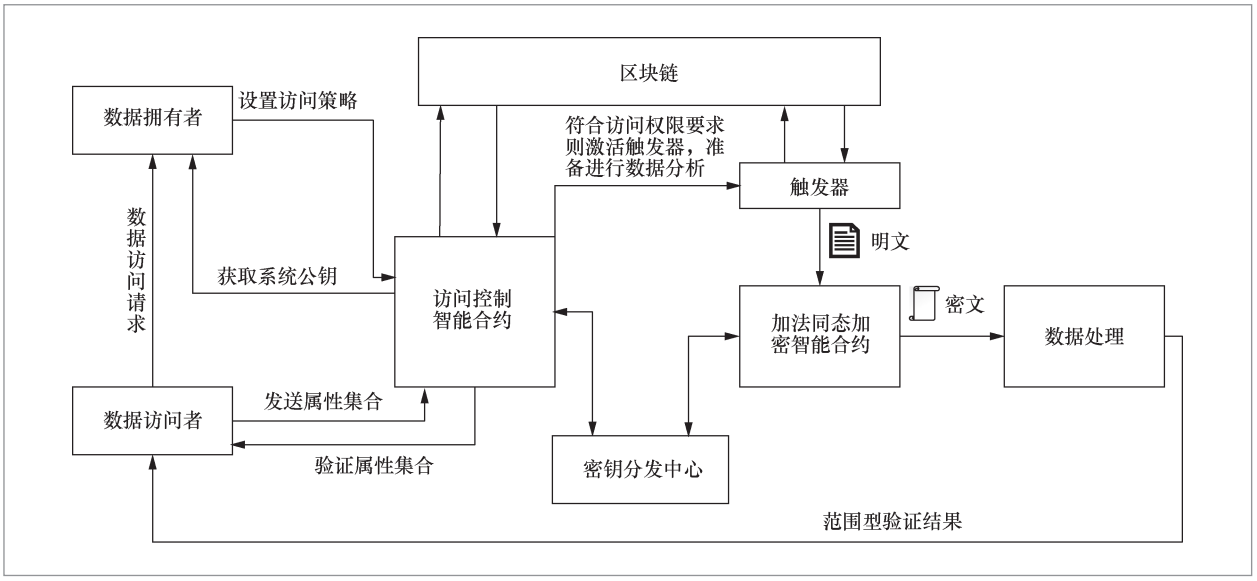


图1 DSOC架构

数据访问者通过调用智能合约从链上获取UK后，基于CP-ABE算法执行解密阶段的功能函数，若符合访问控制策略A_{cp}，则自动获得密钥k_2。

$$\text{Decrypt}(PK,UK,Ek_2)\rightarrow k_2 \quad (7)$$

这时数据访问者将k_2上传至链上，通过智能合约计算SHA192($k_1\oplus k_2$)是否等于Ak。若相等，则数据请求方是合规的用户，激活触发器进入数据分析阶段。

（3）数据分析阶段

触发器从链上取得目标数据集后，通过调用加法同态加密智能合约对数据进行加密处理。这时，密钥分发中心进行密钥生成。首先生成两个大素数p和q，且$n=p\times q$，则欧拉函数$\phi(n)=(p-1)(q-1)$，Carmichael函数$\lambda(n)=\text{lcm}(p-1,q-1)$（为方便描述，用$\lambda$代替$\lambda(n)$），任意选取$g\in Z_{n^2}^*$，满足$\gcd\left(\frac{g^{\lambda}\bmod n^2-1}{n},n\right)=1$，定义集合$S_n=\{u\mid 0<u<n^2,u=1\bmod n\}$，对于任意$u\in S_n$，定义函数$L(u)=\frac{u-1}{n}$，选取哈希函数$h$，$h:\{0,1\}^*\rightarrow\{0,1\}^k\subset Z_{n^2}^*$。因此生成公钥$(n,g)\rightarrow$PAK，生成私钥$(p,q)$或$\lambda\rightarrow$SAK。

密钥分发中心生成随机参数r，对目标数据m_1进行加密，得到密文C_1。

$$\text{Encrypt}(PAK,r,m_1)\rightarrow C_1 \quad (8)$$

同样，对数据将要进行操作的阈值（例如医疗数据中感冒患者的血红蛋白的正常值）m_2进行加密，得到密文C_2。

在数据处理模块中进行加法同态加密运算。

$$\text{Add}(PAK,C_1,C_2) \quad (9)$$

最后用户通过私钥对运算结果进行解密。

$$\text{Decrypt}(SAK,r,C) \quad (10)$$

最终结果通过范围验证$m\leqslant 0$或$m\geqslant 0$（m为目标数据m_1与阈值数据m_2的差值）来判断该患者的特定指标是否在正常范围之外。

4.4 智能合约算法实现

算法1 访问控制算法（Access）

Input: DataSet, DU's address

Output: true/false, error

1: **If** verify(DU's address) **then**

2: **while** DataSet == Acp **do**

```
3:    send UK to DU's address
4:    decrypt(PK, UK, DataSet)
5:    return Data
6:    goto Analyze
7:  end while
8: else
9:  return false, error
10: end
```

在访问控制算法中首先验证数据访问者DU是否为合规节点，若是，则密钥分发中心把用户的解密密钥传给DU。之后验证DU的属性集合，若属性集合与数据拥有者设置的访问控制策略相匹配，则可获得共享数据，并跳转到数据处理模块。

算法2　同态加密算法（Analyze）

```
Input: Data, trigger's status
Output: true/false, error
1: for each trigger's status == true
2:  getPeer(DU's address)
3:  DataSet = Paillier.enc(a, PAK, Data)
4:  Send (abi, address, Data) to DU's address
5: end for
6: Result = Paillier.add(PAK, Data)
7:  If Paillier.enc(a, SAK, Result)>0 & Paillier.enc(a, SAK, Result)==0 then
8:    Return true
9:  else
10:    Return false, error
11: end
```

在同态加密的数据处理算法中，首先需要判定触发器状态，数据经过访问控制模块处理后将激活触发器，使得数据直接传入数据处理模块中。首先需要获得数据访问者DU的地址，然后生成交易，交易包含密文数据Data、合约地址address和合约接口abi，之后对密文数据进行基于Paillier算法的同态加密计算。判断最大正常参考值与结果值的差，若该差值大于0，则返回true，即结果正常；若小于0，则返回false，即该项指标结果异常。

5 实验与分析

5.1 实验环境准备

本文提出的链上数据共享方法基于迅鳐区块链即服务（Ray blockchain as a service，RayBaaS）平台[21]开发智能合约，实现数据的安全共享。RayBaaS平台是拥有自主知识产权的分布式账本系统，是区块链3.0标准下的区块链应用内核，可以让用户简便、快捷、高效地构建基于区块链的服务和应用。硬件设备采用Intel® CoreTM i7、8核CPU以及16 384 MB的RAM的移动工作栈进行实验。区块链底层平台基于CentOS7.6操作系统进行部署，并安装了Java1.8.0、Docker18.09、MySQL5.7.21等组件。本实验所用数据截取自个人用户的体检数据报告。

RayBaaS平台提供了快速部署区块链网络的功能，在用户端中，从区块链管理模块点击创建区块链即可搭建所需的测试联盟链，如图2所示。

其中需要选择共识方式，共识节点用于共识计算，而记账节点则用于存储链上数据。平台支持自定义添加共识节点和记账节点数量，实验中为方便测试与快速搭建，各生成一个节点。

为了使节点间能够相互通信，需要设置节点的http端口以及grpc通信端口，两种端口号需设置得不同。

在配置好节点信息后，RayBaaS平台支持一键组网，可快速构建测试网络。至此区块链网络搭建成功，可进行智能合约的相关测试。

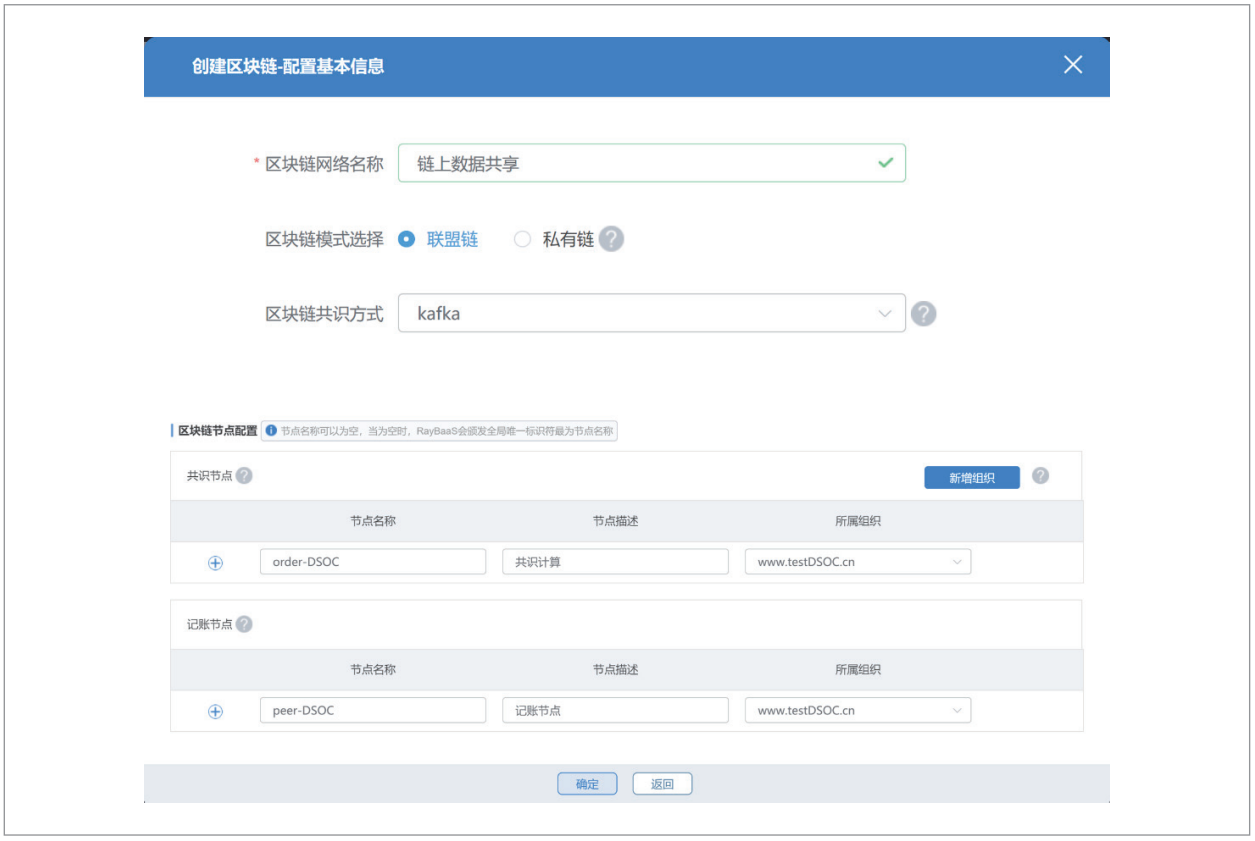


图 2 创建区块链

在区块链网络构建成功后，将访问控制模块以及数据处理模块的智能合约上传至RayBaaS平台即可完成合约的部署，如**图3**所示。RayBaaS平台内置了区块链网络的基础合约，因此只需部署实现系统关键逻辑部分的智能合约。

5.2 数据处理的同态性与范围型验证

原始用户体检报告数据见**表1**。通过

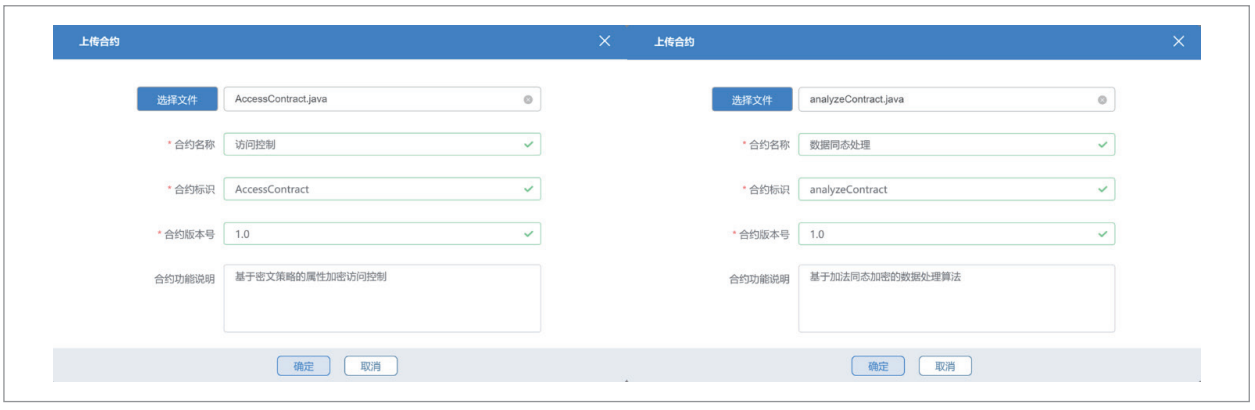


图 3 智能合约部署

访问控制与同态加密模块后，数据访问者调用链上的用户数据时是密文状态，通过区块链浏览器查看链上信息，如**图4**所示。而从链上将数据导出成表后，用户相关属性信息除病历号外皆通过哈希函数加密，而具体指标信息则经过同态加密处理。加密处理后的链上数据如**图5**所示。

数据访问者进行数据处理时，可通过最大正常参考值减去结果值来判断该份报告的用户指标是否正常，处理结果也不显示明文，而是以范围型验证的方式告知用户结果，这样即使数据访问者拿到数据结

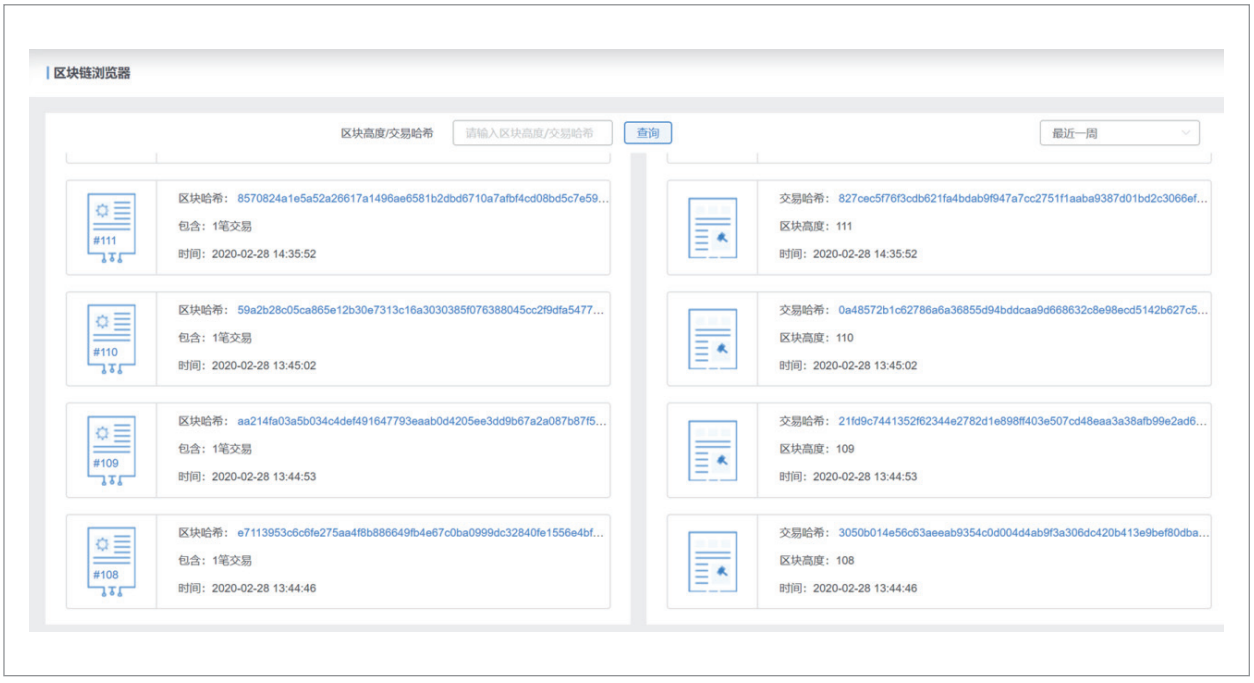


图 4　链上交易信息

患者：	病历号：	年龄：	性别：
1d7f476096e7bd9443513f22ab9af	1234567	8e296a67a37563370ded05f5a3bf3ec	a3d2de7675556553a5f08e4c88d2c228
检查项目	结果值	单位	最大正常参考值
红细胞比容（HCT）	1493122775118091449062147031407013012007365670290919609127535853530808740756789733384816183448761175331317209356596600864979176367082911407050928925222318250567790492554420387391824313591895549072667261170021771602900882827798668143410626034342383830462114468478291596584320774695891054249407954841636907227398	%	6487388667119870639234140229730529112488160220863004411882407464373365744609659011422257091375192119967784057533348922482277459738873412602305785990574592460719174798862578880978317019808366186130135376909984253791023359109716233331086880305552347698976988888370340096941709817799298388361079648219742050285534
血小板（PLT）	1774408570660221904994678007965795767721050067881560167424990514367708677104908752235124366706692071429296992121356931381597085957053799773615257901396064118813455356841244575417806092641775418226428948670477977284386031995946418289327747887803888883972304989774585198102320258453238900301795801004739707480 2	K/uL	1002359457303783956292070402500656564740716680413131988049260729106452311273395606502884523603332240000750422950552374092719689262096255569631796680856891022806421235076860677222982341608774147220344408583623912104909595380979493736859962514115265631324301227769093743105922058318676677632944040589582334445
球蛋白（GLOB）	4084471671661201728725393774396068918970797643917936143027263484525911264550093702355133867739797087582875707488223502742854001894525213606270467834772402906109633831418670544563423983594477705214692919915070789301077854495526674739757018693369915385177079575895977774255554310556061602105557412845523833795 0	g/L	1614789148289979550090198594374220950970966373272415141952574512740901890655213472544586644492040424188739337933069141295945171421234327621492432042998934357852949839266017362633012362424908807111528268400160819921189427174582499425356006791126434565879561957015222590185756565630445491079559237676060303874219

图 5　加密处理后的链上数据

果，也不能得知该指标数值的溢出程度，只知道该指标是否达标。数据结果是否为明文显示则可根据用户需求进行调整。数据处理结果如**图6**所示。

由**图6**可知，数据密文处理后的结果与明文处理后的结果一致（明文结果哈希化后得到的是密文的处理结果，这里直接以表格形式输出），证明了该体系架构的可行性，该体系架构可有效地保障用户隐私以及共享数据的安全。

5.3 改进性分析

在该架构中，目前笔者只实现了加法同态加密的相关分析操作以及一些简单的数据共享处理业务，还不能支持乘法同态加密的相关数据操作，同时在数据量过大情况下的系统并发性还未进行论证，这些将是下一阶段的研究目标。

6 结束语

本文提出的DSOC架构是基于区块链实现的链上数据共享体系。通过CP-ABE算法实现用户的访问控制，之后基于Paillier算法的加法同态加密实现了数据共享中的部分业务场景，通过RayBaaS平台搭建测试联盟链，得到的实验结果论证了数据的同态性以及范围型验证的可行性，说明该架构能够有效地保障数据的安全共享。

参考文献:

[1] 白硕. 浅论区块链的可运维性[J]. 大数据, 2018, 4(1): 85-89.
BAI S. Brief comments on the operability of blockchain[J]. Big Data Research, 2018, 4(1): 85-89.

[2] 马凯航, 高永明, 吴止锾, 等. 大数据时代数据管理技术研究综述[J]. 软件, 2015, 36(10): 46-49, 56.
MA K H, GAO Y M, WU Z H, et al. Data management technology of big data era[J]. Computer Engineering & Software, 2015, 36(10): 46-49, 56.

[3] 唐世渭. 数据管理技术的重要方向[J]. 软件世界, 2003(7): 88-89.
TANG S W. The important direction of

结果值	最大正常参考值	处理结果	处理结果（明文）（可不显示）	范围型验证结果
1493122775118091449062147031407013012007365670290919609127535835308087407567897333848161834487611753313172093565966008649791763670829114070509289252223182505677904925544203873918243135918955490726672611700217716029008828277986681434106260343423838304621144684782915965843207746958910542494079548416369072239 8	6487388667119870639234140229730529112488160220863004411882407464373365744609659011422257091375192119967784057533348922482277459738873412602305785990574592460719174798862578880978317019808366186130135376909984253791023359109716233331086880305552347698976968888370340096941709817799298383610796482197420502853 4	5221665455556481948144399664749796618480344171571355205035615431359753153103482385815779367222270741053675531663572882785326806199373031156652924365754219343011373773471481163194723234019443325912683353264490923453729690468525579283502932353725022498979782931534013978977993405511063496086170877934042613504508	11	result>0
1774408570660221904994678007965795767721050067881560167424990514367708677104908752235124366706629071429296992121356931381597085955057379773615257901396064118813455356841244575417806092641775418226428948670477977284386031995946418289327747887803888883972304989774585198102320258453238900301795801004739707480 2	1002359457303783956292070402500656564740716680413131988049260729106452311273395606502884523603332240000750422950552374092719689262096255569631796680856891022806421235076860677222982341608774147220344408583623912104909595380979493736859962514115265631324301227769093743105922058318676677632944040589582334445	3416328750401573561538877509365059089865957742394896713140482933877797149028995695638092641826040485048368828350075280174299407538943833477033660813661064707713046753533129704033295857136079573131561568373392551088497690841457913382523296855604958072928901698 78	293	result>0
4084471671661201728725393774396068918970797643917936143027263484525911264550093702355133867739797087582875707488223502742854001897452521360627046783477240290610963383141867054456342398359447770521469291991507078930107785449552667473975701869336991538517707957589597777425555431055606160210555741284552383379 50	1614789148289979550090198594374220950970966373272415141952574512740901890655213472544586644492040242188739337933069141295945171421234327621492432042998934357852949839266017362633012362424908807115282684001608199211894271745824994253560067911264345658795619570152225901857565656304454910795592376760603038742 19	4244262055455625563789003185115378044550639375159323094999417853346764943821587356833719782222269721621802295252694060713775686484986876327883030065188505493669797901297652863260643056387995771115255232203824231444001753576759079042174249605640890422549862848949732830715733541761595133762681545178383831286	-5	result<0

图 6 数据处理结果

data management technology[J]. Software and Information Service, 2003(7): 88–89.

[4] 黄如花, 陈闯. 美国政府数据开放共享的合作模式[J]. 图书情报工作, 2016, 60(19): 6–14.
HUANG R H, CHEN C. Study on the sharing cooperation mode of U.S. open government data[J]. Library and Information Service, 2016, 60(19): 6–14.

[5] SUN J J, YAN J Q, ZHANG K Z K. Blockchain-based sharing services: what blockchain technology can contribute to smart cities[J]. Financial Innovation, 2016(26).

[6] YUE X, WANG H J, JIN D W. et al. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control[J]. Journal of Medical Systems, 2016, 40(10): 218.

[7] ZIKRATOV I, KUZMIN A, AKIMENKO V, et al. Ensuring data integrity using blockchain technology[C]//The 20th Conference of Open Innovations Association FRUCT. [S.l.:s.n.], 2017: 534–539.

[8] SHAFAGH H, BURKHALTER L, HITHNAWI A, et al. Towards blockchain-based auditable storage and sharing of IoT data[C]//The 2017 on Cloud Computing Security Workshop. [S.l.:s.n.], 2017: 45–50.

[9] XIA Q, EMMANUEL S, ABLA S, et al. BBDS: blockchain-based data sharing for electronic medical records in cloud environments[J]. Information, 2017, 8(2): 44.

[10] XIA Q, SIFAH E B, ASAMOAH K O, et al. MeDShare: trust-less medical data sharing among cloud service providers via blockchain[J]. IEEE Access, 2017, 5: 14757–14767.

[11] ZHANG P, WHITE J, SCHMIDT D C, et al. FHIRChain: applying blockchain to securely and scalably share clinical data[J]. Computational and Structural Biotechnology Journal, 2018, 16: 267–278.

[12] ZHANG A, LIN X. Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain[J]. Journal of Medical Systems, 2018, 42(8).

[13] 李康, 孙毅, 张珺, 等. 零知识证明应用到区块链中的技术挑战[J]. 大数据, 2018, 4(1): 57–65.
LI K, SUN Y, ZHANG J, et al. Technical challenges in applying zero-knowledge proof to blockchain[J]. Big Data Research, 2018, 4(1): 57–65.

[14] 祝烈煌, 董慧, 沈蒙. 区块链交易数据隐私保护机制[J]. 大数据, 2018, 4(1): 46–56.
ZHU L H, DONG H, SHEN M. Privacy protection mechanism for blockchain transaction data[J]. Big Data Research, 2018, 4(1): 46–56.

[15] MUZAMMAL M, QU Q, NASRULIN B, et al. ChainSQL: a blockchain database application platform[J]. arXiv preprint, 2019, arXiv:1808.05199.

[16] 张超, 李强, 陈子豪, 等. Medical Chain: 联盟式医疗区块链系统[J]. 自动化学报, 2019, 45(8): 1495–1510.
ZHANG C, LI Q, CHEN Z H, et al. Medical chain: alliance medical blockchain system[J]. Acta Automatica Sinica, 2019, 45(8): 1495–1510.

[17] 闫树, 卿苏德, 魏凯. 区块链在数据流通中的应用[J]. 大数据, 2018, 4(1): 1–12.
YAN S, QING S D, WEI K. Application of blockchain in data circulation[J]. Big Data Research, 2018, 4(1): 1–12.

[18] WATERS B. Ciphertext-policy attribute-based encryption: an expressive, efficient, and provably secure realization[C]//The 14th International Conference on Practice and Theory in Public Key Cryptography. Heidelberg: Springer, 2011: 53–70.

[19] 杨攀, 桂小林, 姚婧, 等. 支持同态算术运算的数据加密方法算法研究[J]. 通信学报, 2015, 36(1): 171–182.
YANG P, GUI X L, YAO J, et al. Research on algorithms of data encryption scheme that supports homomorphic arithmetical

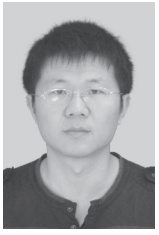
operations[J]. Journal on Communications, 2015, 36(1): 171-182.
[20] PAILLIER P. Public-key cryptosystems based on composite degree residuosity classes[J]. Lecture Notes in Computer Science, 1999, 547(1): 223-238.
[21] 中国信息通信研究院. “链”接未来: 可信区块链应用实践[M]. 北京: 人民邮电出版社, 2018: 29-37.
China Academy of Information and Communications Technology. Chain to the future: trusted blockchain application practice[M]. Beijing: Posts & Telecom Press, 2018: 29-37.

作者简介

刘彦松(1996-),男,电子科技大学计算机科学与工程学院硕士生,主要研究方向为数据安全与区块链。

夏琦(1979-),女,博士,电子科技大学计算机科学与工程学院教授,电子科技大学网络空间安全研究中心副主任/区块链研究所执行所长,四川省大数据共享与安全工程实验室执行主任,中国计算机学会(CCF)区块链专业委员会委员,美国宾夕法尼亚大学访问学者,主要研究方向为网络安全技术及其应用、大数据安全、区块链。

李柱(1995-),男,电子科技大学计算机科学与工程学院硕士生,主要研究方向为数据安全与区块链。



夏虎(1981-),男,博士,电子科技大学计算机科学与工程学院副研究员,主要研究方向为大数据挖掘与分析、数据安全与区块链。

张小松(1968-),男,博士,电子科技大学计算机科学与工程学院长江学者特聘教授,政府治理大数据国家工程实验室专家委员会副主任委员,中国电子学会区块链分会副主任委员,电子科技大学网络空间安全研究院院长,主要研究方向为网络信息技术安全和应用。

高建彬（1976– ），男，博士，电子科技大学资源与环境学院副教授，CCF区块链专业委员会委员，美国宾夕法尼亚大学访问学者，主要研究方向为数据分析与挖掘、图像处理、区块链、智能决策等。

收稿日期： 2020-02-16
基金项目： 国家自然科学基金资助项目（No. 61572115）；四川省科技厅国际合作资助项目（No. 2017HH0028, No. 2018HH0102, No. 2019YFH0014, No. 2020YFH0030）；四川省科技计划资助项目（No. 2017CC0071, No. 2020YFSY0061）
Foundation Items: The National Natural Science Foundation of China(No. 61572115), International Cooperation Project of Science and Technology Department of Sichuan Province (No. 2017HH0028, No. 2018HH0102, No. 2019YFH0014, No. 2020YFH0030), Science and Technology Program of Sichuan Province(No. 2017CC0071, No. 2020YFSY0061)

链上存证、链下传输的可信数据共享平台

张召[1]，田继鑫[2]，金澈清[1]
1. 华东师范大学数据科学与工程学院，上海 200062；2. MCT Technology，上海 200023

摘要
区块链系统可以为分享数据的互不信任的多方之间提供可信的基础设施。但是，将原始分享数据直接上链的方式并不适合大规模的数据分享场景。因此，提出了一种数据共享请求和应答记录上链存证、原始数据链下安全传输的数据共享平台架构，该架构在一定程度上可以缓解系统负载过重以及隐私保护方面的问题。最后总结了随着参与节点的增多，以及每秒需要处理的数据共享请求和应答的增多，已有的区块链技术被应用到数据分享和确权领域时，在分布式存储、共识协议、智能合约执行以及轻客户端查询方面面临的挑战以及改进的方向，以期为已有区块链系统应用于数据共享领域指明需要进一步突破的技术瓶颈。

关键词
数据共享；数据确权；数据追溯；区块链

中图分类号：TP315　　文献标识码：A　　doi: 10.11959/j.issn.2096-0271.2020047

On-chain witness and off-chain transmission trustworthy data sharing platform

ZHANG Zhao[1], TIAN Jixin[2], JIN Cheqing[1]
1. School of Data Science and Engineering, East China Normal University, Shanghai 200062, China
2. MCT Technology, Shanghai 200023, China

Abstract
Blockchain system can build a trusted infrastructure for sharing data between multiple untrusted parties. However, directly uploading original shared data to blockchain is not suitable for large-scale data sharing scenarios. A data sharing architecture where data sharing request and response records are deposited on-chain and original data is transmitted securely off-chain was proposed. The architecture can alleviate the problems of system overload and privacy protection to a certain extent. Finally, with the increase of participating nodes and the data sharing requests and responses to be handled per second, the limitations in distributed storage, consensus protocol, smart contract execution, and query from light clients, directions for further research were proposed, in order to specify the technical bottlenecks that need to be further broken for the existing blockchain system applied to the field of data sharing.

Key words
data sharing, data property rights, data tracking, blockchain