一种基于积分制的改进实用拜占庭容错算法

沈 瑞,李玲娟
(南京邮电大学 计算机学院,江苏 南京 210023)

摘 要:区块链技术是一种融合分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。共识算法是区块链技术中的核心部分之一。该文针对实用拜占庭容错算法(PBFT)存在的可参与节点较少,主节点选举随意,以及节点参与积极性较低的问题,提出一种基于积分制改进的实用拜占庭算法(P-PBFT)。引入委任权益证明算法思想,给每个节点设置积分,通过积分选举出参与共识过程的委员会节点;其次,在主节点出现问题需切换视图的时候,按照积分来切换主节点。最后设置一个时间周期来减少共识节点的积分,避免过度中心化,同时也达到激励节点的效果。通过搭建基于该方案的区块链测试系统并进行实验,证明了 P-PBFT 算法够有效地提高参与节点的数量和吞吐量,具有较好的实用性。

关键词:区块链;共识机制;实用拜占庭容错;委任权益证明;积分制
中图分类号:TP309 **文献标识码**:A **文章编号**:1673-629X(2021)06-0059-06
doi:10.3969/j.issn.1673-629X.2021.06.011

An Improved PBFT Algorithm Based on Point System

SHEN Rui,LI Ling-juan
(School of Computer Science,Nanjing University of Posts and Telecommunications,Nanjing 210023,China)

Abstract:Blockchain technology is a new application mode integrating distributed data storage,point-to-point transmission,consensus mechanism,encryption algorithm and other computer technologies. Consensus algorithm is one of the core parts of blockchain technology. Aiming at the problems existing in the practical Byzantine fault-tolerant algorithm (PBFT),such as fewer nodes to participate in,random election of main nodes,and low enthusiasm of nodes to participate,we propose an improved PBFT algorithm based on the point system,named P-PBFT. Firstly,the idea of delegated proof of stake algorithm is introduced. Each node is set with point,and the committee nodes participating in the consensus process are selected by point. Secondly,when the main node has problems and needs to switch views,it is switched according to point. Finally,set a time period to reduce the integration of consensus nodes,avoiding over centralization,and also achieve the effect of stimulating nodes. By building the blockchain test system based on this scheme and carry out experiments,it is proved that the P-PBFT algorithm can effectively improve the number of participating nodes and throughput with certain practicability.

Key words:blockchain;consensus mechanism;practical Byzantine fault tolerance;delegated proof of stake;point system

0 引 言

比特币诞生于 2008 年,由一个化名为中本聪的学者在论文中首次提出[1],其背后的区块链技术得到越来越多的重视与发展。区块链是一种计算机技术在价值互联网时代的创新应用模式,是数据库、密码学、网络技术等多种技术整合集成的结果,具有去中心化、去信任化、集体维护、不可篡改等特点[2]。从数据的角度,可以把区块链看成由一个个区块组成,区块记录着经过验证的、区块创建过程中发生的所有交易记录。近年来,区块链技术的价值不断被挖掘,在金融、身份验证、社交通讯等多个领域都有着广泛的应用,并得到了许多国家政府的高度重视。

共识算法是区块链技术的核心,用于解决在去中心化的分布式互连网络中,如何判断区块数据正确性和所有权,以使所有的节点达成共识的问题[3]。由于应用场景的不同,共识算法的侧重点也不同。公有链大部分采用 PoW 算法[4]、PoS 算法[5]、DPoS 算法[6]和它们的变形算法,而目前落地的实际应用大多数基于

收稿日期:2020-07-14 **修回日期**:2020-11-18
基金项目:国家自然科学基金(61572260,61872196)
作者简介:沈 瑞(1995-),男,硕士研究生,CCF 会员(C3811G),研究方向为区块链;李玲娟,教授,CCF 会员(E200015276M),研究方向为数据挖掘、分布式计算。

联盟链运行,其采用的共识算法主要是基于消息传递,主流算法有 PBFT 算法[7]、Paxos 算法[8] 和 Raft 算法[9]。

PBFT 算法在区块链中得到广泛应用,但也存在一些不足,不能完全适应所有的应用场景。为了解决 PBFT 算法通信开销大、节点参与共识的积极性不够高等问题,该文提出一种基于积分制的改进的实用拜占庭算法(improved practical Byzantine algorithm based on point system,P-PBFT)。

1 相关知识

1.1 区块链的类型

按照应用场景和设计体系的不同,区块链可分为公有链、联盟链、私有链[10]。

(1)公有链。

公有链所有节点都对外开放,每个人都可以从公有链中读取数据,发送交易。在公有链上,每个节点都可以自由加入或者退出,网络运行时以扁平、无分层的对等网络拓扑结构相互连通,不存在任何中心化的服务节点,是一种完全去中心化的区块链[11]。访问门槛低,数据公开透明且无法篡改、匿名性等是其主要特点,比特币就是公有链的典型代表。

(2)私有链。

私有链是指区块链的开发与维护由一个组织统一管理,各个节点的读取权限与写入权限由该组织决定,不对公共网络开放的区块链[12]。私有链虽然节点权限有所限制,但区块链网络仍然运行在多个节点,其数据的安全性依旧会得到一定的保证。与公有链相比,私有链具有处理交易速度快、交易成本低、隐私保护好等特点。

(3)联盟链。

联盟链是一种介于公有链和私有链之间的区块链,通常应用于不同的企业或组织之间。链上节点都有着相对应的实体,不同的实体组成联盟,节点的加入需要联盟授权,参与者共同维护区块链的运行。联盟链的读取权限对所有节点开放,写入和验证权限则需要联盟内部决定,属于部分去中心化区块链。目前已有很多联盟链,比较知名的有超级账本(hyperledger)项目[13]。

1.2 共识算法

共识算法起源于分布式系统领域,传统共识算法一般面向分布式数据库操作。而在区块链环境下,更多的是面对拜占庭容错问题[14],传统共识算法不能很好地解决这类问题,因此一系列新的共识算法被提出。常用共识算法如下:

(1)工作量证明(proof of work,PoW)。

PoW 的核心思想是通过分布式节点的算力竞争来保证数据的一致性和共识的安全性。而在区块链中,比特币系统则是这一算法的最早实践者,比特币系统的各节点通过计算一个求解复杂但是验证容易的 SHA256 数学难题来竞争记账权,最快解决该难题的节点将获得下一区块的记账权和系统自动生成的比特币奖励。

PoW 算法虽有效地保证了区块链网络的安全性和去中心化性,但其缺点也十分显著。PoW 算法需要节点进行大量的计算,但这种计算不具有现实意义,只会带来大量的电力资源消耗,且需要的交易确认时间过长,不适合一般的商业应用。

(2)权益证明(proof of stake,PoS)。

权益证明是为了弥补工作量证明的一些不足而诞生的,在权益证明机制中,记账权的归属不再是算力最高的节点,而是具有最高权益的节点。权益体现在节点对于虚拟货币的所有权,在最早的 PoS 应用 Peercoin 中,权益被称为币龄。一个节点的币龄越长,其在区块链系统中的权力越大,挖矿的难度越低,所获得奖励也越多。

与 PoW 算法相比,PoS 算法共识过程主要依靠系统内部的权益,而不需要消耗太多外部算力和资源,因此可以有效地解决 PoW 中算力浪费的问题,并且能够在一定程度上缩短达成共识的时间,提升系统运行性能。

(3)委任权益证明(delegated proof of stake,DPoS)。

DPoS 算法由比特股项目提出,是 PoS 算法的一种演化版本。DPoS 算法采用了类似董事会投票的机制,系统中每个节点都是股东,权益相当于选举票,每个节点都可以把自己的选举票投给信任的代表。最后得票最高的一部分节点成为董事会成员,按照既定的时间表轮流对交易进行打包结算、并且生产新区块。相比于 PoS 算法,DPoS 减少了参与验证区块的节点数量,区块可以得到更快的确认,区块链系统的性能得到了进一步的提升。

(4)实用拜占庭容错(practical Byzantine fault tolerance,PBFT)。

拜占庭容错算法 BFT 最早由 Pease 和 Lamport 在 20 世纪 80 年代提出,不同于以上几种共识算法,BFT 类协议是依靠节点之间相互传递消息来对提案达成确定性共识结果,因此早期的拜占庭系统需要指数级的操作,所以未能得到实际应用。直到 1999 年 Miguel Castro 和 Barbara Liskov 提出了 PBFT(实用拜占庭容错)算法,解决了原始 BFT 算法的信息传输复杂度太高的问题,由此实用拜占庭容错算法在实际系统中变

得可行[15]。而在区块链环境下,实用拜占庭容错算法多使用于联盟链中。

1.3　P2P 网络

区块链与 P2P 的出发点都是去中心化。

P2P 网络即对等网络,是在同等地位的节点之间分配计算任务与网络负载的分布式网络架构[16]。P2P 网络模型与客户端/服务器模型不同,P2P 网络中没有客户端或服务器的概念,不存在中心节点,只存在对等的同级节点。每个节点既寻求服务,同时也提供服务。P2P 网络中的节点没有数量、范围、时间或空间上的限制,每个节点都可以自由地加入或退出 P2P 网络。

P2P 网络具有去中心化、可扩展性、健壮性、高性价比、隐私保护和负载均衡的特点[17]。在 P2P 网络中,节点不需要服务器即可提供资源和服务,避免了中心服务器的瓶颈,且起到了负载均衡的效果。

从技术上来讲,区块链就是应用 P2P 的网络架构,通过密码学来保证数据的安全,通过共识算法来保证数据的一致性。P2P 一般存在 4 种网络模型,分别是:集中式、纯分布式、混合式和结构化模型。区块链应用依据自身的实际情况选择不同的 P2P 网络模型,比特币采用的是混合式网络模型,而以太坊采用的则是结构化网络模型。

1.4　Merkle 树

Merkle 树属于区块链数据层,是区块链中重要的数据结构,其基本结构如图 1 所示。

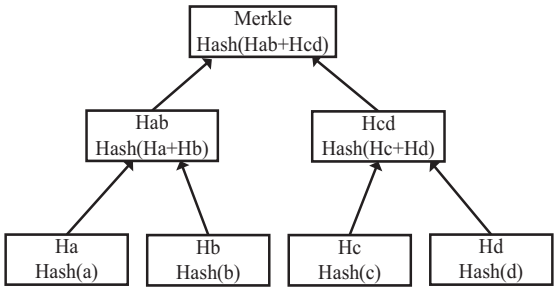

图 1　Merkle 树

该树的每个叶子节点值是对应数据的哈希值,非叶子节点是它的两个子节点合在一起的哈希值,依次叠加,直到计算出整棵树的根节点,最后生成 Merkle 根。Merkle 根是由所有叶子节点值得到的,因此只要验证 Merkle 根是否相等,就可以知道叶子节点的数据是否有改动。在分布式系统中,Merkle 树可以快速验证在传输过程中数据否发生变化,大大降低了计算复杂度。

1.5　加密技术

加密算法是区块链中不可缺少的一个环节,在区块链中所涉及的加密技术主要包括非对称加密、哈希算法和数字签名。

非对称加密算法是指在对数据进行加密和解密过程中使用不同密钥的一种加密算法。非对称加密过程中使用的密钥成对产生,其中公开的密钥叫公钥,任何人都可以获取,非公开的密钥叫私钥,对外是保密的。公钥与私钥都可以用来加解密,如果用私钥对信息进行加密,只能用公钥解密信息。如果用公钥对信息进行加密,只能用私钥解密信息。与对称加密相比,非对称加密无需交换密钥且算法强度高,所以具有更高的安全性,但是其加解密过程复杂度高且耗费时间较长,一般只适合加密少量数据,适用于数字签名、登录验证等场景。常见的非对称加密算法有 ECC(椭圆曲线加密)算法和 RSA 加密算法等。相比 RSA,ECC 优势是可以使用更短的密钥来实现与 RSA 相当或更高的安全性。

哈希函数也称散列函数,是一种单向映射函数。哈希函数将输入数据通过哈希算法生成特定长度的值,该值就称之为哈希值。哈希函数具有单向性、不可逆等特征,逆向求解哈希函数十分困难,几乎不能通过现有的哈希值反推出原文,因此可以有效保证信息的安全性。通过散列算法运算求得的哈希值具有固定长度,且哈希值远远小于输入长度,压缩性保证了任意长度消息压缩映射得到确定长度散列值。哈希函数具有高度灵敏性,如果输入的数据发生字节变化,那通过哈希运算得到的哈希值可能完全不同。哈希函数在区块链中发挥了极其重要的作用,可以进行数据验证、数字签名,保证了区块链系统中数据的安全性和完整性。

数字签名是数字摘要技术和非对称加密技术相结合的应用,为数字信息的完整性和发送者身份的真实性提供了强有力的保障。数字签名的流程是发送方将自己要传输的消息进行哈希,得到摘要,再用私钥将哈希值进行加密,最终得到加密数据。发送方将数字信息原文、加密后的数字摘要和公钥一同发给接收方。同时,接收方会用相同的散列函数生成数字信息的数字摘要。然后接收方使用发送方的公钥对消息和消息摘要进行解密,得到数字摘要。如果发送方的摘要和接收方的摘要相同,则即可证明数字信息的完整性和发送方身份的真实性。如果不同,则说明数字信息已经被篡改。

2　基于积分制的改进实用拜占庭容错算法 P-PBFT 的设计与分析

2.1　PBFT 算法分析

PBFT 算法每次共识发生在一个视图(view)中,视图是连续编号的整数,每个视图对应一个主节点,其余都是备份节点。在总节点数为 n 的系统中,PBFT 算法所能容忍的错误节点数 f 最大为 $(n-1)/3$。

(1)PBFT 算法流程。

PBFT 算法通过五个阶段达成共识,如图 2 所示。

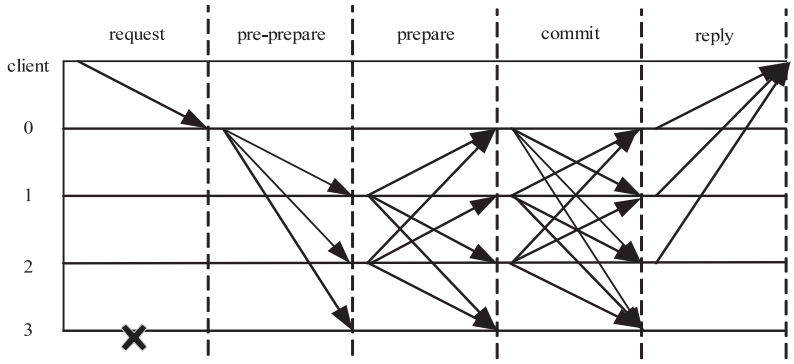

图 2 PBFT 算法共识流程

各阶段的具体流程如下:

请求阶段:客户端节点向主节点发送请求 m。

预准备阶段:主节点接收到来自客户端的请求后,给该请求分配一个序列号 n,生成一个预准备消息,预准备消息的格式为<PRE-PREPARE,v,n,d,m>,这里的 v 是视图编号,d 是客户端发送的请求 m 的摘要。然后向所有备份节点发送预准备消息。

准备阶段:备份节点接收到主节点的消息并对预准备消息进行验证,如果验证通过,生成准备消息<PREPARE,v,n,d,i>并向其他节点广播,同时将预准备和准备消息写入消息日志,i 表示节点编号。

确认阶段:节点收到来自其他备份节点的准备消息,对消息内容进行验证,若验证通过则向所有节点发送确认消息。

回复阶段:当节点完成确认阶段后,向客户端发送回复消息,当客户端接收到 f+1 个不同的节点发来的相同消息时,回复阶段完成,共识流程结束。

(2)视图切换。

PBFT 在视图中执行共识流程时,如果主节点发生宕机或者成为恶意节点,导致共识流程无法进行的时候,系统会运行视图变更协议,根据 p = v mod R 的规则重新选举主节点,其中 v 是视图编号,R 是节点个数。

(3)PBFT 算法的不足。

PBFT 算法在节点数量较多的情况下,通信开销很大,无法满足实际系统的需要。另外,主节点选举随意,在实际应用中无法起到激励节点的作用,且增大了主节点的作恶几率,降低了系统运转的可持续性和安全性。

2.2 P-PBFT 算法设计思想

针对 PBFT 算法的不足,结合实际区块链系统的应用情况,该文提出的 P-PBFT 算法对 PBFT 做了以下几点改进:

(1)结合 DPOS 思想,成立节点委员会,委员会里的节点数目根据实际需要设置,只有委员会里的节点才可以参与共识和竞争主节点,其余的节点只进行投票与结果保存,可以降低系统的通信开销,提升系统性能。

(2)引入积分制的概念,给每个节点设置积分,节点的初始积分可根据实际应用的需要来设置。节点使用积分进行投票,选举参与共识流程的节点。当主节点发生错误从而触发视图切换协议的时候,扣除当前主节点的 5% 积分作为惩罚,然后根据积分的排名顺次选举主节点,这样可以保证主节点的可靠性,提高节点竞争主节点的积极性。

(3)设置积分衰减周期 T,按 70% 的比例减少节点的积分,防止系统过度中心化。

2.3 P-PBFT 算法流程

P-PBFT 算法流程如图 3 所示。

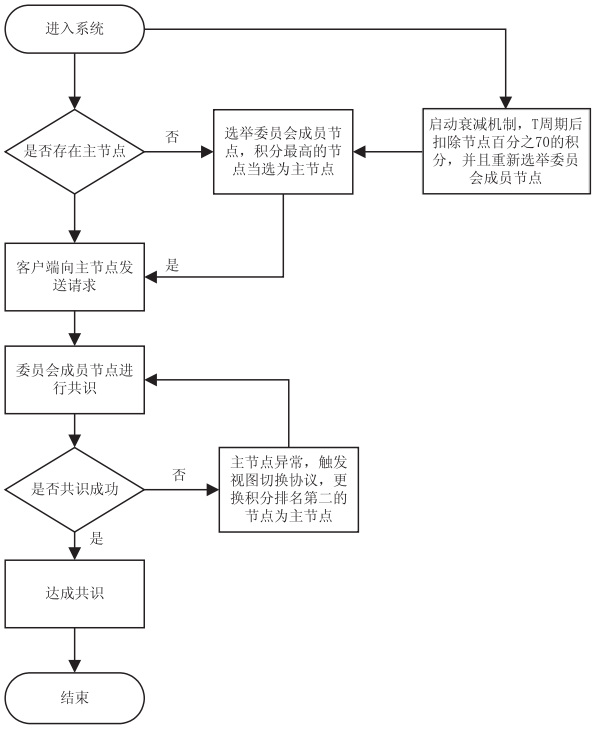

图 3 P-PBFT 算法流程

具体步骤如下:

第一步:客户端发送请求,系统进行响应,如果系

统存在主节点,直接进入第二步。系统不存在主节点的时候,所有节点投票选举出委员会节点和主节点。

第二步:委员会里的节点走共识流程,共识成功,主节点增加 5 积分,委员会节点增加 3 积分,其余节点增加 1 积分。如果主节点发生错误从而触发视图切换,则进行第三步。

第三步:扣除当前主节点的 5% 积分,积分排名第二的节点当选为主节点。

此外,每 T 个周期后,系统进行积分衰减,每个节点减少 70% 的积分,并且重新选举委员会成员。

2.4　P-PBFT 算法通信开销分析

设定系统的节点个数为 N,对于 PBFT 算法,参与共识的节点个数即为 N,三阶段共识的通信开销为 $2N(N-1)$。对于 P-PBFT 算法,设定委员会节点个数为 M,则三阶段共识的通信开销为 $2M(M-1)$,$M \leqslant N$,且随着 M/N 的减小,通信开销的降低会更加明显。

3　仿真实验及结果分析

该文基于 Go 语言实现了一个小型区块链系统,分别运行 PBFT 算法和 P-PBFT 算法,(每秒完成的交易数量,单位为个/秒)进行对比,以检验算法的总体效果。

实验环境是:操作系统为 Windows10,CPU 为 Intel(R) Core(TM) i5 - 6300U 2.30 GHz,内存为 12 GB。算法实现语言为 Go,测试工具为 Apache-JMeter。

(1)固定数量节点在不同时刻的吞吐量测试。

测试基于本地 10 个服务端节点,其中委员会节点数量设置为 4 个,另开一个客户端节点,负责向服务端发送需要共识的交易。开启 Apache-JMeter 进行压力测试,配置 50 个线程,创建 Http 请求,共识完成则记录一次请求成功,结果如图 4 所示,横坐标为时间,纵坐标为吞吐量。

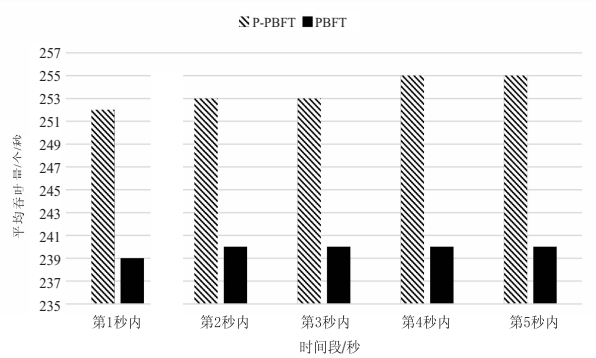

图 4　固定数量节点在不同时刻的吞吐量对比

可以看出,不同时刻吞吐量略有不同,后续测试将观察平均吞吐量。

(2)不同节点数量下的吞吐量测试。

实验分别开启 6 节点、8 节点、10 节点、12 节点、14 节点、16 节点、18 节点,作为服务器端,其中 P-PBFT 算法的委员会节点个数设置为 6,另开客户端接口,取平均数据吞吐量,结果如图 5 所示,其中横坐标为节点数,纵坐标为吞吐量。

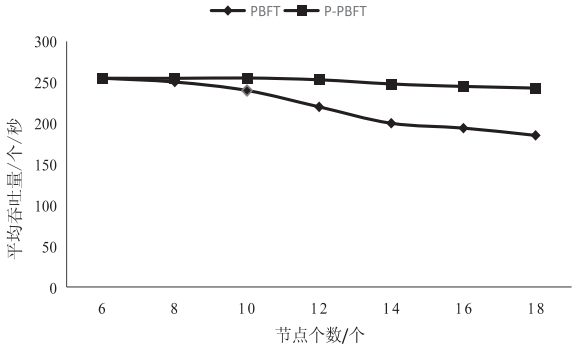

图 5　不同节点数量下的吞吐量对比

可以看出,在节点数量较少的情况下,P-PBFT 算法和 PBFT 算法的吞吐量相差不大。这是因为节点数量较少的时候,P-PBFT 算法参与共识的节点数量和 PBFT 算法相近,且选举委员会节点也需要消耗一定的资源和时间。

(3)积分变化测试。

对 10 个节点分别编号为 $N_1 \sim N_{10}$,其中 N_1、N_2、N_3、N_4 的积分赋值为 10、8、6、4,其余节点的积分赋值为 1,N_1、N_2、N_3、N_4 成为委员会节点,N_1 节点担任主节点,设置衰减周期为 10 分钟。经过十轮交易之后,各节点积分如表 1 所示。

表 1　十轮交易前后积分对比

节点名称	积分初值	十轮交易后积分值
N_1	10	60
N_2	8	38
N_3	6	36
N_4	4	34
N_5	1	11
N_6	1	11
N_7	1	11
N_8	1	11
N_9	1	11
N_{10}	1	11

可以看出,随着共识成功的交易量增加,委员会里的节点与其他节点的积分差距扩大,可以提高参与共识的节点的积极性。

当将主节点端口关闭(即模拟主节点发生错误的情况)时,客户端发送请求,会触发视图切换协议,主节点被扣除 5% 的积分,积分排名第二的 N_2 就成为新的主节点。系统运行十分钟后,所有节点积分全部扣除 70%,并重新进行委员会节点选举。

4 结束语

针对实用拜占庭算法 PBFT 在联盟链节点数较多的情况下性能欠佳的问题,对其加以改进,设计了一种基于积分制的共识算法 P-PBFT。结合 DPOS 思想并引入积分制度,降低了算法通信开销,可以支持更多的节点。改进主节点的选举方式,提高了节点参与共识的积极性;设置积分衰减机制,避免了联盟链中可能出现的过度中心化问题,使得算法更加适应实际的区块链系统。实验与分析表明 P-PBFT 算法能有效地减少系统通信开销,提高系统吞吐量。

参考文献:

[1] NAKAMOTO S. Bticoin:a peer-to-peer electronic cash system[EB/OL]. 2008. http://bitcoin. org//Bitcoin. pdf.

[2] 王元地,李 粒,胡 谍. 区块链研究综述[J]. 中国矿业大学学报:社会科学版,2018,20(3):74-86.

[3] 何渝君,龚国成. 区块链技术在物联网安全相关领域的研究[J]. 电信工程技术与标准化,2017,30(5):12-16.

[4] GERVAIS A,KARAME G O,WÜST K,et al. On the security and performance of proof of work blockchains[C]//Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. Vienna:ACM,2016:3-16.

[5] PKING S,NADAL S. PPCoin:peer-to-peer crypto-currency with p-roof-of-stake[EB/OL]. 2012. https://peercoin. net/assets/paper/peercoin-paper. pdf.

[6] BitShares. Delegated proof of stake[EB/OL]. 2018-04-10. http://d-ocs. bitshares. org/bitshare/d-pos. html.

[7] CASTRO M, LISKOV B. Practical byzantine fault tolerance[J]. ACM Transactions on Computer Systems,2002,20(4):398-461.

[8] LAMPORT L. The part-time parliament[J]. ACM Transactions on Computer Systems,1998,16(2):133-169.

[9] ONGARO D,OUSTERHOUT J K. In search of an understandable consensus algorithm[C]//2014 USENIX annual technical conference. [s. l.]:USENIX,2014:305-319.

[10] 沈 鑫,裴庆祺,刘雪峰. 区块链技术综述[J]. 网络与信息安全学报,2016,2(11):11-20.

[11] 刘敖迪,杜学绘,王 娜,等. 区块链技术及其在信息安全领域的研究进展[J]. 软件学报,2018,29(7):2092-2115.

[12] 吴 映. 基于区块链的考勤记录系统的研究与实现[D]. 西安:西安电子科技大学,2018.

[13] Linux Foundation. Hyperledger fabric[EB/OL]. 2016. https://hyperledger-fabric. readthedocs. io.

[14] LAMPORT L, SHOSTAK R E, PEASE M C. The Byzantine generals problem[J]. ACM Transactions on Programming Languages and Systems,1982,4(3):382-401.

[15] 袁 勇,倪晓春,曾 帅,等. 区块链共识算法的发展现状与展望[J]. 自动化学报,2018,44(11):2011-2022.

[16] KINI U A,SHETTY S M. Peer-to-peer networking[J]. Resonance,2001,6(12):69-79.

[17] 周彬彬,刘 鹏. 计算机对等网络 P2P 技术[J]. 移动信息,2016(2):30.

更 正

本刊 2021 年第 5 期第 209 页刊发的彭云建,欧善国,梁进撰写的《在线气象科普知识竞赛试题的自动组卷方法》,作者简介有误,更正为:彭云建(1974-),男,博士,副教授,硕导,研究方向为随机优化、强化学习控制与信息系统工程;通讯作者:欧善国(1964-),男,高级工程师,研究方向为气象科学传播和教育、应用气象、农业气象。