doi:10.12068/j.issn.1005－3026.2024.02.002

基于贡献度和数据有效性检验的共识机制

时小虎[1,2]，姚 鑫[1]，孙延风[1,2]，马德印[3,4]

（1.吉林大学 计算机科学与技术学院，吉林 长春 130012；2.吉林大学 符号计算与知识工程教育部重点实验室，吉林 长春 130012；3.吉林大学 商学院，吉林 长春 130012；4.长春工业大学 计算机科学与工程学院，吉林 长春 130012）

摘 要：将区块链技术引入到分布式数据维护系统，旨在解决基于传统中心化数据库的分布式系统存在的数据维护不透明、数据易被篡改、历史记录不可追溯等问题，提出一种基于贡献度和数据有效性检验的共识机制．该算法提出一种贡献度优先的随机可验证领导者选举机制，保证记账权分配的随机性及可验证性．进一步引入密度峰值算法对交易数据有效性进行校验，对打包区块的正确性达成共识．最后将所提出的共识机制应用于梅花鹿分布式养殖场场景，结果验证了密度峰值算法在交易数据有效性检测任务中的准确性和高效性．出块时延分析和安全性分析表明，所提出的共识机制能够满足数据有效性验证的实时性需求，能耗较小，具有很强的灾备能力．

关 键 词：区块链；共识机制；离群点检测；分布式数据维护；溯源

中图分类号：TP 391 **文献标志码**：A **文章编号**：1005-3026(2024)02-0160-11

Contribution Degree and Data Validity Verification Based Consensus Algorithm

SHI Xiao-hu[1,2], *YAO Xin*[1], *SUN Yan-feng*[1,2], *MA De-yin*[3,4]

(1. College of Computer Science and Technology, Jilin University, Changchun 130012, China; 2. Key Laboratory of Symbol Computation and Knowledge Engineering (Ministry of Education), Jilin University, Changchun 130012, China; 3. School of Business and Management, Jilin University, Changchun 130012, China; 4. College of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China. Corresponding author: MA De-yin, E-mail: madeyin_jlu@163.com)

Abstract: The blockchain technology into the distributed data maintenance system is introduced and a contribution degree and data validity verification based consensus algorithm is proposed. In the algorithm, a random and verifiable leader election mechanism is designed according to contribution priority, ensuring the randomness and verifiability of the assignment of accounting rights. Furthermore, it introduces the density peak algorithm to reach the consensus on the correctness of packaged blocks. Finally, the proposed consensus algorithm is applied to the distributed deer farms for experimental comparison. The results validate the accuracy and efficiency of the density peak algorithm in the task of data validity testing. The analysis of block delay analysis and safety shows that the consensus algorithm satisfies the real-time requirements, consumes less energy and has strong disaster preparedness ability.

Key words: blockchain; consensus algorithm; outlier detection; distributed data maintenance; trace back

区块链技术起源于比特币系统，由中本聪在2008年发表的比特币白皮书[1]中提出．比特币系统一经提出后获得了成功，也推动了区块链技术的飞速发展．从区块链1.0比特币系统到区块链2.0智能合约，再到如今的区块链3.0多领域应用场景实现，区块链技术已经广泛应用于金融[2]、电子健康[3]、门禁系统[4]、车联网(IoV)[5]、工业IoT[6-7]等多个领域．

收稿日期：2022-10-31

基金项目：吉林省科技发展计划项目(20210201080GX, 20200101146FG)；吉林省发展改革委员会资助项目(2021C045-9, 2021C044-1)．

作者简介：时小虎(1974-)，男，河北玉田人，吉林大学教授．

区块链技术综合密码学、数学、算法和经济模型等多领域技术,结合P2P网络,采用分布式共识机制解决传统分布式数据库同步问题[8],具有去中心化、匿名性、交易透明、不可篡改、可追溯等特征. 区块链可以被视为一种按照时间顺序将数据区块以链式方式组成特定数据结构的分布式账本[9]. 网络中的所有节点都维护着区块链的完整副本,交易等信息存储在区块中. 当新的区块生成时,所有节点必须就新的区块达成一致,才能完成出块[10]. 在区块链网络中,为了实现去中心化和全局的一致性,共识机制发挥着至关重要的作用,各节点通过共识机制验证和更新区块链账本数据,从而实现区块链账本的统一[11]. 区块链系统共识基于分布式系统共识,区块由获得记账权的节点生成,部分算法在节点获得记账权后同时出块,称为记账权竞选共识机制,另一部分算法在选出记账节点后结合传统共识算法完成出块过程,称为区块生成共识机制[12]. 本文将介绍这两类共识机制的原理及其代表算法.

在记账权竞选算法中,记账权竞选节点主动参与记账权竞选并付出较高竞选代价,其余节点验证竞选结果时只需付出较低验证代价. 该类共识机制的主要代表有工作量证明(proof of work, POW)[13]、权益证明(proof of stack, POS)[14]、授权股权证明(delegated proof of stack, DPOS)[15]等. POW共识算法是已知的第一个区块链共识算法,其主要思想是用算力换取记账权,矿工节点需要不断消耗算力计算符合出块条件的区块散列值,最先满足出块条件的节点获得记账权. POW共识算法具有节点随时加入、可拓展性强、验证简单、容易实现等优点,被广泛应用于电子货币系统中. 但POW算法因所有矿工节点需要不断消耗算力竞选记账权,导致大量资源被无意义消耗,同时存在系统吞吐量低,出块速度慢等问题. 为了避免共识算法造成的资源浪费,以太坊中提出了一种POS共识算法,该共识算法利用数字货币的价值,使用数字货币相关信息参与计算,持币数量和持币时间乘积越高的节点越容易获得记账权. POS共识算法减少了资源消耗,缩短了出块时延. 但POS共识算法容易遭受“无利害关系”攻击而产生分叉,且存在“中心化”问题,即获得记账权概率越高的节点会不断增加被选中的概率. 此外,POS算法要求所有节点都参与共识,对网络的吞吐量要求较高,为解决此问题,一种更加快速、安全且能源消耗小的DPOS算法被提出[16]. 在DPOS算法中,节点通过将持币委托给受托人,拥有持币最多的少部分受托人负责打包区块、维持系统的运转并获得相应的奖励. 然而DPOS并不是完全的去中心化,因此更适用于联盟链或私有链.

区块生成共识机制的主要代表为实用拜占庭容错(practical Byzantine fault tolerance, PBFT)[17]. PBFT共识算法是最早用来解决拜占庭将军问题[18]的分布式一致性算法之一. PBFT算法的最大作恶节点容错数量为$(m-1)/3$,其中m为系统节点总数. 算法的主要流程包括请求、预准备、准备、提交和回复5个阶段,每个阶段中各节点都要向其他节点广播验证结果,所以随着网络中节点个数的增加,共识时延也会随之增加,因此PBFT算法不适用于节点规模较大的系统. 每种共识算法都有其独特的优势及适合的应用场景,同时也存在一定的局限性,因此共识算法的研究与应用仍然充满挑战.

分布式数据维护场景广泛存在于实际生产和生活中,典型的包括联邦学习[19]、精准农业[20]、基于物联网的畜牧养殖等应用场景. 针对此类数据分布式生成或存储场景,采用传统中心化的数据更新或聚合的维护策略会面临数据维护不透明、数据易篡改、历史记录不可追溯等问题. 引入区块链技术可以有效解决上述问题,但传统POW类共识算法大量重复无意义计算,资源消耗较大. 针对此问题,本文提出了一种基于贡献度和数据有效性检验的共识机制,在确定记账权的同时进行了数据的实时性校验,避免了资源的无意义消耗. 同时,算法对于每个节点只需单次计算即可确定一次记账权,计算复杂度较小.

1　基于贡献度和数据有效性检验的共识机制

本文提出一种基于贡献度和数据有效性检验的共识机制,旨在解决分布式数据维护时数据不透明、易被篡改、历史记录不可追溯等问题. 该共识机制有以下设计目标:①保证记账权分配的公平性和可验证性并减少资源消耗;②节点可以从交易数据中筛除异常数据,将有效交易数据打包出块;③抵御可能存在的区块链网络攻击.

1.1　算法框架

基于贡献度和数据有效性检验的共识机制的核心设计思想为:

1) 在每个出块轮次中,各节点查询本地维护

的全局节点贡献度降序列表,获取当前轮次全网节点贡献度值,调用随机可验证的领导者选举算法,确定当前轮次记账权节点,获得记账权的节点作为当前轮次的领导者节点. 如图 1 中的步骤①所示.

2) 领导者节点整理交易数据,调用改进的密度峰值算法筛除交易数据中的无效交易,打包有效交易并广播区块. 如图 1 中的步骤②③④所示.

3) 区块链网络中的节点接收到当前轮次区块,在确认记账权节点身份后调用区块验证算法验证区块,将通过验证的区块加入区块链账本,同时更新贡献度列表. 如图 1 中的步骤⑤⑥所示.

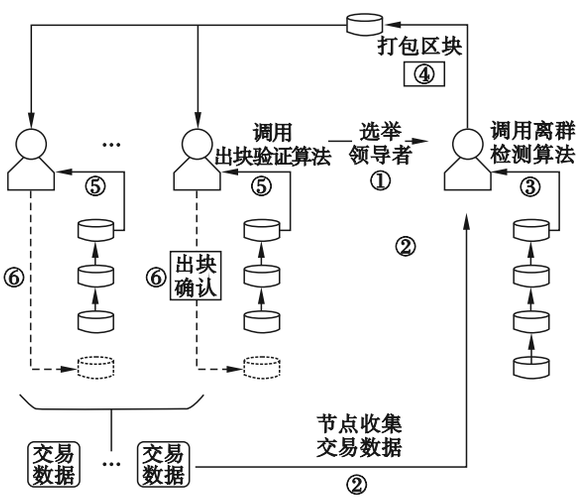

图 1 算法架构图

Fig. 1 Algorithm architecture diagram

1.2 实现方案

1.2.1 贡献度机制

贡献度定义为节点通过诚实维护数据和正确出块对区块链网络所做贡献程度的数值体现. 贡献度越高代表节点对区块链网络的贡献程度越大,该节点越希望区块链网络朝着更优的方向发展. 因此,贡献度可以作为网络中节点可信程度的度量. 本文将贡献度值作为记账权分配依据,尽可能将记账权分配给贡献度值高的节点,以保障系统的安全性. 网络中的所有节点同时维护一个全局节点列表,记录当前网络中的所有节点及其贡献度值. 对于每一个新加入的节点,初始贡献度由抵押机制确定,具体描述参见 2.3.1 节.

在算法中,增加贡献度的方式有两种:①领导者节点正确出块奖励;②数据更新维护奖励. 节点每次被选中为领导者并正确出块时可以获得 λ_1 原始贡献度值奖励,提供一条有效数据更新交易可以获得 λ_2 原始贡献度值奖励. 节点原始贡献度值更新如式(1)所示:

$$L_{t+1}=\alpha L_t+k_1\lambda_1+k_2\lambda_2. \tag{1}$$

其中:L_t 表示 t 时刻节点的原始贡献度值;α 表示遗忘系数,为一小于 1 的正数;k_1 表示节点正确出块的次数;k_2 表示有效数据更新交易数量. 为了使得贡献度有界,利用 Sigmoid 函数对其进行压缩,即有

$$C_{t+1}=\frac{1}{1+\mathrm{e}^{-L_{t+1}}}. \tag{2}$$

其中,C_{t+1} 表示节点在 t+1 时刻的贡献度值.

在领导者选举算法中需要对节点按照贡献度排序,当有不同节点具有相同的贡献度时,为了保证网络中所有节点在本地维护的贡献度降序列表的一致性,本文预设了一种贡献度比较算法,通过比较两节点公钥地址的散列值来确定两个节点在贡献度降序列表中的先后顺序,伪代码如算法 1 所示. 其中 hash( )表示散列函数.

算法 1 贡献度比较算法

conCompare(C_i, A_i, C_j, A_j)

输入:节点 i 的贡献度值 C_i,节点 i 的公钥地址 A_i,节点 j 的贡献度值 C_j,节点 j 的公钥地址 A_j;

输出:两个节点中最大贡献度值节点公钥地址 $A_{\max}$;

① if $C_i \neq C_j$ then

② $A_{\max} \leftarrow A_{\arg\max(C_i, C_j)}$

③ if $C_i = C_j$ then

④ $A_{\max} \leftarrow A_{\mathrm{argmax}(\mathrm{hash}(C_i),\mathrm{hash}(C_j))}$ /*当贡献度值相同时,比较两节点公钥地址的哈希值*/

1.2.2 随机可验证的领导者选举算法

记账权分配是共识机制设计的核心部分,每一类共识机制都具有其独有的记账权分配方式. 优秀的共识算法记账权分配过程必须具备公平性、随机性、可验证性等特性. 本文将获得记账权的节点称为领导者节点,从以上角度出发提出一种随机可验证的领导者选举算法. 其基本思路是将节点贡献度值映射为每一个共识轮次中该节点被选为领导者节点的概率,应用轮盘赌方式选出领导者节点. 该算法既保证了可信节点的高选中率,同时又保证了记账权分配的公平性、随机性和易验证性.

随机可验证的领导者选举算法伪代码如算法 2 所示,其主要步骤包括:

1) 各节点查询本地维护的全局节点贡献度降序列表,计算得到所有节点在当前共识轮次中被选为领导者节点的概率p_i.

2) 各节点根据概率数组计算所有节点在轮盘赌算法中的累计概率,构建累计概率数组Q.

3) 调用散列函数返回最新区块内容散列的散列值,根据最大散列值maxHash将其映射到[0,1]区间,得到累计概率数组的种子值.

4) 获得累计概率数组和种子值后,根据轮盘赌法则确定获得记账权的节点.

算法在利用轮盘赌方式选择领导者节点时,随机种子是通过对最新区块取双重散列函数得到的. 由于在未出块之前任何节点都无法正确预测未生成区块的散列值,而当区块生成后区块链网络中所有节点只需要简单地利用散列函数即可验证所生成的随机种子,因此该方式既保证了领导者选举的随机性,又保证了易验证性. 同时,轮盘赌选择方式在公平性的基础上保证了可信节点以较高概率被选中为领导者节点.

算法2　随机可验证领导者选举算法 ldS(hashVal,C)

输入:最新区块散列值hashVal,全局节点贡献度降序列表C;

输出:当前轮次领导者节点leader;

① for C中的每一个节点i

② $p_i \leftarrow \dfrac{C_i}{\sum_{j=1}^{n} C_j}$　/*根据节点贡献度值计算节点被选中领导者节点的概率P_i*/

③ for 每一个节点i

④ $Q_i \leftarrow \sum_{k=1}^{i} P_k$　/*根据P_i计算每个节点在轮盘赌算法中的累计概率Q_i*/

⑤ feed $\leftarrow$ hash(hashVal)/ max Hash　/* 计算轮盘赌算法的随机种子值*/

⑥ leader $\leftarrow \arg\max\{Q_i | Q_i < \text{feed}\}$　/* 根据随机种子feed选出当前轮次领导者节点leader */

1.2.3　基于密度峰值的数据有效性验证算法

分布式数据维护场景中的有效数据应该具有相同的变化趋势,分属于相应数据类簇中,因此基于聚类的离群点检测方法最符合分布式数据维护场景的离群数据点检测需求. 本文基于快速搜索密度峰值聚类算法[21]的局部密度和更高局部密度点最近距离,提出交易数据有效性评分的计算规则,进行离群点检测,实现交易数据的有效性验证.

在快速搜索密度峰值聚类算法中假设每个类簇中心都由局部密度较低的数据点包围,并且每个类簇中心与具有更高局部密度的点有较大距离. 该算法定义了“局部密度”以及“到更高局部密度点的最近距离”两个值,分别如式(3)和式(4)所示.

$$\rho_i = \sum_j c\left(d_{ij} - d_c\right), \tag{3}$$

$$\delta_i = \min_{j:\rho_j > \rho_i}\left(d_{ij}\right). \tag{4}$$

其中:$c(x)$为条件函数,当$x<0$时,$c(x)=1$,其他情况时,$c(x)=0$;d_{ij}为节点i与节点j之间的欧氏距离;d_c为预定义的截断距离. ρ_i相当于计算与点i距离小于截断距离的点的个数,称之为“局部密度”. δ_i表示所有局部密度大于节点i的点与节点i的最近距离,而对于密度最大的点,定义其值为

$$\delta_{\max} = \max_{ij}(d_{ij}). \tag{5}$$

聚类中心一般具有相对较大的“局部密度”和“到更高局部密度点最近距离”,而具有相对较小“局部密度”和相对较大“到更高局部密度点最近距离”的点常被看做“离群点”,即异常数据. 因此,本文利用“局部密度”和“到更高局部密度点最近距离”设计交易数据的有效性评分,如式(6)所示.

$$S_i = \left(\delta_{\max} - \delta_i + 1\right)^{\text{parameter}} \times \mathrm{e}^{\rho_i}. \tag{6}$$

式中,

$$\text{parameter} = \frac{\rho_i}{\dfrac{1}{k}\sum\limits_{j \in N_k(i)} \rho_j}. \tag{7}$$

其中,$N_k(i)$表示节点i的k近邻节点集. 可以看出,当节点具有相对较小“局部密度”和相对较大“到更高局部密度点最近距离”时,其有效性评分较低,说明该数据点越可能为离群数据. 因此只需要计算所有交易数据的有效性评分,将有效性评分小于给定阈值的交易数据点判定为离群点. 为了减小计算量,选择一个参考交易数据集,对所有交易数据的“局部密度”和“到更高局部密度点最近距离”提前进行离线计算,之后对新交易数据可以实现快速在线离群点检测,其伪代码如算法3所示.

系统将密度峰值离群点检测算法模型、参考交易数据集和有效性评分阈值保存在区块链的初始块中. 加入网络的节点可以通过其他节点获取历史区块,并从初始块中下载离群点检测算法模型、参考交易数据集和有效性评分阈值. 该算法可以实现快速地检测交易数据的有效性,一方面保证了系统的准确性和稳定性,另一方面增加

了共识机制的可拓展性,不同场景下可以使用不同的离群点检测算法维持系统的性能.

算法3 基于密度峰值离群点检测算法

predict(X,y)

输入:参考交易数据集X,新交易数据y;

输出:新交易数据有效性Validity;

%参考交易数据集X的离线计算

① 计算参考交易数据集X中所有数据点之间的欧氏距离$D=\{d_{ij}=\mathrm{dis}(x_i,x_j),x_i,x_j\in X\}$;

② 将D中元素按照升序排序,取r%位置的距离作为截断距离d_c;

③ 根据式(3)计算每个数据点的局部密度ρ;

④ 根据式(4)计算每个数据点到更高局部密度数据点的最近距离δ;

%新交易数据y的在线计算

⑤ 计算y到参考交易数据集X中所有数据点之间的欧氏距离$T=\{t_j=\mathrm{dis}(y,x_j),x_j\in X\}$;

⑥ 根据式(3)计算y的局部密度ρ_y;

⑦ 根据式(4)计算y到更高局部密度数据点的最近距离δ_y;

⑧ 根据式(6)计算y的有效性评分S_y;

⑨ 若$S_y<S_c$,Validity=True,否则Validity=False.

其中,dis()表示欧氏距离函数,S_c表示有效性评分阈值.

1.2.4 区块验证算法

领导者节点收集给定时间段的所有交易数据集合Y,调用密度峰值离群点检测算法计算集合Y的有效性标识:

$$\mathrm{Flag}=\left\{\mathrm{flag}_i=\mathrm{predict}\left(X,y_i\right),y_i\in Y\right\}. \quad (8)$$

之后将其与交易数据Y一起打包出块.

验证节点在接收到新区块后利用密度峰值离群点检测算法的可验证性,对领导者打包区块正确性进行验证,伪代码如算法4所示.

算法4 区块验证算法verify(Y,Flag)

输入:区块中所有交易构成的数据集合Y,区块中交易标志位数组Flag;

输出:区块验证结果res;

%验证节点计算集合Y的有效性标识

① Flag_val = { $\mathrm{flag\_val}_i=\mathrm{predict}(X,y_i),y_i\in Y$}

%校验Flag和Flag_val是否完全一致

② if $\mathrm{flag\_val}_i=\mathrm{flag}_i$ for all i, then

③ res←true /*区块通过验证*/

④ else

⑤ res←false /*拒绝出块*/

通过验证的区块加入到本地区块链中,同时更新节点贡献度列表.

1.3 区块设计

根据前文所述算法的实现方案,对区块结构进行设计. 区块链分布式账本的特征是将只读区块设计成链式结构,每一个区块都保存指向前一区块的指针,即前一区块内容的散列值. 此外,每个区块中还包含时间戳和由数据交易构建的默克尔树. 除了上述三个传统区块链分布式账本常用的结构之外,针对本文的具体方案进行了如下设计.

随机可验证的领导者选举算法是基于全局节点贡献度值的一种共识机制,由于贡献度是动态更新的变量并且与节点的初始贡献度值、发起有效交易次数和成为领导者正确出块的次数有关,将每一轮次的领导者节点公钥地址保存到当前块中;同时需要存储对最新区块取双重散列函数得到的随机种子以实现对领导者节点选取算法的验证. 最后,为了实现区块的验证,保存了当前打包交易的有效性标识. 区块的具体结构设计如图2所示. 区块链网络中的所有节点在接收到最新区块时,调用计算节点贡献度的智能合约,根据区块链账本,按照式(1)和式(2)进行计算,更新在本地维护的所有节点贡献度值列表. 当有新节点加入网络中时,其初始贡献度值通过构建相应交易的方式广播给网络中的所有节点,并被维护到所有节点的本地贡献度值列表中. 同时,新节点的加入将触发计算节点贡献度的智能合约,即根据区块链账本,按照式(1)和式(2)进行计算,构建所有节点的贡献度本地列表.

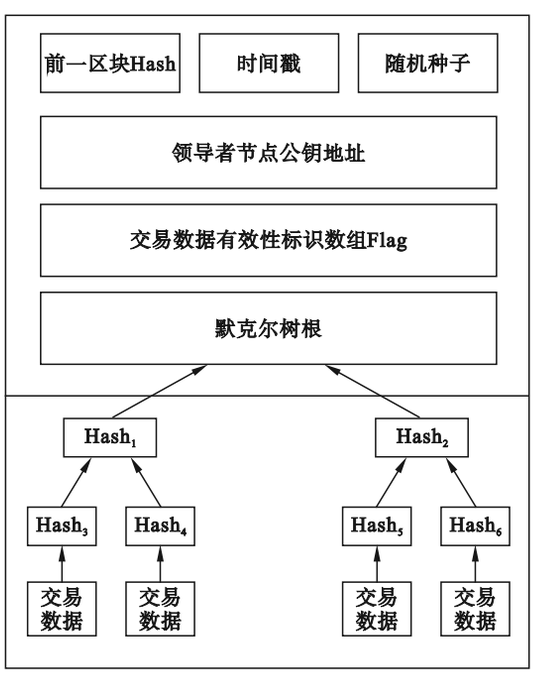

图2 区块结构设计

Fig. 2 Block structure design

2　安全性分析

2.1　攻击方式

2.1.1　女巫攻击

女巫攻击是指网络中的恶意节点通过创建不同的虚拟身份对区块链网络发起攻击从而达到获利的目的. 在本文的公有链网络中,恶意节点可能通过以下两种方式对网络发起女巫攻击:第一种方式为恶意节点不断伪造虚拟节点加入区块链网络,从而增加被选为领导者节点的概率;第二种方式为恶意节点冒充领导者节点并随意出块以达到干扰出块进程的目的.

2.1.2　区块截留攻击

区块截留攻击是指被选中的领导者节点没有按照预设规则按时出块,导致区块链网络中所有节点一直处于区块等待状态的攻击方式. 在本文的公有链网络中,恶意节点可能在被选为领导者节点后实施区块截留攻击或正常节点因网络和通信等原因宕机时遭到区块截留攻击.

2.2　系统健壮性

系统具备健壮性是指系统可以判断出规范要求以外的输入,并执行合理的处理方式. 在本文的公有链网络中节点可能因为设备或人为因素整理得到包含未知误差的交易数据,对系统的健壮性发起挑战.

2.3　安全策略

2.3.1　抵押机制

基于贡献度和数据有效性检验的共识机制将贡献度进行量化,根据贡献度值确定节点获得记账权的概率. 为了防止恶意节点通过伪造虚拟节点加入网络提升被选为领导者的概率,本文提出一种抵押机制. 新节点需要抵押资本获取加入网络的资格,新节点的初始贡献度值由抵押资本的数量决定,通过资本验证的节点才可以将其公钥地址和初始贡献度值广播全网,被其他节点加入到本地贡献度列表,获得参与共识的权利. 抵押机制可以有效防止恶意节点伪造虚拟节点实施女巫攻击. 节点贡献度初始化如式(9)所示.

$$\mathrm{CI}_{\mathrm{new}}=\frac{\sum_{i=1}^{n}\mathrm{CI}_i\times M_{\mathrm{new}}}{\sum_{j=1}^{n}M_j}. \tag{9}$$

其中:CI_i为节点i的初始贡献度;M_{new}为新节点抵押资产数量;M_j为节点j抵押资产数量.

2.3.2　引入非对称加密算法

为了防止恶意节点伪造身份干扰出块进程,引入非对称加密算法进行身份验证. 区块链网络中每一个节点都有一对公私秘钥,私钥用于节点签名,而公钥用来其他节点对私钥签名进行身份验证,确定节点身份. 如图3所示,在数字签名过程中,节点调用散列函数得到数据摘要并使用私钥对数据摘要加密,打包数据和加密的数据摘要后发送给其他节点. 其他节点接收到带有数字签名的数据后,使用签名节点的公钥对加密的数字摘要解密,比较解密得到的摘要和调用散列函数得到的数据摘要完成签名的验证. 由于私钥是私有的,任何节点无法根据公钥获取其他节点的私钥,所以恶意节点无法伪造身份欺骗正常节点.

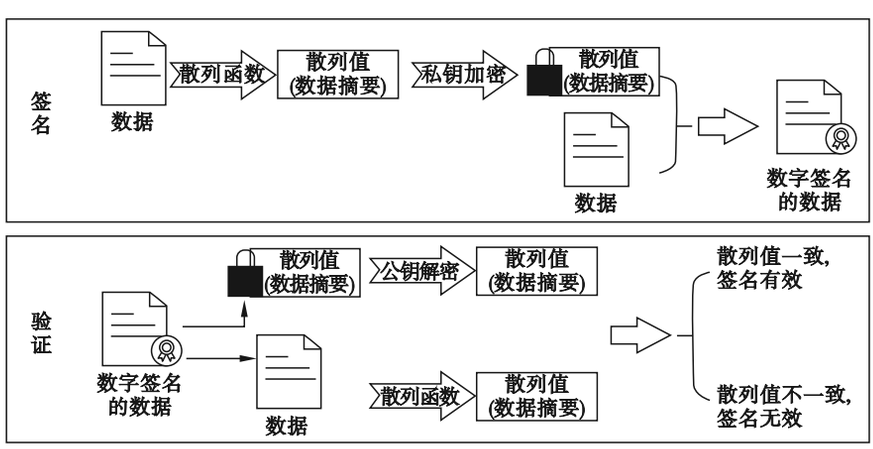

图3　非对称加密

Fig. 3　Asymmetric encryption

2.3.3　区块等待约束机制

为了避免遭受区块截留攻击造成所有节点处于出块等待状态,导致系统瘫痪的情况发生,本文提出区块等待约束机制,设置固定的区块等待时延阈值. 每一轮次的领导者选举结束后所有节点处于出块等待状态,节点启用监听并在区块

等待约束时间内不断计算区块截留信号位,区块截留信号位的计算公式如式(10)所示,当超过区块等待时延时,若节点还未接收到更新区块则默认区块被截留,所有节点将本轮次的领导者节点排除本轮,重新调用随机可验证领导者选举算法确定本轮领导者. 经过实验分析表明,加入区块等待时延后出块时间开销依然在可以接受的范围.

$$\text{signal}=\begin{cases}0, & 0\leqslant T_{\mathrm{r}}-T_{\mathrm{l\_S}}\leqslant T_{\mathrm{h}};\\ 1, & T_{\mathrm{c}}-T_{\mathrm{l\_S}}>T_{\mathrm{h}};\\ 2, & T_{\mathrm{r}}-T_{\mathrm{l\_S}}<0.\end{cases}\tag{10}$$

其中:signal表示当前轮次区块截留信号位,信号位为0时表示已接收当前轮次区块,信号位为1时表示遭受区块截留攻击,信号位为2时表示节点处于最新区块等待状态;T_{r}表示接收到最新区块的时间;$T_{\mathrm{l_s}}$表示当前轮次领导者确定的时间;T_{h}表示区块等待时延阈值;T_{c}表示当前时间.

3 实验分析

本文从3个方面进行实验分析:第1部分实验的目的是验证密度峰值离群点检测模型的性能,第2部分对共识算法的性能进行分析,第3部分通过实验论证本文提出的共识机制具备抵御安全性攻击的能力.

3.1 密度峰值离群点检测算法

3.1.1 实验数据

实验数据选用了分布式鹿场养殖数据,每个鹿场作为一个节点加入,每一条数据共有4种特征,分别为体重、体高、体长、体温,其中体重、体高、体长均为一段时间的变化量. 实验数据共有正常数据50 000条,异常数据10 000条. 列举4个节点部分数据如表1所示,其中标识为0的数据表示有效数据,标识为1的数据表示无效的异常数据.

表1 部分实验数据

Table 1 Part of experimental data

编号	体重/kg	体高/cm	体长/cm	体温/℃	标识
A01	4.098	0.906	1.051	36.9	0
A02	2.682	0.301	0.319	36.8	0
A03	3.682	0.821	0.811	36.5	0
B04	3.392	0.598	0.616	36.1	0
B05	3.639	0.746	0.796	36.8	0
B06	4.727	1.282	1.132	36.9	0
C07	2.786	0.347	0.362	36.5	0
C08	2.374	0.294	0.315	36.4	0
⋮	⋮	⋮	⋮	⋮	⋮
D01	5.388	0.430	0.883	36.7	1
D02	3.715	0.836	2.342	36.3	1

3.1.2 数据有效性验证参与共识的可行性分析

利用密度峰值离群点检测算法在分布式鹿场养殖数据集上进行数据有效性验证,查准率及召回率结果如图4所示. 有效数据和离群数据的识别精度分别达到了99.1%和95.9%,而它们的召回率分别为98.9%和96.1%. 实验结果表明,引入密度峰值离群点检测算法满足交易数据有效性验证任务的需求,其高查准率和召回率为算法的稳定性提供保障.

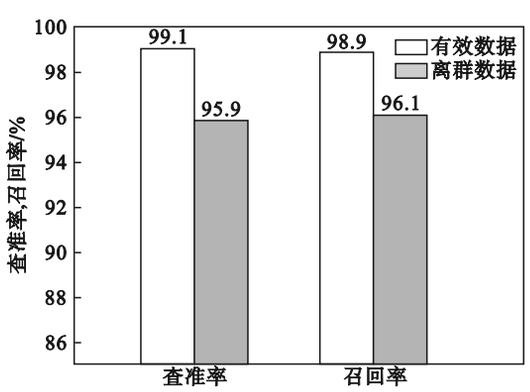

图4 离群点检测算法性能

Fig. 4 Outlier detection algorithm performance

密度峰值离群点检测算法的时间开销与计算的交易数据量有关,该算法的时间开销也影响着共识算法出块的时间开销. 本文进行了交易数据量对密度峰值离群点检测时间开销影响的实验,结果如图5所示. 从图中可以看出,随着数据量的不断增大,密度峰值离群点检测时延与数据量成正相关,可以根据时间开销需求设置区块链交易数据量阈值.

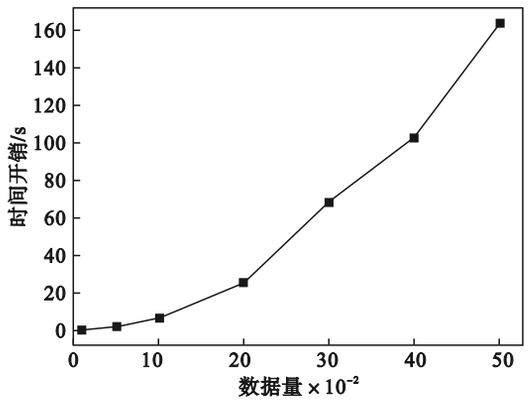

图5　离群检测时间开销与交易数据量的关系

Fig. 5　Relationship between the time cost of outlier detection and the amount of transaction data

3.2　共识机制性能分析

3.2.1　贡献度值对领导者选举算法的影响

为了分析节点贡献度值与其被选中领导者概率之间的关系,设置10个节点,进行4轮实验,每轮进行10 000次领导者选举. 在所有轮次中节点2到节点10的贡献度均设置为5并保持不变,而节点1在4个轮次中贡献度分别设置为10,15,20,25,实验结果如图6所示. 可以看出,随着节点1贡献度值不断增加,其被选中领导者节点的概率不断升高,其余节点被选中的概率则不断降低且几乎相同. 实验结果表明,随机可验证领导者选举算法完全具有随机性,并且节点被选中的概率与贡献度值正相关,可保证每一次共识过程中诚实节点以高概率获得记账权.

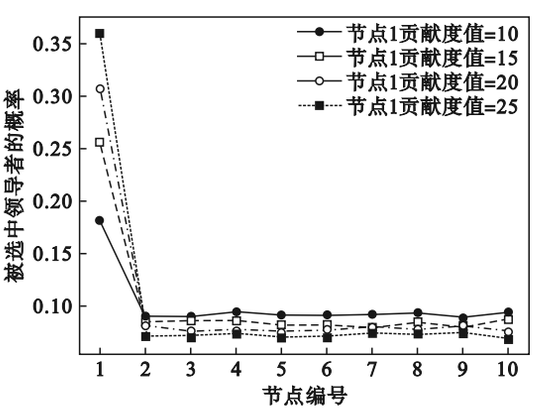

图6　节点贡献度值与被选中领导者概率的关系

Fig. 6　Relationship between the node contribution value and the probability of the selected leader

3.2.2　出块时间开销

本文提出的基于贡献度和数据有效性检验的共识机制出块时间开销包括领导者选举时间开销、交易数据离群点检测时间开销、打包区块时间开销、区块确认时间开销. 其中,随机可验证领导者选举时间开销与区块链网络中节点数量有关,随机可验证的领导者选举算法每个共识轮次记账权分配的时间开销和节点数量关系如图7所示. 领导者选举时间开销随节点数量增加而增加,节点数量为50 000时的时间开销仅为43.7 ms,因此领导者选举时间开销对出块时间开销的影响可以忽略不计.

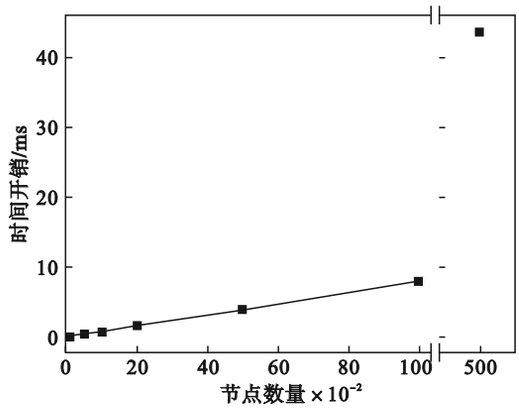

图7　领导者选举时间开销与节点数量关系

Fig. 7　Relationship between the time cost of leader election and the number of nodes

交易离群数据点检测时间开销、打包区块时间开销和区块确认时间开销均与区块打包的交易数据数量有关. 基于贡献度和数据有效性检验的共识机制出块时间开销与交易数据数量的关系如图8所示. 在实验中节点数量设置为5 000. 实验结果表明,该共识算法有较高的吞吐量和较低的时间开销,在数据量为5 000时,出块时间开销为5.3 min. 为了平衡单位时间内的交易确认数量和单次共识更新的交易量,在共识过程中设置区块打包交易数据上限阈值为4 000,由图8可以看出,数据量为4 000时出块时间开销约为3.3 min,换算后可知数据交易吞吐量约为20条/s.

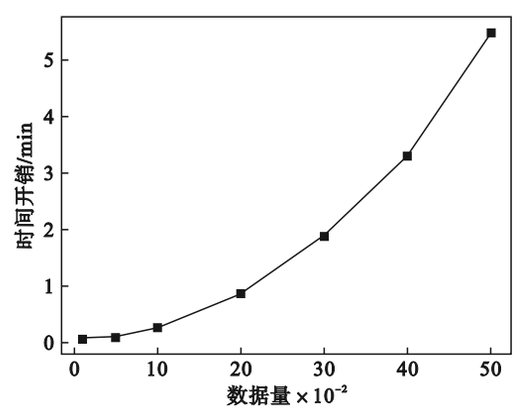

图8　出块时间开销

Fig. 8　Block time overhead

3.3 安全性分析

3.3.1 系统健壮性分析

本文提出的基于贡献度和数据有效性检验的共识机制中,系统健壮性保证区块中数据交易的有效性. 因此影响本系统健壮性的因素为具有未知误差数据的节点是否会将误差数据作为有效数据打包出块,以上情况发生必须满足两个条件:①节点被选为领导者节点;②误差交易数据被离群点检测算法检测为有效数据. 因此节点将误差数据作为有效数据打包出块的概率如式(11)所示.

$$\varepsilon=(1-P)\frac{C_i}{\sum_{j=1}^{n}C_j}. \tag{11}$$

其中:P表示离群点检测算法查准率,由实验得到离群点检测算法的查准率为99%;C_i表示恶意节点i贡献度值;C_j表示节点j贡献度值;n表示节点总数.

由式(11)可以看出,节点将误差数据作为有效数据打包出块的概率与节点数量和节点的贡献度有关.

本文对节点将误差数据作为有效数据打包出块的概率与节点数量的关系进行实验分析,从而验证该系统的健壮性. 实验中设置每个节点的贡献度值均为10. 实验结果如图9所示,节点将误差数据作为有效数据打包出块的概率随着节点数量增加不断降低,当节点数量为5时,概率仅为0.2%,实验结果表明,基于贡献度和数据有效性检验的共识机制具有良好的健壮性.

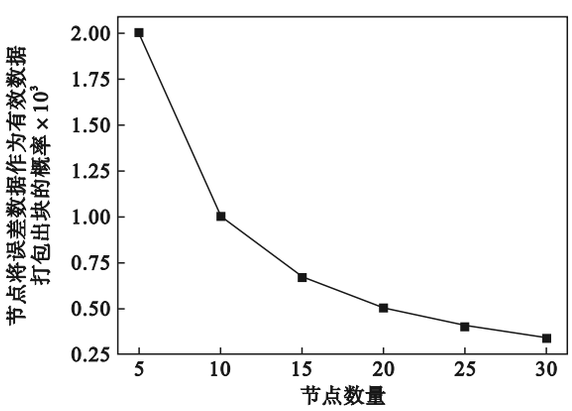

图9 健壮性分析

Fig. 9 Robustness analysis

3.3.2 区块截留

为了验证所提出的共识机制在受到区块截留攻击时的灾备能力,对遭受区块截留攻击时的出块时间开销和未遭受区块截留攻击时的正常出块时间进行了对比实验,结果如图10所示. 由实验结果可以看出当遭受区块截留攻击时,出块的时间开销与区块中的交易数据数量正相关,其曲线与未遭受区块截留攻击时的正常时间开销曲线大致平行,相同数据量下遭受区块截留攻击时出块时间比未遭受区块截留攻击时的正常出块时间平均增加1.2 min,当数据量为4 000时,遭受区块截留攻击时的时间开销为4.5 min. 实验结果表明,该共识机制在遭遇区块截留攻击时仍具有较低出块时间开销,具有应对区块截留攻击的灾备能力.

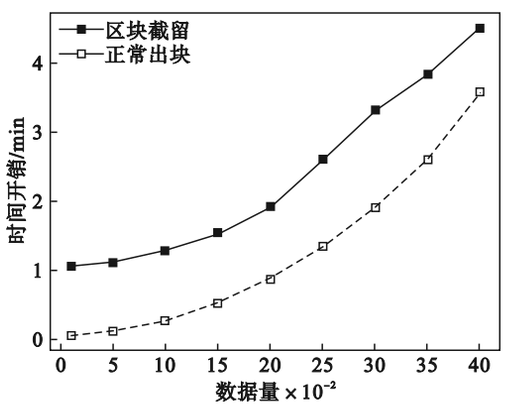

图10 区块截留重新出块时间开销

Fig. 10 Block interception reblock time overhead

3.4 与主流共识算法对比

PoW和PoS是适用于大规模节点数量的公有链主流共识算法,根据官方文档[1,22]所提供的数据,选取在比特币系统中应用的PoW共识算法和Peercoin中应用的PoS共识算法,在本地相同实验环境中与本文提出的基于贡献度和数据有效性检验的共识机制在出块时间和交易吞吐量两个指标上进行了对比,结果分别如图11和图12所示.

Target值是影响PoW和PoS共识速度的主

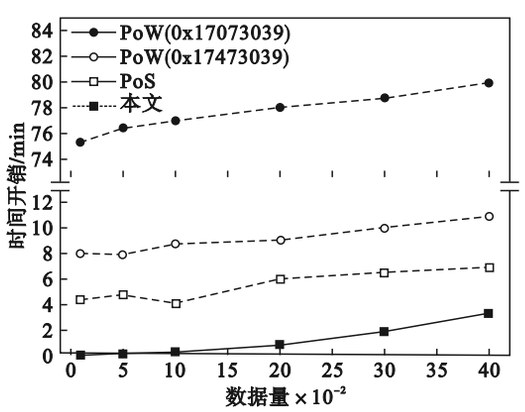

图11 出块时间对比

Fig. 11 Block time comparison

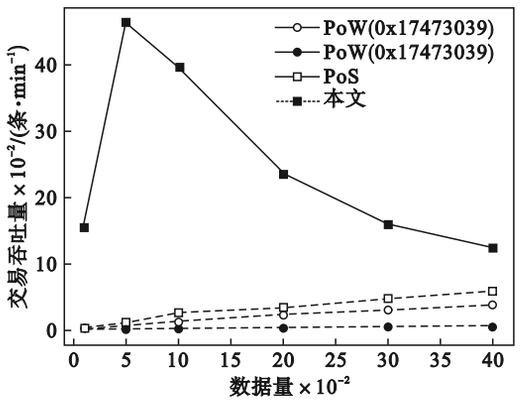

图12　交易吞吐量对比

Fig. 12　Transaction throughput comparison

要因素,除了target值之外,币龄是影响PoS共识速度的另一个主要因素. 针对PoW,首先选取当前比特币系统中的target值0x17073039进行实验,发现其在出块时间远大于本文方法,在吞吐量上则远低于本文方法. 当数据量为1 000时,PoW的出块时间约为77 min,大约是本文方法(0. 3 min)的256倍;PoW的交易吞吐量为13条/min,大约是本文方法(3 900条/min)的1/300. 因此将target值提高到0x17473039降低挖块难度,也进行了对比. 同时将PoS的target值也取为0x17473039,并将币龄均设置为2. 可以看出此时PoS的出块时间小于PoW,它们都明显高于本文方法. 当数据量为1 000时,PoS的出块时间为4 min,为PoW(8. 7 min)的1/2,大约是本文方法的13倍. 在交易吞吐量方面,PoS为250条/min,约为PoW(115条/min)的2. 2倍,但仅为本文方法的1/15左右.

此外,在交易确认时间、资源消耗、时间复杂度、是否会分叉4个指标上将PoW和PoS和本文方法进行了对比,结果如表2所示. 其中,交易确认时间为作者在0x17473039条件下实验所得,其他3项指标的结果由文献[11]给出. 由此可以清晰看出本文方法在交易吞吐量、出块时间、交易确认时间、资源消耗、时间复杂度等方面都具有优势,并且解决了区块链产生分叉的问题.

表2　与主流共识算法对比

Table 2　Comparison with mainstream consensus algorithms

共识机制	交易确认时间/min	资源消耗	时间复杂度	是否会分叉
PoW	54	高	$O(n)$	是
PoS	33	中	$O(n)$	是
本文	3	低	$O(1)$	否

4　结　　论

本文提出了一种基于贡献度和数据有效性检验的共识机制,有效解决了分布式数据维护场景下数据易被篡改、历史数据不可追溯等传统数据库存在的问题. 基于密度峰值离群点检测算法的引用满足了分布式数据维护对数据有效性验证的需求,同时提升了该算法的可拓展性. 定义贡献度机制和提出随机可验证领导者算法保证了记账权分配的随机性和公平性,并且避免了无意义计算带来的资源消耗. 实验结果表明,所提出的基于贡献度和数据有效性检验的共识机制高效可行,具有广阔的应用前景.

参考文献:

[1] Nakamoto S. Bitcoin: a peer-to-peer electronic cash system [EB/OL]. [2022-04-22]. https://bitcoin.org/bitcoin.pdf.

[2] Ali O, Ally M, Clutterbuck P, et al. The state of play of blockchain technology in the financial services sector: a systematic literature review [J]. *International Journal of Information Management*, 2020, 54: 102199.

[3] Alonso S G, Arambarri J, Coronado M L, et al. Proposing new blockchain challenges in eHealth [J]. *Journal of Medical Systems*, 2019, 43(3): 64.

[4] Novo O. Blockchain meets IoT: an architecture for scalable access management in IoT [J]. *IEEE Internet of Things Journal*, 2018, 5(2): 1184-1195.

[5] Lu Z J, Wang Q, Qu G, et al. BARS: a blockchain-based anonymous reputation system for trust management in VANETs [C]//IEEE Trustcom BigDataSE ISPA. New York, 2018: 98-103.

[6] Liu D X, Alahmadi A, Ni J B, et al. Anonymous reputation system for IIoT-enabled retail marketing Atop PoS blockchain [J]. *IEEE Transactions on Industrial Informatics*, 2019, 15(6): 3527-3537.

[7] Meshcheryakov Y, Melman A, Evsutin O, et al. On performance of PBFT blockchain consensus algorithm for IoT-applications with constrained devices [J]. *IEEE Access*, 2021, 9: 80559-80570.

[8] Lin I C, Liao T C. A survey of blockchain security issues and challenges [J]. *International Journal of Network Security*, 2017, 19(5): 653-659.

[9] Zheng Z B, Xie S, et al. Blockchain challenges and opportunities: a survey [J]. *International Journal of Web and Grid Services*, 2018, 14(4): 352-375.

[10] Yi H B, Li Y P, Wang M, et al. An efficient blockchain consensus algorithm based on post-quantum threshold signature [J]. *Big Data Research*, 2021, 26: 100268.

[11] 王缵,田有亮,岳朝跃,等. 基于门限密码方案的共识机制 [J]. 计算机研究与发展, 2019, 56(4): 692-707.
(Wang Zan, Tian You-liang, Yue Chao-yue, et al. Consensus mechanism based on threshold cryptography scheme [J]. *Journal of Computer Research and Development*, 2019, 56(4): 692-707.)

[12] 蔡晓晴,邓尧,张亮,等. 区块链原理及其核心技术 [J]. 计算机学报, 2021, 44(1): 84-131.

(下转第178页)