计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190400001

融入区块链技术的医疗数据存储机制

王 辉 刘玉祥 曹顺湘 周明明
南京工业大学计算机科学与技术学院 南京 211816
(wanghui@njtech.edu.cn)

摘 要 医疗机构现有数据库存储的单一性和集中性,使得电子医疗数据的安全性、完整性和可追溯性无法得到保证,从而导致患者的医疗隐私受到威胁。虽然已有研究提出了基于云存储等数据安全存储的方案,但是其需要依赖一个完全可信的第三方来保证交互的可靠性。为此,文中提出了去中心化的区块链信息管理方案来实现医疗数据的安全存储。该方案采用了改进PBFT共识算法和优化Hash加密算法,将医疗数据安全有效地存储于分布式数据库中,保证了医疗数据的完整性和可追溯性;同时,设计了全新的数据交互系统来阻止第三方与数据库的直接交互,以防止不可信的第三方恶意破坏医疗数据,保证了数据的安全性;最后,通过访问控制与Lucene检索机制保护患者的隐私并实现医疗数据的快速检索。实验分析表明,相较于工作量证明(Proof of Work,POW)、股份授权证明(Delegated Proof of Stake,DPOS)等算法,改进的PBFT共识算法为医疗区块链系统提供了更优的稳定性和吞吐量;相比于普通的数据库交互,数据交互系统有效地阻止了对数据库的直接操作,具有较好的安全性和防篡改性。实验数据表明,去中心化的医疗数据存储系统、改进的PBFT共识算法以及数据交互系统的架构,实现了医疗数据的安全、可追溯和防篡改,解决了医疗数据集中存储、不可追溯和易受攻击等难点,为进一步推动区块链技术应用于医疗信息行业的发展奠定了基础。

关键词:医疗区块链;共识算法;隐私保护;数据交互;访问控制
中图法分类号 TP393

Medical Data Storage Mechanism Integrating Blockchain Technology

WANG Hui,LIU Yu-xiang,CAO Shun-xiang and ZHOU Ming-ming
School of Computer Science and Technology,Nanjing University of Technology,Nanjing 211816,China

Abstract The Singularity and centrality of Medical Institutions' existing database storage makes the security,integrity and traceability of electronic medical data impossible to be guaranteed,as a result,the medical privacy of patients is threatened. Although existing research has proposed a secure data storage scheme based on cloud storage,it needs to rely on a fully trusted third party to ensure the reliability of interaction. Therefore,this paper proposed a decentralized block chain information management scheme to achieve the safe storage of medical data. This scheme adopts improved PBFT consensus algorithm and optimized Hash encryption algorithm to store medical data safely and effectively in distributed database to ensure the integrity and traceability of medical data. At the same time,it proposes and designs a new data interaction system to prevent the direct interaction between the third party and the database,prevent the untrustworthy third party from maliciously destroying medical data and ensure the data. Finally,through access control and Lucene search mechanism to ensure patient privacy and achieve rapid retrieval of medical data. Experiments show that the improved PBFT consensus algorithm provides better stability and throughput than proof of work(POW) and delegated proof of stake (DPOS). Compared with the common database interaction,the data interaction system in this paper effectively prevents the direct operation of the database and has better security and tamper resistance. The experimental data show that the decentralized medical data storage system,the improved PBFT consensus algorithm and the data interaction system architecture have realized the security,traceability and tamper-proof of medical data,solved the difficulties of centralized storage,traceability and vulnerability of medical data,and laid a foundation for further promoting the application of block chain technology in the development of medical information industry.

Keywords Medical blockchain,Consensus algorithm,Privacy protection,Data interaction,Access control

1 引言

全球在解决电子医疗数据(Electronic Medical Record,EMR)的存储、校验和可追溯性等问题上一直存在难点,患者和医生在访问EMR时会受到严格的限制,需要花费大量的资源和时间进行许可校验[1]。由于EMR需要经常在医院、

到稿日期:2019-04-01 返修日期:2019-07-21 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。
通信作者:刘玉祥(lyx_echo@njtech.edu.cn)

研究人员和患者等之间进行分发和共享，因此保证 EMR 的安全性十分重要。其次，EMR 对保持病人病史的完整性和获取的实时性也意义重大。患有严重疾病或者慢性疾病的患者必须保持治疗过程和治疗后的不间断监测和康复，因此获得完整和最新的病史记录对他们的治疗至关重要。EMR 通常存储在医院单独的数据库中，集中存储导致其信息价值较大，易成为攻击重点。EMR 一旦被第三方恶意攻击，则数据的安全性、完整性和不可更改性可能无法得到保证，因此医疗数据的安全问题亟待解决。

近年来，加密技术、大数据、云存储等相关技术得到快速发展，研究人员已提出了基于同态加密的数据安全存储和基于云存储服务数据安全机制等方案。但是，这些安全服务方案都依赖于一个完全可信的第三方来保证交互的可靠性，一旦第三方信任机构遭到攻击，则所有的服务都不再安全[2]。区块链技术的发展，为解决医疗数据的安全存储和复杂的权限限制结构提供了一种全新的去中心化模式。国内区块链与医院相结合的研究刚刚起步，国外则相对成熟。Azaria[3] 等提出了 MedRec——一个基于 POW 的医疗区块链，用于对医疗数据的访问许可进行管理。Esposito 等[4] 提出了基于云存储的医疗区块链系统来保护数据的隐私安全，但是采用的 POW 共识机制需要消耗大量的算力。Patel[5] 通过区块链共识机制，提出并设计了一个安全去中心化的医疗图片数据的分享。在现有研究的基础上，结合目前国内的 EMR 存储系统，本文利用以太坊联盟链设计了一个使用区块链技术的医疗数据安全存储方案。该方案利用区块链的去中心化、不可篡改等特点保证了 EMR 的安全存储与共享，解决了 EMR 存储集中、可追溯性难等问题，达到了隐私规则和安全规则的标准。相比其他 EMR 安全研究，本文建立了访问控制机制，使患者能够更好地控制自己的医疗数据；在安全存储上，将 EMR 的 Hash、索引等组成的医疗元数据块存储在区块链的分布式账本中，通过数据交互系统(Data Interactive System, DIS)将 EMR 等组成的医疗数据块加密存储在链下分布式数据库中，这样既减轻了主链的压力，又增强了系统的可扩展性；通过改进现有的 PBFT 共识算法以及 Hash 算法保障了 EMR 的安全性和不可篡改，最后通过 Lucene 快速检索机制保证了用户对 EMR 的快速检索查看。

本文第 2 节介绍了区块链的相关背景知识；第 3 节提出了基于区块链的 EMR 存储机制研究方案，从系统框架的不同层次介绍了研究方案和改进技术；第 4 节介绍了医疗数据存储方案模型的具体存储过程和访问控制等；第 5 节从系统性能和安全性等方面对系统模型进行了实验和理论的评估，并通过对比分析了系统模型的优劣。

2 理论背景

区块链是一个链表的数据结构，每个区块相互连接，存储了一系列的有序事物，是一个特殊的数据库。它与普通数据库最主要的区别是数据加入数据库的方式不同，其具有一致性、不变性、典型性和去中心化[6]。其中，去中心化的特性最为关键，它意味着整个系统中不会出现单一故障点，没有任何第三方可以私自篡改数据。区块链系统中的数据审查过程对用户完全透明，用户不需要知道系统如何审核数据，也不需要确定应该信任系统中的哪一个节点[7]。

2.1 工作原理

区块链交易的工作原理和流程可以分成两部分来理解：交易和区块。区块链的工作原理如图 1 所示。

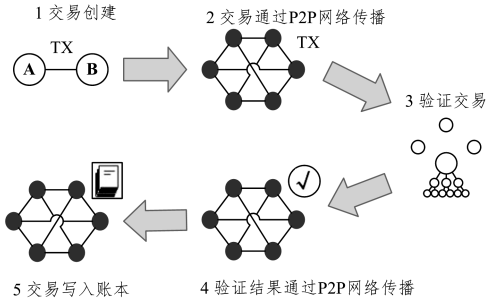

图 1 区块链的工作原理

Fig. 1 Working principle of blockchain

步骤 1 交易生成

当前所有者使用私钥为前一次交易和下一个所有者进行数字签名，并将数字签名附加到本次交易尾部以制作成交易单[8]。新的交易产生时，会将其广播到区块链网络中的其他参与节点。

步骤 2 交易传播

当前所有者将交易广播至整个网络，并且每个节点都会将大量未验证的交易哈希值收集到块中。最先完成 POW 的节点会将自己的区块传播到其他节点[9]。

步骤 3 共识机制

共识机制有 POW，PBFT 和 DPOS 等，如比特币是以 POW 为共识机制。工作量证明机制的核心思想是分布式网络中的每个节点都去竞争一道数学题，由最快算出结果的节点获得创建新块的权利，并验证交易[10]。

步骤 4 全节点验证

当一个节点找到满足要求的数字时，它会将该区块记录的所有带时间戳的交易广播至全网，由全网其他节点进行审查。其他节点将确认该区块中所包含交易的有效性，确认其没被双花(重复消费)并且具有有效的数字签名后，接受该区块，正式链接到区块链中且保证其不再被篡改[10]。

步骤 5 区块链记录

全网其他节点对该区块记账的正确性进行审核，在没有错误之后，它们将在该合法块之后竞争下一个块，从而形成一个合法记账的区块[11]。为保证系统的稳定性，全网的算力随着区块创建的时间不断变化，以保证大约每 10 min 产生一个区块[12]。

2.2 共识算法

区块链中的核心技术是共识算法：比特币使用的是依赖于算力的 POW；以太币使用的是股权证明(Proof of Stake, POS)，减轻了对算力的依赖；DPOS 则进一步减少了算力的浪费，同时也加强了区块链的安全性。本文研究方案是基于联盟链的医疗区块链，需要提供高吞吐和低时延的特性，因此选择 PBFT 作为改进算法。对由 n 个共识节点组成的共识系统，PBFT 共识算法提供了 $f=\lfloor(n-1)/3\rfloor$ 的容错能力。假

设系统要求区块每次产生的时间间隔为 t，则算法的流程图如图 2 所示，具体执行流程如下：

1）任意节点向全网广播带有数字签名的交易数据；

2）所有共识节点独立监听和记录全网的交易数据；

3）主节点在经过时间 t 后，发送〈PrepareRequest，h，v，p，$block$，$\langle block\rangle_{\sigma p}$〉；

4）副节点 i 在收到提案后，发送〈PrepareResponse，h，v，i，$\langle block\rangle_{\sigma i}$〉；

5）任意节点在收到至少 $n-f$ 个$\langle block\rangle_{\sigma i}$后，达成共识并发布完整区块；

6）收到完整区块后，节点都将删除内存中包含的交易，并开始下一轮共识。

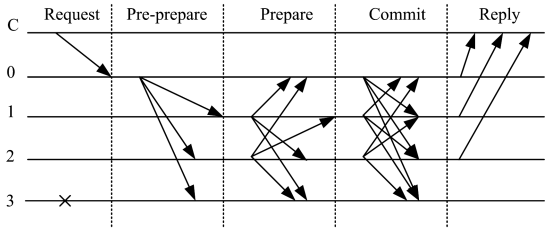

图 2　PBFT 共识过程

Fig. 2　PBFT Consensus process

该算法要求参与共识的节点中至少有 $n-f$ 个节点具有相同的初始状态，即所有的节点 i 具有相同的区块高度 h 和视图编号 v。节点监听全网交易，在收到提案后，需要对交易的合法性进行验证。如果发现非法交易，则节点禁止将其写入内存池；如果提案包含非法交易，则共识算法放弃本次共识，并立即更换视图。

2.3　加密算法

区块链技术需要 Hash 算法和非对称加密算法等多种加密算法共存，以进行数据加密及隐私保护[17]。不同的加密算法性能不同，单一算法很难满足互联网中各个应用的需求，一般需要组合使用。目前国际上主要的 Hash 算法是 MD5 和 SHA，如比特币区块链使用的是 SHA-256 。非对称加密时产生两把密钥，分别用于数据加密和数据解密：公钥对外公开，用于加密；私钥用于解密，由使用者自己保管，不对外公开[18]。常用的非对称加密算法有 ECC 和 RSA 椭圆曲线算法，比特币和以太坊中均使用了 ECC 椭圆曲线算法[18]。

3　系统描述

图 3 显示了系统的整体架构流程。系统由用户层和系统处理层两个结构组成，包含 6 个主要部分，即请求连接池、密钥中心、各级医院组成的区块链网络结构、共识中心、数据交互系统和分布式数据库。

用户层由所有从系统获取数据的用户组成，如患者、医生等。在医院就诊后，患者更加关心个人医疗数据的隐私性、完整性和快速可查性。患者可以使用自己的私钥对个人医疗数据进行签名，以保证医疗数据的隐私性；医生通过医疗区块链将患者的医疗元数据块和医疗数据块上传至区块链和分布式数据库，以保证数据的完整性和不可篡改；之后，患者可以通过自己的私钥检索电子医疗数据。

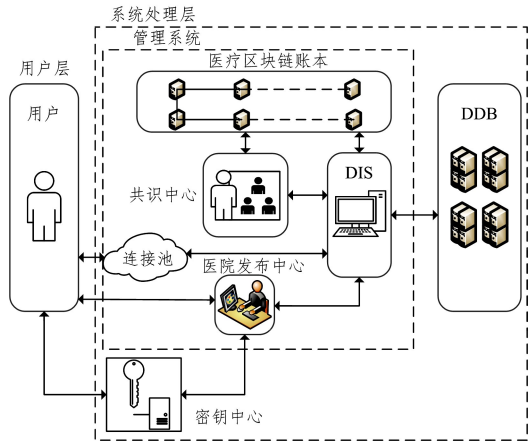

图 3　医疗区块链系统架构

Fig. 3　Medical blockchain system architecture

系统处理层由医院的管理系统和分布式数据库组成。管理系统主要包括区块链分布式账本、共识中心、发布中心及数据交互系统。医疗区块链中的网络节点由各级医院组成，根据医院等级，本文将各医院在链中的节点设定为普通节点和共识节点。调查研究发现，医疗大数据和服务中心一般都建立在高水平医疗机构，如三甲医院和高水平医疗研究所等，因此本文设定共识节点主要由全国各三甲医院等高水平医疗机构组成。普通节点不参与全局账本的记账，但是需要同步整个账本，并且可以用其私钥将患者的医疗数据进行签名，然后提交至参与共识的上级医院来发布。参与共识的医院节点需要将数据打包成医疗元数据块和医疗数据块，并向医疗区块链中的共识中心发送请求，主节点将所有医疗元数据块打包进区块，通过共识添加至医疗区块链，之后再将医疗数据块加密存储在链下分布式数据库中。

如图 4 所示，医疗数据分为医疗元数据块和医疗数据块。其中，Patient ID 为患者的公钥，通过公钥字段可以很方便地查到患者的医疗数据记录。

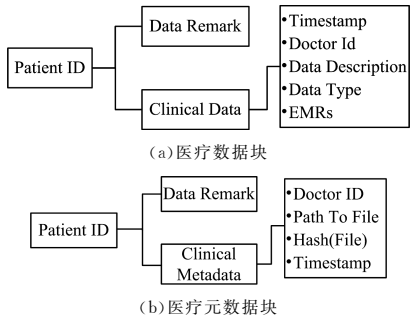

图 4　医疗数据的数据结构

Fig. 4　Data structure of medical data

图 4(a)所示为患者的医疗数据块结构，包含诊断文档、医疗图片、视频等，加密存储在链下分布式数据库中，以减轻主链压力。其包含临床数据块和数据备注。

(1)临床数据块

1）时间戳 Timestamp。

2）Doctor ID：主治医师的 ID，即医师的公钥。

3）Data Description：医疗数据的主要描述。

4）Data Type：医疗数据的数据类别。

5)EMR:医疗数据的数据文件。

(2)数据备注:数据备注是医师记录该次诊疗的相关备注,用于患者在下次治疗时为医师提供上次治疗的效果和注意事项。

图 4(b)是患者的医疗元数据块结构,发布至区块链账本中。医疗区块链网络中所有节点同步备份,其中包含临床元数据块和数据备注。

(1)临床元数据块中包含了主治医师上传到分布式数据库中的所有医疗数据文件的信息。

1)时间戳 Timestamp。

2)Doctor ID:主治医师 ID,即医师的公钥。

3)Path To File:文件路径,指向存储在分布式数据库中的文件的指针。文件路径使用患者的私钥进行签名,签名后本身是需要进行隐藏的,因为账本是公开的,所有人都可以查阅到,所以对未授权认证的用户隐藏文件路径。用户提供自己的私钥签名作为主要的身份认证,数据安全交互系统对私钥签名进行验证,仅当匹配到正确的签名时才会向授权用户展示被隐藏的内容。

4)Hash(File):数据文件的 Hash,确保存储在分布式数据库中的数据文件的不可伪造和篡改。

(2)数据备注块:与存储在分布式数据库中的数据备注类似,医师可以添加诊疗的相关备注信息,方便后期跟进治疗。

3.1 改进的 PBFT 算法

PBFT 共识算法的共识节点承担着医疗数据的发布和存储工作,以保证医疗数据的隐私性、不可篡改和完整性。传统的 PBFT 属于典型的三阶段协议,由式(1)可知:随着广播消息数量的增加,网络宽带会急剧减小[19]。

$$BandWidth = N \cdot (N-1) \cdot Blockchain \tag{1}$$

本方案中的 PBFT 共识算法主要是针对医疗区块链而设计,目的在于全网对医疗数据的有效性达成共识,不涉及请求排序等问题,因此可以取消确认阶段,保留预准备和准备阶段。这将提高信息的传输速率,缩短传输时间,从而实现高吞吐和低时延。

为了减小医疗数据在共识过程中受错误节点的影响,采用信任管理方法对 PBFT 共识算法中的共识节点 N_i 进行可靠性考查。在该模型中,信任被定义为节点成为共识节点后发布医疗数据至医疗区块链的可靠程度,用 0～1 之间的数值来对该信任节点进行量化与评估。对医院节点共识过程中发生的错误进行监测,利用 beta 信誉系统构建 t 阶段 N_i 节点的信任值 $T_i(t)$,阶段 t 以月为单位,一年共 12 个阶段。节点的信任值越高,其发生错误的概率就越低,即可靠性越高,则在下一阶段被选为共识节点的概率就越大。系统采用改进的"Watchdog"监控技术来监测节点。

beta 分布通常用来表示一个二元事件的后验概率,主要用参数 α,β 和 Γ 函数来表示,如式(2)所示:

$$f(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \tag{2}$$

其中,$0 \leqslant p \leqslant 1$,$\alpha > 0$,$\beta > 0$。

beta 分布的概率期望如式(3)所示:

$$E(f(p|\alpha,\beta)) = \frac{\alpha}{\alpha+\beta} \tag{3}$$

对于二元事件$\{X, \overline{X}\}$,用 s,f 分别表示 X 和$\overline{X}$出现的次数,变量 p 表示 X 出现的概率。通过对式(3)中的 α 和 β 进行设置,事件 X 出现的概率密度可以表示为历史统计数据的函数。

$$\alpha = s+1, \beta = f+1 \tag{4}$$

在 PBFT 共识算法中,将节点参与共识发布医疗数据的行为看作独立同分布的二项事件,在共识过程中没有发生错误的事件记为 X,发生错误的事件记为$\overline{X}$。在每一个阶段 t 内,用 s_i 和 f_i 分别表示监测到共识节点未发生错误和发生错误的次数。利用二项事件后验概率服从 beta 分布的特性,计算出节点 N_i 的信任值 $T_i(\tau)$,如式(5)所示:

$$T_i(\tau) = \frac{s_i+1}{s_i+f_i+2} \tag{5}$$

3.2 改进的 Hash 算法

在区块链中,信息是完全公开的,第三方可以轻易地获取到存储在分布式账本中的数据的 Hash 值。不同于一般用户数据,医疗数据具有很大的价值,所以存在恶意的第三方会暴力破解医疗数据。2004 年国际密码学会议也宣布 MD5 不再是一种安全的加密算法,因此简单的 MD5 加密是不可能实现绝对安全的。

本方案研究使用 MD5 加密算法重新设计密文,截取密文中的一段数据并丢弃,然后用随机函数填充丢弃的部分,并且整个过程不改变 MD5 加密后的位数。加密过程用算法描述如下:

(1)对明文 *message* 进行 MD5 加密,获得密文 *MD5*(*message*);

(2)对密文从开始位置截取到位数 *number* 的 1/2 处,得到密文 A,其中 $A = left(MD5(message), number/2)$,密文的剩余部分丢弃;

(3)使用随机函数 *gen_key*(*number*/2)填充密文丢失部分;

(4)变换后的密码值为 $encrypt\_message = A \& gen\_key(message/2)$。

其中,*encrypt_message* 是经过处理后的密文,*number* 是 MD5 加密后密文的位数。

经过修改的密文与原先的密文完全不同,即使算暴力破解,得到的结果也不是真正的数据。但是,修改后的密文仍然具有验证数据是否被篡改的功能,验证过程如下:

(1)先对 *message* 进行加密;

(2)然后从 *beginNumber* 处截取前半部分得到 A';

(3)与 *encrypt_message* 中的 A 进行对比,如果 $A' = A$,则认为数据没有被篡改。

3.3 数据交互系统

本方案提出并设计了一种新的数据交互系统(Data Interaction System,DIS)来保证患者对医疗数据具有完全的所有权。它是一种采用对区块链进行监测的方式,来控制分布式数据库所有操作行为的系统。DIS 通过交易 ID 对区块链进行监测,从而授权第三方医疗机构对分布式数据库进行读、写,避免了第三方对数据库的直接操作,防止了不可信的第三方对数据库直接操作时恶意盗取和破坏医疗数据。

4　医疗数据存储方案

4.1　数据发布与存储

密钥中心是密钥的分发机构，并且对患者的医疗数据执行加密操作。当患者就诊后，医生使用患者和主治医生的私钥对医疗数据进行数字签名并发送至发布中心。对于加密过程，本文提出的用于实现加密哈希生成器的改进 MD5 加密技术使用密文拼接来提升安全级别。在固定的时间段内，发布中心将所有上传的医疗元数据块打包成块，而后通过共识机制上传到区块链；成功上传至区块链后，发布中心向数据交互系统申请将医疗数据块存储到分布式数据库中；数据交互系统根据对区块链的监测，判断是否通过医疗区块链的共识认证，通过后则允许医疗数据存储至分布式数据库中。具体存储过程如下：

(1)患者就诊后，主治医生将使用患者和自己的私钥对医疗数据进行数字签名，生成医疗数据的 Hash 值，并生成医疗数据块和医疗元数据块，然后将其一起发送至医院系统的发布中心。

(2)发布中心在固定的时间内收集上传的医疗数据，并将医疗元数据块打包后提交至医疗区块链系统等待共识认证。

(3)共识节点基于改进的 PBFT 共识算法对医疗数据块进行共识认证。

(4)通过共识认证后，医疗元数据块将被上传到医疗区块链的分布式账本中。医疗元数据块成功发布至区块链后，发布中心通过数据交互系统将医疗数据块安全地存储到医院的分布式数据库中。

(5)返回第(1)步循环进行。

4.2　访问控制

区块链的分布式账本是公开的，这意味着每个人都可以搜索和浏览数据。为了保护患者的隐私，本方案将医疗区块链中的文件路径进行了加密隐藏，需要用户提供私钥作为身份认证和解密密钥，数据交互系统再根据文件路径在分布式数据库中查询用户的医疗数据并返回。为解决用户频繁查找数据所带来的效率问题，本文提供了访问连接池，用户在短时间内多次通过数据交互系统访问数据库时可以更加高效地对数据进行读取。

4.3　Lucene 检索机制

本文通过 Lucene 提供两种检索方案：1)通过用户的公钥 ID 进行精确检索；2)对输入的用户姓名等关键词进行模糊检索。

(1)计算查找词权重(Term Weight)

1)Term Frequency(tf)：Term 在数据文件中出现的频次。

2)Document Frequency(df)：包含 Term 文档的频次。

计算公式如式(6)所示：

$$w_{t,d}=tf_{t,d}\times\log(n/df_t) \tag{6}$$

(2)向量控件模型算法

将数据中所有的词(Term)和权重(Term Weight)表示为一个向量。

$$\boldsymbol{Document}=\{term_1,term_2,\cdots,term_N\}$$

$$\boldsymbol{DocumentVector}=\{weight_1,weight_2,\cdots,weight_N\}$$

将查询的查找词等也看作数据，用向量来表示。

$$\boldsymbol{Query}=\{term_1,term_2,\cdots,term_N\}$$

$$\boldsymbol{Query\ Vector}=\{weight_1,weight_2,\cdots,weight_N\}$$

图 5 将所有的搜索数据结果向量和查询向量放入 N 维空间，每个词(Term)是一维。

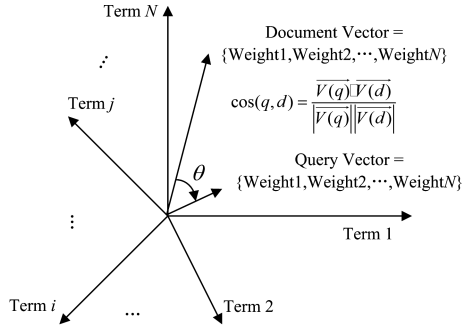

图 5　向量空间模型

Fig. 5　Vector space model

计算夹角的余弦值作为相关性判断的依据，并根据数据的最高相关性依次返回搜索结果。

相关性计算如式(7)所示：

$$correlation=\frac{\vec{V_q}\cdot\vec{V_d}}{|\vec{V_q}||\vec{V_d}|}=\frac{\sum_{i=1}^{n}w_{i,q}w_{i,d}}{\sqrt{\sum_{i-1}^{n}w_{i,q}^2}\sqrt{\sum_{i-1}^{n}w_{i,d}^2}} \tag{7}$$

5　系统分析

5.1　实验准备

实验测试环境为：Windows 10 操作系统，8 GB RAM，Inter Core i7-5930K。为了保证交易可以快速地进行并更好地获取数据，设定每次交易的量为 1。

5.2　性能分析

5.2.1　系统稳定性

区块链系统中区块的分叉会影响区块链的可靠性，所以比特币区块链为了消除分叉的影响会保持块的生成时间是 10 min。因此，区块生成过程中分叉节点对系统的影响是至关重要的，本方案在测试中选取 POW 和 DPOS 共识算法进行分析比较。为了模拟现实网络中的网络延迟现象，测试过程中添加延迟算法以保持每个实验的网络延迟不变，并且允许系统的区块一致增长到相同的高度(100)。实验结果通过 Matlab 仿真，如图 6 所示。

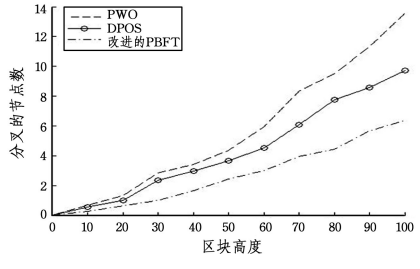

图 6　区块高度与分叉节点数的关系

Fig. 6　Relationship between block height and number of fork nodes

分析实验结果得知，当区块高于 20 时，本方案分叉节点数开始明显低于 POW 和 DPOS 共识算法。随着区块高度的

增加，POW 和 DPOS 分叉节点数开始急剧增加，而基于改进的 PBFT 算法的分叉节点数较 POW 和 DPOS 增加缓慢，医疗区块链受分叉节点的影响较小，系统更加稳定。

5.2.2 吞吐量

吞吐量衡量了系统在单位时间内处理交易的能力，是系统处理并发能力的重要指标，本文用 TPS(Transaction Per Second)表示。医疗区块链中吞吐量是单位时间内医疗数据写入区块链中的总交易数，如式(8)所示：

$$TPS_{\Delta_t}=\frac{SumTransaction_{\Delta_t}}{\Delta_t} \tag{8}$$

其中，Δ_t 为医疗数据通过共识的时间间隔，$SunTransaction_{\Delta_t}$ 为该时间间隔内区块中总的医疗数据块。

本次实验取 Δ_t 分别为 10 s，20 s，40 s，60 s，80 s，100 s 这 6 个不同的时间间隔，测试取出各个时间段的 TPS。测试所得 TPS 总数的 Matlab 仿真如图 7 所示。

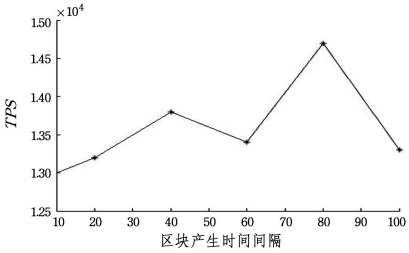

图 7 TPS 与出块时间的折线图

Fig. 7 Line chart of TPS and block time

选取测试的平均值作为改进 PBFT 算法的 TPS 值，并将其与目前较为成熟的区块链共识算法的 TPS 进行比较，结果如图 8 所示。实验结果显示，改进 PBFT 算法在提供 $f=\lfloor(n-1)/3\rfloor$ 的容错前提下，吞吐量较原始的 PBFT 算法提升了 29%。公有链中的 POW 和 POS 等算法为了保证系统较高的容错性，单位时间内的吞吐量低于 10，无法在短时间内完成大量的交易；本方案保证了系统足够的容错性能，并且在吞吐量上明显优于 POW 和 POS 等，为医疗区块链提供了很好的交易吞吐性能。

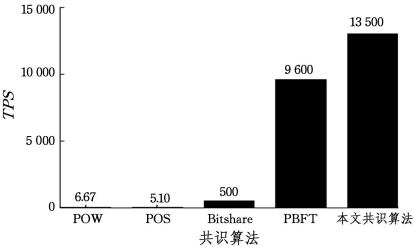

图 8 共识算法的 TPS 比较

Fig. 8 TPS comparison of consensus algorithms

5.3 安全性分析

本文采用改进的 PBFT 共识算法以及改进的加密技术等来保证系统的安全性，下面从防篡改、容错能力分析以及网络攻击等方面来分析系统的安全性。

5.3.1 隐私保护、防篡改

存储在区块链中的医疗元数据块存储了时间戳、存储路径和 Hash 值等，在访问控制协议中，由于数据存储路径被隐藏，身份未认证的第三方无法获取数据存储路径；并且区块链交易中都是通过患者和医生的私钥进行数字签名，再通过共识上传至区块链，而私钥的隐秘性可以保证交易的安全性和匿名性。

在现有普通的 Hash 加密技术无法完全保证可靠的安全性情况下，本文区块链系统改进了传统的 Hash 加密技术，保证了数据无法被暴力破解，但是依然可以验证数据是否被篡改。在数据库连接上，设计了一种全新的数据交互系统(DIS)，对区块链进行监测，认证有效才允许通过 DIS 对数据库进行读写操作，阻止了第三方对数据库的直接操作，避免了恶意的第三方对数据库进行暴力攻击。

5.3.2 网络攻击

本文假设攻击链(Attack Chain，AC)在系统主链(Main Chain，MC)上进行攻击，导致主链节点分叉。根据区块链系统的机制，当分叉的 AC 长度超过 MC 时，攻击成功。AC 攻击成功的概率为：

$$P=\begin{cases}1, & p_{MC}\leqslant p_{AC}\\ \dfrac{p_{AC}}{p_{MC}}, & p_{MC}>p_{AC}\end{cases} \tag{9}$$

其中，P 是 AC 攻击成功的概率；p_{AC} 为攻击节点造出下一个区块的概率；p_{MC} 为诚实节点造出下一个区块的概率。

当 MC 已经产生了 k 个区块后，AC 超过 MC 攻击成功的概率为：

$$P=1-\sum_{x=0}^{k}\frac{\alpha^x e^{-\alpha}}{x!}\cdot\left(1-\left(\frac{p_{AC}}{p_{MC}}\right)^{(k-x)}\right),x=0,1,2,\cdots \tag{10}$$

其中，P 是 MC 已经产生 k 个区块后攻击成功的概率；$\alpha=k\left(\dfrac{p_{AC}}{p_{MC}}\right)$ 是攻击者进展的泊松分布期望。

实验设定 p_{AC} 的值为 0.1，图 9 给出了攻击节点在主链先产生区块后攻击成功的概率。从实验结果可以得知，当 MC 已经产生的区块数 k 在不断增加时，p_{AC} 攻击成功的概率呈指数下降，当主链 MC 产生的区块数超过 10 时，恶意节点攻击成功的概率几乎为 0。在医疗区块链系统中，主链中完成交易的区块数量远多于 10，因此恶意节点攻击成功的概率几乎为 0，不会对系统的安全性造成影响。

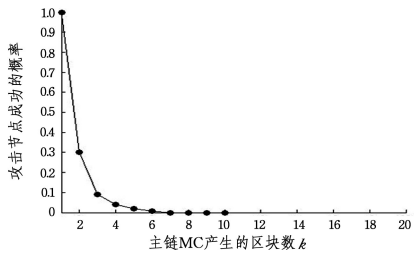

图 9 攻击节点攻击成功的概率

Fig. 9 Probability of succeessful attack of attacking node

5.4 对比分析

表 1 从 4 个方面将本文方案与其他研究方案做出了比较。从表中可以看出，本方案设计并使用了改进的 PBFT 共识算法和加密算法，在系统安全性和吞吐量上具有一定优势；同时通过数据交互系统阻止了第三方对数据库的直接操作，避免了恶意第三方对数据库直接操作时的破坏。但是与其他

研究成果相比，本文方案也有很多不足，如智能合约等。

表1　方案对比

Table 1　Scheme comparison

	MedRec[3]	文献[15]	本文
共识机制	PoW	PoI	改进的 PBFT
主链压力	大	小	小
数据库直接访问	是	是	否
智能合约	否	是	否

结束语　在医疗存储领域，数据的安全存储是非常重要的。区块链可以为患者提供安全、去中心化、防篡改的数据库技术，可以在实现医疗数据存储等方面发挥基础性作用。本文提出的医疗数据存储模型可以帮助患者安全、有效地存储和控制他们的医疗数据，包括对他们医疗数据进行永久的存储以保证病史的完整，并且定义了区块链这项新技术应用于医疗行业所需要的架构和层次。区块链技术作为信心技术，世界各国均在实验研究中，本文研究在智能合约等相关技术方面还有待完善和改进。

参 考 文 献

[1] FAN K, WANG S Y, REN Y H, et al. MedBlock: Efficient and Secure Medical Data Sharing Via Blockchain [J]. Journal of Medical Systems, 2018, 42(8): 136.

[2] WANG H, SONG Y J. Secure Cloud-Based EHR System Using Attribute-Based Cryptosystem and Blockchain [J]. Journal of Medical Systems, 2018, 42(8): 152.

[3] AZARIA A, EKBLAW A, VIEIRA T, et al. MedRec: Using Blockchain for Medical Data Access and Permission Management [C] // International Conference on Open and Big Data. IEEE, 2016: 25-30.

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

[5] PATEL V. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus[J]. Health Informatics Journal, 2018, 25(4): 146045821876969.

[6] LI H Y, ZHU L H, SHEN M, et al. Blockchain-Based Data Preservation System for Medical Data [J]. Journal of Medical Systems, 2018, 42(8): 1-13.

[7] CHEN Y, DING S, XU Z, et al. Blockchain-Based Medical Records Secure Storage and Medical Service Framework[J]. Journal of Medical Systems, 2018, 43(1).

[8] BROGAN J, BASKARAN I, RAMACHANDRAN N. Authenticating Health Activity Data Using Distributed Ledger Technologies [J]. Computational and Structural Biotechnology Journal, 2018, 16(7): 257-266.

[9] XIA Q, SIFAH E B, ASAMOAH K O, et al. MeDShare: Trustless Medical Data Sharing Among Cloud Service Providers Via Blockchain[J]. IEEE Access, 2017, PP(99): 1-1.

[10] LIU P T S. Medical Record System Using Blockchain, Big Data and Tokenization[C] // International Conference on Information and Communications Security. Springer International Publishing, 2016.

[11] KIM K J, HONG S P. A Trusted Sharing Model for Patient Records based on Permissioned Blockchain[J]. Journal of Internet Computing and Services, 2017, 6: 75-84.

[12] METTLER M. Blockchain technology in healthcare: The revolution starts here[C] // 2016 IEEE 18th International Conference on E-Health Networking, Applications and Services. IEEE, 2016: 1-3.

[13] LIU P T S. Medical Record System Using Blockchain, Big Data and Tokenization[C] // International Conference on Information and Communications Security. Springer International Publishing, 2016.

[14] 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.

[15] HOY M B. An Introduction to the Blockchain and Its Implications for Libraries and Medicine[J]. Medical Reference Services Quarterly, 2017, 36(3): 273-279.

[16] ZHOU L J, WANG L C, SUN Y R. MIStore: a Blockchain-Based Medical Insurance Storage System[J]. Journal of Medical Systems, 2018, 42(8): 149.

[17] LESLIE M. (Block) Chain Reaction: A Blockchain Revolution Sweeps into Health Care, Offering the Possibility for a Much-Needed Data Solution[J]. IEEE Pulse, 2018, 9(3): 4-7.

[18] HÖLBL, MARKO, KOMPARA M, et al. A Systematic Review of the Use of Blockchain in Healthcare [J]. Symmetry, 2018, 10(10): 470.

[19] XUE T F, FU Q C, WANG W, et al. Research on medical data sharing model based on blockchain[J]. Acta Automatica Sinica, 2017, 43(9): 1555-1562.

WANG Hui, born in 1962, Ph. D, professor. Her main research interests include optical communication and signal processing.

LIU Yu-xiang, born in 1996, postgraduate, is member of China Computer Federation(CCF). His main research interests include blockchain and data security.