计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190300087

基于联盟区块链的医疗健康数据安全模型

冯 涛 焦 滢 方君丽 田 野
兰州理工大学计算机与通信学院 兰州 730000

摘 要 传统医疗信息系统中总是存在医疗健康数据安全存储难和共享难的问题,不同身份的人员在访问和共享医疗健康数据时都受到比较严格的限制,且验证身份和数据的真实性需要大量的资源和时间。针对传统医疗信息系统中存在的存储集中、共享安全性低和达成一致困难等问题,提出了一个基于联盟区块链的医疗健康数据安全模型。该模型根据目前的医疗资源分布情况将医疗机构划分等级,使用股份授权证明机制(Delegate Proot of Stake,DPOS)和实用拜占庭机制(Practical Byzantine Fautt Tolerante,PBFT)结合的混合共识机制保证在没有中心节点的情况下联盟中医疗机构可以快速达成一致,共享医疗健康数据;并根据区块链去中心、安全可信和防止篡改等特点,将数据记录及其他重要信息存储在区块链上,而将完整医疗数据加密存储在分布式数据库(Distributed Database,DDB)中,在安全存储用户医疗健康数据的同时,提高了数据在各医疗机构间的共享效率。安全性分析表明,该模型在容错范围内可以保护医疗健康数据,防止其被篡改和共谋;一致性分析表明,该模型有99%的概率保证联盟中医疗机构达成共识并共享医疗数据。

关键词:医疗健康数据;区块链;混合共识机制;代理重加密;安全模型
中图法分类号 TP309.2

Medical Health Data Security Model Based on Alliance Blockchain

FENG Tao,JIAO Ying,FANG Jun-li and TIAN Ye
School of Computer and Communication,Lanzhou University of Technology,Lanzhou 730000,China

Abstract In traditional medical information system,medical health data security storage and sharing have been becoming a challenging task. There are many restrictions in process of health data accessing and sharing for different people of identity,which spends a lot of resources and time on identity verification and data authentication. Aiming at these problems such as storage of the high concentration,unreliable data sharing security and the difficulty of reaching agreement,this paper proposed an alliance blockchain-based medical health data security model. According to the distribution of medical resources in reality,the medical institutions are ranked in the security model,and then combine DPOS with PBFT to ensure that the medical institutions can reach an agreement rapidly without a central node and share medical data in alliance. The security model has the advantages of decentralization,high security and tamper resistance,so it can store data records and other important information on the blockchain,but the original medical data is stored in Distributed database. The use's medical health data is stored securely,meanwhile the sharing efficiency among the medical institutions is improved. Security analysis shows that the proposed model can protect medical health data within the scope of fault tolerance,prevent the data from tampering and the collusion problem. The proposed model has a 99% probability to ensure that the medical institutions can reach a consensus and share medical data in alliance by the consistency analysis.

Keywords Medical health data,Blockchain,Hybrid consensus mechanism,Proxy re-encryption,Security model

1 引言

1.1 研究背景

在医疗健康数据系统中,医疗健康数据精确记录着每个人的用药情况、过敏药物、化验结果以及各项生命体征,这些都是十分宝贵的数据资产。如何保证数据在安全存储的情况下被各医疗机构高效共享一直是一个难点。传统的方法利用云端来存储和共享医疗健康数据,医疗机构可以通过云端上传或下载患者的电子病历,这在一定程度上提高了存储、检索和共享的效率,但同时存在着数据被篡改、传输不安全等风险。例如,拥有病人医疗健康数据的特定医疗机构可以控制其他医疗机构是否能访问数据;病人的医疗健康数据由医疗机构托管在半可信的第三方服务器上,如果攻击者攻击第三方服务器或者医疗机构受到高价值敏感信息的诱惑,将导致服务器上所有用户的医疗信息被完全泄露。

作为一个分布式可验证公共账本,区块链可以为医疗机

到稿日期:2019-03-20 返修日期:2019-06-11 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61462060)
This work was supported by the National Natural Science Foundation of China (61462060).
通信作者:冯涛(fengt@lut.cn)

构和患者提供安全的存储和数据共享服务。

1.2 本文贡献

本文详细介绍并设计了一个基于联盟区块链的医疗健康数据安全模型，帮助解决医疗健康数据易被垄断和篡改、各机构间共享难以及第三方不可信等问题，以真正达到分布式、去中心化、可追溯、不可篡改[1]的医疗健康数据安全存储和共享的目的。本文模型的主要具有如下特点。

1)医院一级节点联盟群(Hospital Level 1 Node Alliance Group，HL1)和医院二级节点联盟群(Hospital Level 2 Node Alliance Group，HL2)：根据目前的医疗资源分布情况以及医疗机构做出贡献的大小和设备的先进程度，可以将医疗机构划分等级。贡献大且设备先进的高等级医疗机构加入一级节点，这些高等级医疗机构可以充分利用共享的医疗健康数据以提高工作效率；贡献小且设备陈旧的医疗机构加入二级节点，发挥验证和监督的作用。如果一级节点在一段时间后出现了错误或做出的贡献较低，整个联盟会通过“升降级”制度将贡献较低的医院节点删除，禁止其共享数据，这种做法间接提高了医疗机构的竞争力。本文模型中采用股份授权证明机制与实用拜占庭机制混合的共识机制来达成节点之间的共识。

2)医疗健康记录存储结构：利用 Merkle 树的结构，将每一次记录、贡献节点签名、地址加密哈希后放在 Merkle 树的叶子节点，逐级向上哈希后得到 Merkle 根，将这个 Merkle 根作为一个区块放入区块链中。这种存储方式达到了不可篡改以及不可抵赖的目的。

3)分布式数据库：由于区块链的存储容量有限，因此将所有医疗数据存储在区块链上是不现实的。本文模型利用分布式数据库存储完整的医疗健康数据密文，而将数据记录放入区块链中。这不仅解决了数据集中存储在某个医疗机构云服务器上容易被攻击的问题，同时也解决了区块链容量受限的问题，减轻了区块链被高频率访问的压力。

本文第 2 节首先介绍医疗健康数据中区块链的研究现状和所提模型中的相关技术；第 3 节详细介绍医疗健康数据安全模型的组成部分；第 4 节和第 5 节通过安全性证明及一致性分析说明本文模型可以达到安全存储和共享医疗健康数据的目的；第 6 节通过与现有医疗健康数据安全模型进行对比来评估所提模型的性能；最后对本文的工作进行总结并对下一步研究工作进行展望。

2 相关工作

2.1 研究现状

区块链技术自 2008 年出现以来，一直在金融领域受到持续关注并快速发展。医疗健康领域是仅次于金融领域的区块链第二大应用场景，其中医疗健康数据的安全存储和共享成为了研究的重点。Azaria[2]等提出了 MedRec 方案，该方案利用区块链的特性以及 POS 共识机制进行身份管理和验证，确保了共享医疗数据的机密性。Zhao[3]等利用身体传感器网络(Body Sensor Networks，BSN)为健康数据区块链的密钥设计了一个轻量级备份且有效的恢复模型，该模型可以有效保护区块链上健康数据的隐私。Shrier 和 Chang 等提出采用麻省理工大学的 OPAL/Enigma 加密平台，配合使用区块链技术创建了一个用于存储和分析医疗数据的安全环境[4]。还有一些研究对利用区块链存储电子病历的模型[5-6]进行了预测和评估。本文提出并设计了一个基于联盟区块链的医疗健康数据安全模型。

2.2 区块链和共识机制

区块链的基本原理如图 1 所示。区块链将记录或交易成批地存放在带有时间标记的数据块中。每个数据块使用其自身的哈希值进行标识，并与其前面产生数据块的哈希值相连，从而形成一个完整的链条。正是这种独特的数据结构，使得区块链具有高冗余、无法篡改、低成本和可以进行多签名复杂权限管理的特性[7]。

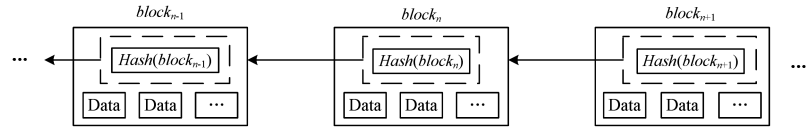

图 1 区块链的基本原理图

Fig. 1 Basic schematic diagram of Blockchain

股份授权证明(DPOS)是在股权证明(Proot of Stake，POS)基础上将记账人角色专业化的共识机制[8]。DPOS 先将权益作为选票选出记录者，然后记录者之间轮流进行验证和记账，但必须要保证有 90%的记账人在线。这种共识机制大幅减少了参与验证和记账的节点数量，可以达到秒级的一致性验证，避免了 POW 的高成本以及 POS 的高度中心化[8]。根据 DPOS 的工作原理，通过选举选择记录者，再在记录者之间进行验证和记账的方式对于快速达成共识后共享医疗健康数据是十分适用的；但 DPOS 仍过于依赖代币，很多实际医疗应用中都不需要代币的存在。

在区块链的共识机制中，POW 延时太大，POS 和 DPOS 过于集中化，所以这 3 种共识机制都是在公有链或者部分私有链中被广泛应用。但是，面对特定的行业，例如医疗健康领域，大部分还是使用以 PBFT 为代表的共识机制。Castro 等在 1999 年提出了 PBFT，它是一种基于状态机副本复制的算法，可以在保证安全性和活性的前提下提供失效节点不超过 $\left\lfloor\frac{n-1}{3}\right\rfloor$ 的容错保证[9]。也就是说，PBFT 在 $N\geqslant 3F+1$ 情况下的一致性问题是可以解决的。其中，N 为总节点数，F 为恶意节点数量。

2.3 哈希算法和默克尔树结构

哈希算法是安全领域中重要的密码学技术，简单的说就是将任意长度的输入通过算法转变成固定长度的输出，输出的值称为哈希值[10]。哈希算法的一个重要特点就是单向性，这种单向性对于医疗健康区块链是十分重要的。Merkle 树是基于哈希算法的数据结构，其中，数据会被分成很多小的数据块放到最底层，首先对单独的小数据块进行哈希处理，然后将两个子节点的哈希值合并之后再次哈希，将其表示为上一

层的父节点，重复这种方法，直到只有一个根哈希[11-12]。区块链中利用 Merkle 树结构验证数据是否被损坏或篡改，是十分有用的。

3　医疗健康数据安全模型

3.1　基本结构

医疗健康数据安全模型包括很多重要部分，如图 2 所示。

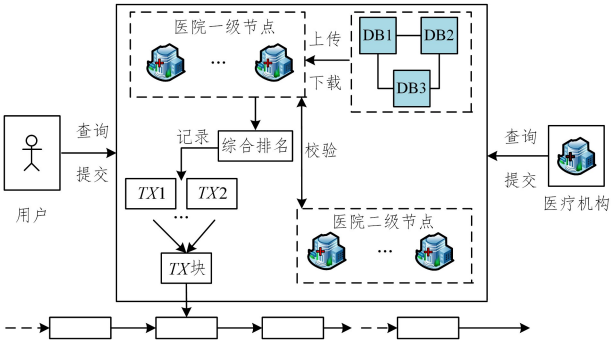

图 2　医疗健康数据安全模型

Fig. 2　Security model about medical health data

根据目前医疗行业发展的情况，本文模型可以分为 3 种不同的服务：存储服务、查询服务、上传服务。

1）存储服务：存储所有医疗健康数据（医院可以对外提供接口进行存储服务）。

2）查询服务：医院一级节点联盟群向用户或者医疗机构提供查询（联盟群内的医疗机构提供查询接口，用户可以进行对医疗机构授权等操作）。

3）上传服务：用户在医院访问结束时请求将数据加密并上传，医院可以通过存储服务向医院一级节点联盟提出存储请求，用户去联盟内的其他医院就诊时可以通过查询服务获得自己的历史数据。

该模型将医疗机构联盟分成两级，即医院一级节点联盟群（HL1）和医院二级节点联盟群（HL2），并配合使用混合共识机制。为了保证数据是可信的、未被篡改过的，该模型将数据摘要哈希值、数据上传者公钥、上传时间统称为一个 TX，计算 TX 的哈希值并将其存放在 Merkle 树的叶子节点，叶子节点两两相加并哈希后放入根节点中。这样不仅可以保证数据的真实性，同时提高了二级节点对数据真实性的验证效率。

如图 3 所示，叶子节点由多个 TX 块组成，之后层层向下计算哈希值直到得到 Merkle 树的树根，每 10 s 进行一次。

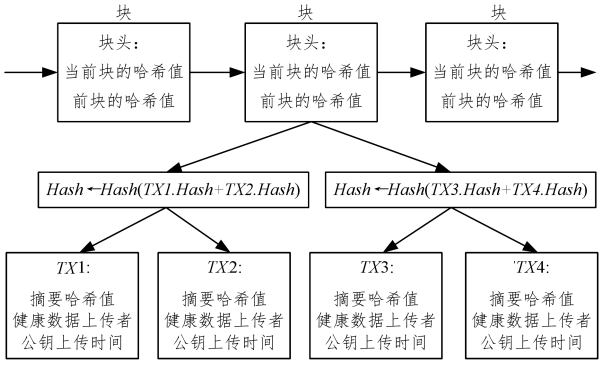

图 3　医疗健康数据的区块链结构

Fig. 3　Blockchain structure of medical health data

由于本文模型是以联盟链为基础的，因此混合共识机制可以稳定在每分钟内生成一个区块，所以模型每分钟冻结一次数据。由医院一级节点联盟群中的当值节点将已生成的 Merkle 根提交到区块链中，这样可以真正达到不可篡改的目的。本文模型首先让成员利用混合共识机制快速达成一致后共享数据，提高一致性；再利用代理重加密的方式将原始健康数据存储在分布式数据库中，提高安全性。

3.2　医疗健康数据安全模型的混合共识机制

针对 DPOS 共识机制的缺陷，本文将 DPOS 与 PBFT 结合，创建了一种新的适合医疗健康数据安全模型的共识机制。根据最新的全国医院综合排名，将重点医疗机构划分到 HL1 中，它们需要先通过投票选出可以代表其权益的当值节点，由当值节点将提交上来的原始医疗健康数据加密存储在 DDB 中，并在用户需要时从 DDB 中下载这些数据，同时当值节点使用自己的私钥对数据记录签名，联合数据摘要、上传者公钥和上传时间一并放入 TX 块中。将排名靠后的医疗机构划分到 HL2 中轮流验证当值节点签署的区块的正确性及合法性。在整个网络中，本文模型采用“升降级”方式，HL1 中的医院节点每隔一段时间根据贡献进行一次排名，排在最后的几位会被移出一级节点联盟群，由二级节点联盟中排名靠前的医院节点进行补位。如果一级节点联盟群中医疗机构的服务器因为外界原因出现故障，并不会使整个共识机制失效，因为混合共识机制会在容错范围内使模型可以正常工作。“升降级”方式的加入，充分调动了医疗机构为病人提供更好服务的积极性，未来也将成为评估医疗机构的重要标准。

3.2.1　混合共识机制的威胁模型

在本文提出的共识机制中，只有已经存在的区块链是可信的。HL1 中的当值节点可能为恶意节点，其故意签署错误区块或延迟发起共识，继而导致发布错误共识；HL1 中的其他节点可能伪装成当值节点故意签署错误区块；HL1 中的其他节点共谋，故意延迟发送，或发送的确认共识消息不一致，导致共识失败。

3.2.2　混合共识机制的符号表示

本文模型将 HL1 和 HL2 分别用集合 N_1^* 和集合 N_2^* 表示，有：

$$N_1^* \geqslant 3f+1 \tag{1}$$

其中，f 是 HL1 中可以容忍的最大错误节点数。为了方便计算，设置 $N_1^*=3f+1$。

HL2 是 HL1 的补位集合。一段时间后，根据贡献排名剔除 HL1 中排名在末尾的 n 个授权代表放入 HL2 的末尾，并从 HL2 中选取前 n 个代表进入 HL1 中。有如下关系：$N_2^*=\lfloor 2f \rfloor$，$n=\left|\frac{1}{2}f\right|$。

对 HL1 和 HL2 中各个医疗机构服务器的信息进行调整，使得它们首先在配置上达成共识。然后将这些服务器信息进行统一编号，令编号为 c（编号从 1 开始）并依次增加；令 HL1 中的当值节点为 R，当值节点的选取需要兼顾此时的区块高度和服务器编号，满足公式：$R=(H+c) \bmod N_1^*$（其中 H 为当前区块的高度）。

令 HL1 中的其他节点为 O，用编号 $\{0,1,\cdots,N_1^*-1\}$ 表

示，发生一次“升降级”后开始重新随机编号；令 HL2 中的检查节点为 C。

本文模型达成共识需要分为两步进行：发起共识(Initiating Consensus)和确认共识(Confirm Consensus)。

3.2.3 混合共识机制的具体过程

当有用户或者医院提出上传数据记录的请求并发送医疗数据时，由 HL1 中的当值节点发起共识。为了使本文模型更有效，定义 Δ 为当值节点开始一次共识的时间间隔。整个模型中，当值节点收到医疗数据后，先将所收到数据的信息附上发送者签名，然后将其以 Flooding 方式在全网广播，最后将医疗数据加密存储在 DDB 中。此外，当值节点将数据记录放在自己的内存中，并开始构建区块。

在区块链网络中，所有的节点必须使它们的原始区块高度保持一致才能达成共识。为了达到这个要求，各个医院节点必须具备独立监听网络中医疗数据的能力，并将摘要信息写入自己的内存中。当当值节点发起一次共识时，需要保证 HL1 中所有诚实节点的医疗数据信息、医疗机构服务器信息、区块高度 H 以及上一区块的哈希索引、版本号是一致的。

基于以上逻辑，得出本文模型的混合共识机制的具体过程。

1)令 p 为大素数，G_1 和 G_2 分别是阶数为 p 的加法群和乘法群，并从中选择生成元 P 和 Q，G_T 为 p 阶循环群，有双线性映射 $e:G_1\times G_2\rightarrow G_T$。定义 SHA 函数：

$$\begin{cases}H_0:\{0,1\}^*\rightarrow N_p^*\\H_1:\{0,1\}^*\rightarrow G_1\\H_2:\{0,1\}^*\times G_1\rightarrow G_2\end{cases}\tag{2}$$

在整个当前模型下，计算：$P_c=c\cdot P$。

现得到系统参数：$Params=\{G_1,G_2,e,p,P,Q,P_c,H_0,H_1,H_2\}$。

2)用户或者医院 u_i 选择一个随机数 $x_i\in N_p^*$ 作为其随机秘密值，将 GID_i 作为其全局标识符，计算：

$$PK_i=x_iP\tag{3}$$

$$Q_i=H_2(GID_i\parallel P_i)\tag{4}$$

$$W_i=cQ_i\tag{5}$$

$$SK_i=(W_i,x_i)\tag{6}$$

其中，PK_i 和 SK_i 为用户的公私钥。

3)用户上传医疗数据并提出记录请求。如果该请求是 HL1 的其他节点 O 收到的，则将数据继续泛洪转发。如果该请求是 HL1 中的当值节点 R 收到的，则 R 首先验证数据是否合法，若合法，则用户利用 SK_i 将医疗数据加密后存入 DDB 中，随后当值节点 R 将数据摘要的哈希值、上传者公钥以及上传时间写入其内存中并记录到 TX 块上；若不合法，直接将其丢掉。

4)每隔 10 s 检查 TX 块的数量，达到 10 个时就根据 Merkle 树的结构计算最终的 Merkle 根，即得到一个新的区块 B。

5)当值节点 R 经过 Δ 时间后，向除了自身以外的其他所有节点发送发起共识的消息 M_R。消息格式为〈〈Initiating Consensus，c，H，TX，σ_R〉，$BLOCK$〉，其中 $BLOCK$ 是传播的区块信息，$TX=\langle H(Digest(data)),PK_i,Time\rangle$，$\sigma_R$ 是 R 对 M_R 的签名。签名算法的执行过程如下。

①当值节点 R 选择随机数 $r_i\in N_p^*$，计算：

$$R_i=r_iP\tag{7}$$

$$h_i=H_0(GID_R\parallel M_R\parallel PK_i\parallel R_i)\tag{8}$$

$$T=H_1(P_c)\tag{9}$$

②计算：

$$X_i=W_i+h_ir_iT+x_iQ\tag{10}$$

6)输出当值节点 R 对 M_R 的签名 $\sigma_R=(X_i,R_i)$。将消息分发给其他节点之前，检查节点 C 需要验证当值节点 R 对 M_R 签名的有效性，计算过程如下：

$$h_i=H_0(GID_R\parallel M_R\parallel PK_i\parallel R_i)\tag{11}$$

$$Q_i=H_2(GID_R\parallel PK_i)\tag{12}$$

$$T=H_1(P_c)\tag{13}$$

然后验证式(14)是否成立：

$$e(X_i,P)=e(Q_i,P_c)e(T,h_iR_i)e(Q,PK_i)\tag{14}$$

7)如果 M_R 被验证有效，HL1 中的每个其他节点 O_i($i\in\{0,1,\cdots,N_1^*-1\}$)在收到当值节点送过来的消息后，开始给 R 发送确认共识的消息 M_{O_i}，消息格式为〈Confirm Consesus，c，H，σ_{o_i}，$BLOCK_i$〉，其中 $BLOCK_i$ 是其他节点 O_i 转发的由当值节点 R 生成的区块信息，σ_{o_i} 为其他节点 O_i 对 M_{o_i} 的签名。O_i 的签名过程如下。

①选择随机数 $b_i\in N_p^*$，计算：

$$P_{o_i}=b_iP\tag{15}$$

$$Q_{o_i}=H_2(GID_{o_i}\parallel P_{O_i})\tag{16}$$

$$W_{o_i}=cQ_{o_i}\tag{17}$$

$$O_i=b_iP\tag{18}$$

$$h_{o_i}=H_0(GID_{o_i}\parallel M_{o_i}\parallel PK_i\parallel Q_i)\tag{19}$$

②计算：

$$X_{o_i}=W_{o_i}+h_{o_i}b_iT+x_iQ\tag{20}$$

最后输出 O_i 对 M_{o_i} 的签名：

$\{(M_{O_1},\sigma_{O_1}),\cdots,(M_{O_n},\sigma_{O_n}(X_{O_n},O_n))\}$。

8)当值节点 R 收到来自其他节点的 $2f$ 个确认共识的消息 M_{o_i} 后，如果逐条发送给检查节点 C 验证消息，那么达成共识的时间会大大增加。对此，本文将收到消息中的签名聚合为一条签名，从而减少验证时间，加快达成共识的效率，过程如下。

①R 计算 X_O 和 O，使得：

$$X_O=\sum_{i=0}^{N_1^*-1}X_{o_i}\tag{21}$$

$$O=\sum_{i=1}^{N_1^*-1}h_{o_i}O_i\tag{22}$$

②得到聚合签名 $\sigma_O=(X_O,O)$，R 收到的确认共识消息为 M_O。

9)检查节点 C 验证 M_O 的有效性，执行以下算法。

①输入系统参数：$Params=\{G_1,G_2,e,p,P,Q,P_c,H_0,H_1,H_2\}$，其他节点 O_i 的身份列表 $GID_O=\{GID_{O_1},\cdots,GID_{O_{N_1^*-1}}\}$，其他节点 O_i 的公钥列表 $P_{o_i}=\{P_{O_0},\cdots,P_{O_{N_1^*-1}}\}$，确认共识的消息列表 $M_O=\{M_{O_0},\cdots,M_{O_{N_1^*-1}}\}$，签名列表 $\sigma_O=\{\sigma_{O_0},\cdots,\sigma_{O_{N_1^*-1}}\}$。

②计算：

$$Q_{o_i}=H_2(GID_{o_i}\parallel P_{o_i})\tag{23}$$

$$T=H_1(P_c) \tag{24}$$

③验证式(25)是否成立：

$$e(X_O,O)=e(\sum_{i=0}^{N_1^*-1}Q_{o_i},P_c)e(T,O_i)e(Q,\sum_{i=0}^{N_1^*-1}P_{o_i}) \tag{25}$$

10)如果验证通过，则 R 认定共识达成，可以将区块 B 锚定到区块链上。该轮共识任务完成后，R 将存在自己内存中的 TX 删除掉，同时服务器标号复位为 0，开始新一轮共识。

3.3　医疗健康数据的安全存储和访问

当值节点在检查医疗健康数据无误后，允许用户或医院利用其公钥将医疗数据加密存储在 DDB 中，本文模型利用代理重加密(Proxy Re-encryption)[13]来完成安全存储和访问。当医生提出查询病人数据的服务请求时，HL1 中的当值节点 R 对医生需要查询的那部分数据在 DDB 中做正常的加密操作，并在医生和病人之间生成与他们自身相对应的代理重加密密钥。HL1 中的其他节点 O 和 HL2 中的检查节点 C 都可以申请重加密权利，HL1 可以从中选取一个节点并将重加密密钥发送给它，该节点根据密文和重加密密钥完成重加密操作，然后将数据记录到区块链中，而索引标识则是提出请求的医生的公钥。医生可以利用自己的私钥，通过区块链到 DDB 中获取病人的医疗数据。本文模型采用文献[14]中的基于网格的代理重加密方案。

1)初始化。在输入安全参数 n 时，设置参数 $q=poly(n)$ 与 $m=O(n\lg n)$，选择两个随机矩阵 $\boldsymbol{A}\in Z_q^{n\times n}$ 与 $\boldsymbol{X}\in Z_q^{nk\times n}$，其中 $k=\lceil \lg q \rceil$；公共参数(PP)由矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{X}$ 组成。

2)生成公私钥对。选择 3 个噪声矩阵 $\boldsymbol{R},\boldsymbol{S}\in\psi_s^{n\times l}$ 和 $\boldsymbol{E}\in\psi_s^{nk\times l}$，其中 l 是信息长度，s 满足 $s=\partial q$，$0<\partial<1$，计算 $\boldsymbol{P}_1=\boldsymbol{R}-\boldsymbol{AS}$ 和 $\boldsymbol{P}_2=-\boldsymbol{XS}+\boldsymbol{E}$，得到私钥为 $\boldsymbol{S}$，公钥为 $\boldsymbol{P}=(\boldsymbol{P}_1,\boldsymbol{P}_2)\in(Z_q^{n\times l},Z_q^{nk\times l})$。

3)加密算法。选择 3 个噪声矢量 $\boldsymbol{e}_1,\boldsymbol{e}_2\in\psi_s^{1\times n}$ 和 $\boldsymbol{e}_3\in\psi_s^{1\times l}$，其中 ψ_s 是高斯分布；并计算：$c_1=\boldsymbol{e}_1\boldsymbol{A}+\boldsymbol{e}_2\in Z_q^{1\times n}$，$c_2=\boldsymbol{e}_1\boldsymbol{P}_1+\boldsymbol{e}_3+m\lfloor q/2\rfloor$。

4)m 为需要加密的信息，且 $m\in\{0,1\}^l$；输出密文 $C=(c_1,c_2)\in Z_q^{1\times(n+l)}$。

5)重加密密钥生成算法。选择两个噪声矢量 $\boldsymbol{e}_4\in\psi_s^{nk\times nk}$ 和 $\boldsymbol{e}_5\in\psi_s^{nk\times l}$，其中 ψ_s 是高斯分布；并计算重加密密钥：

$$rk_{A\to B}=\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{e}_4\boldsymbol{X} & \boldsymbol{e}_4\boldsymbol{P}_2+\boldsymbol{e}_5+Power2(S_A)\\ 0_{l\times n} & \boldsymbol{I}_{l\times l}\end{bmatrix} \tag{26}$$

$$Power2(S_A)=\begin{bmatrix}S_1 & \cdots & \cdots & \cdots & S_l\\ 2S_1 & \cdots & \cdots & \cdots & 2S_l\\ \vdots & \cdots & \cdots & \cdots & \vdots\\ 2^{k-1} & S_1 & \cdots & 2^{k-1} & S_l\end{bmatrix}\in Z_k^{nk\times l} \tag{27}$$

6)重加密算法。利用重加密密钥 $rk_{A\to B}$ 将密文 C_A 转换为密文 $C_B=(c_{1B},c_{2B})=[Bits(c_1)\mid c_2]\cdot rk_{A\to B}\in z_q^{1\times(n+l)}$。其中 $Bits(c_1)=[b_{1,1}\cdots b_{n,1}\mid b_{1,2}\cdots b_{n,2}\mid\cdots\mid b_{1,k}\cdots b_{n,k}]\in\{0,1\}^{1\times nk}$。

7)解密算法。计算：

$$m=[c_1\,c_2]\begin{bmatrix}\boldsymbol{S}_B\\ \boldsymbol{I}_{l\times l}\end{bmatrix} \tag{28}$$

其中，$m=(m_1,\cdots,m_l)$。如果 $m_i\leqslant\left\lfloor\frac{q}{4}\right\rfloor \bmod q$，则 $m_i=0$；否则 $m_i=1$。

4　安全性分析

4.1　区块链安全性分析

本文模型无需全局可信的第三方实体。传统的医疗健康数据模型中均存在垄断、脆弱和隐私问题，本文利用医疗健康联盟链的方式来保证数据的安全存储和安全共享，不依赖于全局可信的第三方实体，节点都以对等的方式进行通信。值节点加密存储原始数据，将数据记录放置于链上，从而避免了传统集中化的存储方式导致的中心节点被恶意攻击的问题，同时又提高了共享效率。本文安全模型具有良好的可扩展性和可靠性。

4.2　混合共识机制的安全性分析

1)正确性

定理 1　HL2 中检查节点 C 对 M_R 签名的验证是正确的。

证明：检查节点 C 对 M_R 的签名 $\sigma_R=(X_i,R_i)$ 进行验证的过程如下：

$$\begin{aligned}e(X_i,P)&=e(W_i+h_ir_iT+x_iQ,P)\\&=e(W_i,P)e(T,h_ir_iP)e(x_iQ,P)\\&=e(Q_i,P_c)e(T,h_iR_i)e(Q,PK_i)\end{aligned} \tag{29}$$

定理 2　HL2 中检查节点 C 对 M_O 聚合签名的验证是正确的。

证明：检查节点 C 对 M_O 的聚合签名 $\sigma_O=(X_O,O)$ 进行验证的过程如下：

$$\begin{aligned}e(X_O,O)&=e(\sum_{i=0}^{N_1^*-1}X_{o_i},O)\\&=e(\sum_{i=0}^{N_1^*-1}W_i+h_{o_i}b_iT+x_iQ,O)\\&=e(\sum_{i=0}^{N_1^*-1}W_i,O)e(\sum_{i=0}^{N_1^*-1}h_{o_i}b_iT,O)e(\sum_{i=0}^{N_1^*-1}x_iQ,O)\\&=e(\sum_{i=0}^{N_1^*-1}Q_{o_i},P_c)\prod_{i=0}^{N_1^*-1}e(T,h_{o_i}O_i)\prod_{i=0}^{N_1^*-1}e(Q,P_{o_i})\\&=e(\sum_{i=0}^{N_1^*-1}Q_{o_i},P_c)e(T,O_i)e(Q,\sum_{i=0}^{N_1^*-1}P_{o_i})\end{aligned} \tag{30}$$

2)抗共谋

定理 3　混合共识机制可以抵抗攻击者与 R 共谋，从而解决 R 不发起共识或未按时发起共识消息，HL1 中的其他节点可以撤销当前的当值节点并发起服务器信息需要更新的消息。

证明：令新的服务器信息编号为 $c_{\text{new}}=c_{\text{old}}+1$，发送更新消息的其他节点为 O_i'，发送的消息格式为〈Change Confirm，c_{old}，c_{new}，$\sigma_{O_i'}$，H〉。除 O_i' 以外的其他节点收到更新消息后，会拒绝 R 发起的共识，但是依旧会监听医疗健康数据记录。这些节点验证更新消息中的 c_{old} 和 H，查看是否与自己目前的服务器信息一致，并在此基础上加 1，变为 $c_{\text{new}}=c_{\text{old}}+1$；接着将自己的服务器消息已变更完成的消息广播给其他节点，格式为〈Change Confirm，c_{old}，c_{new}，H，i〉，其中 i 为已完成变更的其他节点编号。HL1 中除错误节点外的所有节点收到 $2f$ 个变更完成的消息后，服务器信息更新完成，编号从 c_{new} 开始。

(3)抗分叉攻击

定理 4 在本文模型的容错范围内,达成共识的网络无法被分叉。

证明:在混合共识机制达成一致的网络中,共识节点分为4个部分,分别为 R,O,F,C,即:$N_1^* = R \cup O \cup F$,$N_2^* = C$,且 $R \cap O = \varnothing$,$F \subseteq O$,$R \cap F = \varnothing$。R 和 O 中的部分节点是诚实节点;F 由恶意节点组成并且可以与 R 和 O 中的节点通信。

F 想使网络发生分叉,需要在 R,O,C 达成共识,并将新区块加入链之后,在未通知 O 少部分节点的情况下与 R 和 C 达成第二次共识,而且还能取消第一次达成的共识。满足以上条件的公式为:

$$\begin{cases} R + |F| \geqslant N_1^* - f \\ |O| + |F| \geqslant N_1^* - f \end{cases} \tag{31}$$

当恶意节点是网络可以容忍的最大节点数,即 $|F| = f$ 时,式(31)变为:

$$\begin{cases} R \geqslant N_1^* - 2f \\ |O| \geqslant N_1^* - 2f \end{cases} \tag{32}$$

则:

$$R + |O| \geqslant 2N_1^* - 4f \Leftrightarrow N_1^* \leqslant 3f \left(\text{已知 } f = \frac{N_1^* - 1}{3}\right) \tag{33}$$

得到:

$$N_1^* \leqslant N_1^* - 1 \tag{34}$$

这与事实矛盾,因此模型在容错范围内无法被分叉。

5 一致性分析

定理 5 基于蒙特卡洛一致性协议[15-17],本文提出的混合共识机制在容错范围内可以达到一致。

证明:假设有 $f \leqslant \frac{1}{3} N_1^*$ 个恶意节点可以共谋向当值节点 R 发送错误消息,出现恶意节点消息反对诚实节点消息,使得 R 接收恶意节点消息,从而导致共识失败。那么,正常情况下,节点同意达成共识的概率为:

$$a = \sum_{k=\frac{N_1^*}{3}}^{N_1^* - f} C_k^{N_1^* - f} p'^k (1 - p')^{N_1^* - f - k} \tag{35}$$

非正常情况下恶意节点不同意达成共识的概率为:

$$b = \sum_{k=\frac{N_1^*}{3} - f}^{N_1^* - f} C_k^{N_1^* - f} p'^k (1 - p')^{N_1^* - f - k} \tag{36}$$

其中,N_1^* 中的节点对达成共识消息的赞成或反对是随机的,赞成的概率为 p' 且服从二项分 $B(k, N_1^*, p')$。

假设模型出现难题可以被解决的概率为 p,那么正常情况下直接达成共识的概率为:$p_a = (1 - (1-p)^{N_1^* - f}) \times a$。出现恶意节点联合难题但是可以被解决后达成共识的概率为:$p_b = fpb$。经过 Δ 时间后,正常节点返回正确消息的过程服从二项分布 $B_a = B(p_a, \Delta)$,恶意节点返回错误消息的过程服从二项分布 $B_b = B(p_b, \Delta)$。

当 $N_1^* p > 5$,$N_1^* (1-p) > 5$ 时,使用正态分布 $N(N_1^* p, N_1^* p(1-p))$ 与二项分布 $B(N_1^*, p)$ 拟合可得:

$$B_a = N(\Delta p_a, \Delta p_a (1 - p_a)) \tag{37}$$

$$B_b = N(\Delta p_b, \Delta p_b (1 - p_b)) \tag{38}$$

根据一致性协议,需要保证 $B_a - B_b > 0$。由正态分布的性质可得:

$$B_a - B_b = N(\Delta(p_a - p_b), \Delta p_a (1 - p_a) + \Delta p_b (1 - p_b)) \tag{39}$$

由标准正态分布性质可得:

1) $X \sim N(0,1)$,$p = P\{-k < X < k\}$。

2)当 $k=3$ 时,$p = 99.73\%$;当 $k=4$ 时,$p = 99.99\%$。为了使 $\frac{\Delta(p_a - p_b)}{\sqrt{\Delta p_a (1-p_a) + \Delta p_b (1-p_b)}} > k$ 成立,则有:$\Delta > k^2 \frac{p_a(1-p_a) + p_b(1-p_b)}{(p_a - p_b)^2}$。当 $k=4$,Δ 取大于 $16 \frac{p_a(1-p_a) + p_b(1-p_b)}{(p_a - p_b)^2}$ 的值时,就有 99% 的概率达成一致,因此蒙特卡洛一致性得证。

6 与现有医疗安全模型的安全性对比

本节通过与现有医疗健康数据安全模型研究中的一些解决方案进行对比,来验证本文模型的安全性,并分析其优缺点。相关对比情况如表1所列。

表1 安全模型与部分解决方案

Table 1 Security model vs. partial solution

	基于区块链	关注医疗问题	共识机制	联盟链
SNOW[18]	×	√	无	×
DEPR[19]	×	√	无	×
MedRec[2]	√	√	POW	√
BBSN[3]	√	√	POW	×
Model Chain[4]	√	√	POI	×
HDG[20]	√	√	POS	×
MDSM[21]	√	√	改进 DPOS	×
AMDCM	√	√	DPOS+PBFT	√

结束语 随着智慧医疗的快速发展,中心化存储和共享医疗健康数据的方式终将会被淘汰,区块链技术的兴起是解决这一问题的途径之一。利用区块链技术的分布式、防篡改等特点,可以在医疗健康数据的安全存储和安全共享过程中降低管理成本,这会对医疗行业的未来发展产生重大影响。文章提出的医疗健康数据安全模型可以满足医疗机构在现有医疗条件的基础上对数据进行安全存储和安全共享,通过混合共识机制与传统加密方式的结合,为区块链技术与医疗大数据搭建了基础平台,这在一定程度保护了医疗健康数据的隐私。但是,利用分布式数据库及单一加密机制存储和加密原始医疗数据的方式,在安全性方面还是具有一定的局限性。下一步,我们致力于将区块链与云存储相结合,使用混合加密机制保护医疗健康数据隐私的研究,进一步使得所提安全模型具有隐私保护的功能。

参 考 文 献

[1] CHRISTIDIS K, DEVETSIKIOTIS M. Blockchains and smart contracts for the Internet of Things[J]. IEEE Access, 2016, 4: 2292-2303.

[2] AZARIA A, EKBLAW A, VIEIRA T, et al. MedRec: using blockchain for medical data access and permission management

[C]// 2016 2nd International Conference on Open and Big Data (OBD). Vienna, Austria. IEEE, 2016: 25-30.

[3] ZHAO H W, ZHANG Y, PENG Y, et al. Lightweight backup and efficient recovery scheme for health blockchain keys[C]// 2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS). Bangkok, Thailand: IEEE, 2017.

[4] SHRIER A A, CHANG A, DIAKUN-THIBAULT N, et al. Blockchain and health IT: algorithms, privacy, and data[OL]. http://www.truevaluemetrics.org/DBpdfs/Technology/Blockchain/1-78-block-chainandhealthitalgorithmsprivacydata_whitepaper.pdf.

[5] ICHIKAWA D, KASHIYAMA M, UENO T. Tamper-resistant mobile health using blockchain technology[J]. JMIR MHealth and UHealth, 2017, 5(7): e111.

[6] CHEN L, XU L, GAO Z, et al. Protecting Early Stage Proof-of-Work Based Public Blockchain[C]// 2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W). IEEE, 2018: 122-127.

[7] LI D, WEI J W. Theory, application fields and challenge of the blockchain technology[J]. Telecommunications Science, 2016, 32(12): 20-25.

[8] HE P, YU G, ZHANG Y F, et al. Survey on blockchain technology and its application prospect[J]. Computer Science, 2017, 44(4): 1-7, 15.

[9] CASTRO M, LISKOV B. Practical byzantine fault tolerance [C]// Proceedings of the Third Symposium on Operating Systems Design and Implementation. New Orleans: Usenix Association, 1999: 173-186.

[10] EASTLAKE D, JONES P. Message digest (MD5) algorithm and secure hash algorithm (SHA)[M]// Encyclopedia of Multimedia. Boston, MA: Springer US, 2006: 407-408.

[11] SZYDLO M. Merkle tree traversal in log space and time[M]// Advances in Cryptology - EUROCRYPT 2004. Berlin: Springer, 2004: 541-554.

[12] CORON J S, DODIS Y, MALINAUD C, et al. Merkle-damgard revisited: how to construct a hash function[M]// Advances in Cryptology - CRYPTO 2005. Berlin: Springer, 2005: 430-448.

[13] BLAZE M, BLEUMER G, STRAUSS M. Divertible protocols and atomic proxy cryptography[M]// Lecture Notes in Computer Science. Berlin: Springer, 1998: 127-144.

[14] NISHIMAKI R, XAGAWA K. Key-private proxy Re-encryption from lattices, revisited[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2015, E98. A(1): 100-116.

[15] ATTIYA H, CENSOR-HILLEL K. Lower bounds for randomized consensus under a weak adversary[J]. SIAM Journal on Computing, 2010, 39(8): 3885-3904.

[16] CAO B, LIN L, LI Y, et al. Review of blockchain research[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2020: 32(1): 1-14.

[17] WU T, HUANG K, ZHOU L L, et al. Research on Blockchain Consistency Algorithm with State Legality Verification [J]. Computer Engineering, 2018, 44(1): 160-164.

[18] HAILEMICHAEL M A, MARCORUIZ L, BELLIKA J G. Privacy-preserving Statistical Query and Processing on Distributed OpenEHR Data[J]. Studies in Health Technology & Informatics, 2015, 210: 766-770.

[19] KEMKAR O S, KALODE P. Formulation of Distributed Electronic Patient Record (DEPR) System Using Openemr Concept [J]. International Journal of Engineering Innovations and Research, 2015, 4(1): 85-89.

[20] 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.

[21] XUE T F, FU Q C, WANG C, et al. A medical data sharing model via blockchain[J]. Acta Automatica Sinica, 2017, 43(9): 1555-1562.

FENG Tao, born in 1970, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include network and information security.