文章编号:2096-1472(2023)12-0001-05
DOI:10.19644/j.cnki.issn2096-1472.2023.012.001

基于区块链的众包系统研究综述

吉 原,蒋凌云

(南京邮电大学计算机学院,江苏 南京 210023)
✉ niojiy@163.com; jianglingyun@njupt.com

摘 要:基于区块链的众包系统可以为用户提供不依赖集中式中央平台且安全可信的交易环境。文章系统地梳理了近五年来基于区块链的众包系统的研究工作。首先,介绍了众包的概念以及区块链技术的基本原理。其次,分析了传统基于集中式中央平台的众包系统存在的问题,并针对这些问题提出基于区块链技术的解决思路。再次,从架构设计、安全隐私、共识机制、数据存储四个方面分析了基于区块链的众包系统设计所面临的问题以及对应的解决方案。最后,展望了基于区块链的众包系统领域未来的研究方向。

关键词:区块链;众包;共识机制;智能合约
中图分类号:TP399 **文献标志码**:A

Overview of Crowdsourcing Systems Based on Blockchain

JI Yuan, JIANG Lingyun

(*School of Computer, Nanjing University of Posts and Telecommunications, Nanjing* 210023, *China*)
✉ niojiy@163.com; jianglingyun@njupt.com

Abstract: Blockchain-based crowdsourcing systems can provide users with a secure and trustworthy trading environment that does not rely on a centralized central platform. The paper systematically reviews the research work on blockchain-based crowdsourcing systems in the past five years. First of all, the concept of crowdsourcing and the basic principles of blockchain technology are introduced. Then, the problems of traditional crowdsourcing systems based on centralized central platforms and solutions based on blockchain technology are analyzed. Next, the problems in blockchain-based crowdsourcing systems and corresponding solutions have been analyzed from the perspectives of architecture design, security and privacy, consensus mechanism, and data storage. Finally, the future research directions in the field of blockchain-based crowdsourcing systems are discussed.

Key words: blockchain; crowdsourcing; consensus mechanism; smart contracts

0 引言(Introduction)

众包(Crowdsourcing)[1]是一种面向大众公开招募任务工作者的问题解决模式,通过招募具有相应技能的个体,完成一些复杂难解的任务,这种模式近年受到了工业界和学术界的广泛关注。然而,传统的基于集中式中央平台的众包模型存在单点故障、隐私泄露、信任风险等问题。区块链作为一种去中心化的分布式账本技术,为解决传统基于集中式中央平台的众包模型存在的问题提供了一种可行方案。本文调研了基于区块链的众包系统的相关研究工作,首先介绍了众包和区块链技术的概念,其次分析了传统基于集中式中央平台的众包模型存在的问题,并从架构设计、安全隐私、共识机制、数据存储等方面分析了基于区块链的众包系统面临的主要问题及针对这些问题常用的解决方案,同时针对众包数据质量以及共识协议通信复杂度问题,本文给出了基于信誉系统的解决思路。最后本文

收稿日期:2023-03-28
基金项目:国家自然科学基金项目(62072252)

对基于区块链的众包系统未来的研究方向进行了展望。

1 相关技术介绍(Introduction to relevant technologies)

1.1 众包技术

HOWE[1]于2006年提出了众包这一概念,并将其描述为一种将传统上由员工完成的任务以公开招募的形式外包给一大群人的想法。通过众包这种商业模式,企业或个人可以在众包平台上发布任务并将任务分配给有意愿完成此任务的互联网上的大众,以此寻求任务解决方案。

典型的众包系统通常由三组角色组成:任务请求者、工作者和众包平台(如图1所示)。任务请求者在众包平台上描述任务需求并发布任务,平台中对该任务感兴趣的、符合技能要求的平台用户报名参加任务,平台或者任务请求者选择合适的工作者。被选中的工作者在完成任务后将结果提交给平台,当任务执行结果满足任务请求者设置的要求后,平台将最终结果发送给任务请求者,并向相应的工作者发放奖励。设计好的激励机制可以吸引更多的平台用户参与众包任务,因此成为众包领域的研究热点。为了保证激励机制的有效执行,需要选择一个可信的第三方平台(也可称为中介)对交易和服务进行托管,保证众包服务与奖励之间的顺利交换。因此,传统的众包系统普遍基于集中式的架构完成众包各个阶段任务的执行。

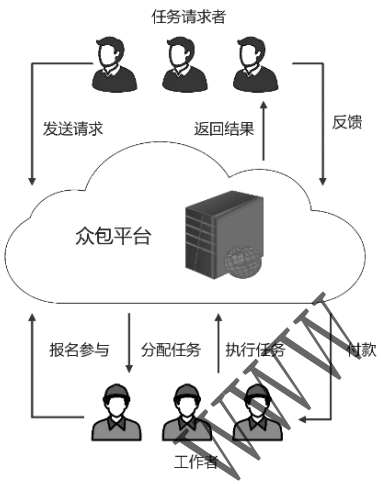

图1 典型的众包系统架构

Fig. 1 Typical architecture of crowdsourcing system

1.2 区块链技术

区块链的概念首次出现在比特币白皮书中,但是该白皮书中并未对区块链做出精确定义[2]。目前,学术界将区块链视为一种去中心化、集体维护的分布式数据库,其数据结构由按时间顺序排列的数据区块组成,每个区块包含一段时间内的事务信息、相应的时间戳和指向前一个区块的哈希指针等信息,并通过密码学技术加强事务信息的完整性,避免信息被篡改和伪造[3]。

区块链账本由网络中的众多节点共同维护,系统中的节点使用点对点(Peer-To-Peer,P2P)网络连接通信,P2P网络中的节点地位平等,不需要中央服务器集中协调。为了保证分布式网络中各节点数据的一致性,需要设计合理的共识机制。共识的过程就是节点验证事务信息并更新区块链的过程。共识机制作为区块链技术的核心,对区块链的安全性、公平性及效率等方面起着关键作用。

区块链系统中一个典型的运作流程如下:用户在区块链客户端发起请求消息,各节点将消息数据在网络中广播,网络中参与共识的节点验证请求数据,各节点根据共识机制完成用户请求并将一段时间内的请求数据打包生成区块,节点将新区块在网络中广播,其余节点验证新区块并更新其本地区块链。

1.3 智能合约

智能合约是一种不需要中介、自我验证、自动执行合约条款的计算机程序,近年来随着区块链技术的广泛应用而备受关注[4]。智能合约的概念最早于1994年被提出,它被定义为一种能够按照事先写好的规则自主执行且不受外界人为干预的计算机程序,但当时无法为智能合约提供可信的执行环境,所以在很长的一段时间内没有得到广泛应用[5]。随着区块链技术的出现,人们发现区块链系统可以为智能合约提供去中心化的可信执行环境,使得智能合约的概念得到实际应用。同时,区块链系统也可以借助智能合约的可编程特性实现一些复杂的交易功能。

区块链系统中相关用户对规则进行协商,达成一致后创建智能合约代码,并将该合约代码部署到区块链上。一旦满足了合约的触发条件,预定义的合约代码将自主执行,并将执行后的结果打包进区块,经共识验证之后发布到区块链中。通过智能合约,平台用户和陌生人可以在去中心化的环境中安全、公平地进行交易。

2 基于区块链技术的众包系统方案(Crowdsourcing system scheme based on blockchain technology)

基于集中式中央平台的众包模型存在一些问题:首先,中央平台容易因分布式拒绝服务攻击(Distributed Denial of Service,DDoS)或系统故障导致服务不可用,损害平台用户的利益;其次,中央平台并不一定完全公平,可能会偏向支付服务费的任务请求者,当任务请求者和工作者之间发生冲突时,需要依赖平台的主观仲裁,这会引起公平性方面的风险;最后,中央平台的数据库中一般存储了大量的用户隐私信息(如姓名、年龄、地理位置、职业信息等),可能存在用户隐私泄露的风险。

区块链为解决上述问题提供了可行的方案。首先,区块链去中心化机制可以有效地解决单点故障问题,在没有集中式中央平台的情况下,支持相互不信任的用户安全地完成交易,解决了中央平台存在的公平性问题;其次,区块链的匿名机制能在一定程度上保护用户的隐私。因此,基于区块链的众包系统成为众包领域重要的研究方向。

本章节首先讨论了基于区块链的众包系统的架构设计和实现流程,其次从安全隐私、共识机制和数据存储方面介绍了目前区块链领域的相关工作。

2.1 架构设计

一个典型的基于区块链的众包系统流程如图2所示。任务请求者一般通过以下步骤发布需求,进而获得解决方案。

(1)注册:任务请求者和工作者在平台上进行身份注册,获

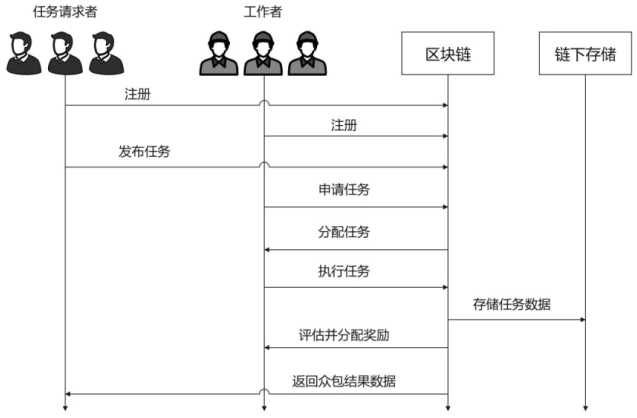

图 2　一个典型的基于区块链的众包系统流程

Fig. 2 A typical blockchain-based crowdsourcing system process

得身份信息,如公私钥、数字证书等。

(2)发布任务:任务请求者向区块链中部署众包智能合约,合约内容包括众包需求信息、任务请求者公钥、工作者报名截止时间、任务截止时间、任务结果评价规则、奖励分配规则等,合约内容透明、公开,并且任务请求者需提前将奖励资金转到智能合约账户地址。

(3)申请任务:愿意执行众包任务的工作者通过任务请求者发布的智能合约进行报名、支付押金的操作。

(4)分配任务:智能合约根据预定的规则选择工作者参与众包,并通知被选择的工作者。

(5)执行任务:被选中的工作者执行众包任务,使用任务请求者的公钥加密任务数据,然后发送至分布式链下存储,并将数据哈希值和指针存储在区块链上。

(6)评估并分配奖励:任务请求者可以通过指针找到任务数据,并用自己的私钥解密;智能合约根据事先设定的规则评价工作者的任务完成质量,确定奖励分配,同时将押金返还给工作者。

CrowdBC 是一个典型的基于区块链的众包平台,其系统架构如图 3 所示[6]。CrowdBC 架构分为应用层、区块链层和存储层。对众包任务感兴趣的工作者可以在应用层查询任务请求者发布的任务,并根据自己具备的相应技能报名相应的任务。区块链层负责对来自应用层的输入执行共识协议以达成共识,并对众包任务最终状态达成一致。在数据存储方面,由于区块链上的数据存储容量有限,因此 CrowdBC 单独抽离出存储层,将众包任务元数据(如数据大小、哈希值、所有者、指针)放在链上存储,原始数据则放在链下存储层,用户可以通过链上的元数据验证存储层数据的完整性和真实性。

CrowdBC 实现了三种类型的智能合约:用户注册合约、用户摘要合约、任务请求者与工作者关系合约。任务请求者或工作者使用用户注册合约进行注册,用户摘要合约为注册成功的用户创建用户信息(包括技能、职业、声誉等)。任务请求者与工作者关系合约则实现了任务请求者与工作者之间的协议,描述了任务发布、任务分配、方案收集和奖励分配的过程。

使用去中心化的智能合约描述复杂的众包逻辑,可以降低对集中式中央平台的依赖性,并增强众包的灵活性。例如,

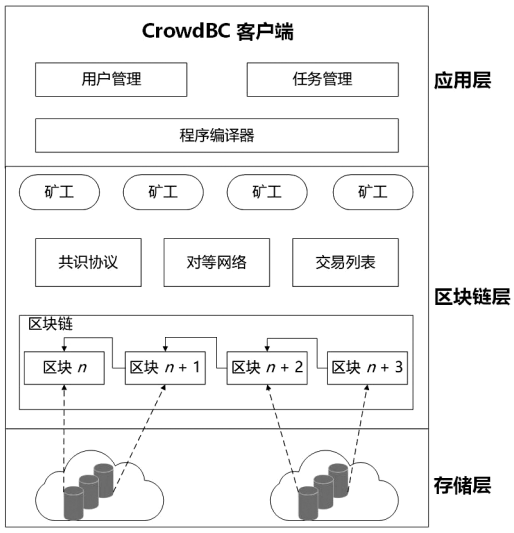

图 3　CrowdBC 系统架构

Fig. 3 System architecture of CrowdBC

GAO 等[7]构建了一种基于区块链技术的新型众包模型,在智能合约中部署工作者选择策略,实现可靠的任务分配。余春堂等[8]提出了基于区块链的众包物流分级多层智能服务交易监管架构,通过编写智能合约实现与区块链网络的交互。KADADHA 等[9]提出了基于区块链和拍卖机制的众包框架 ABCrowd,设计了拍卖算法激励工作者公布其真实成本以获得最高效用,通过以太坊智能合约保证任务请求者和工作者之间的可信交互以及拍卖机制的可信执行。

为了解决工作者能力水平参差带来的众包任务数据质量问题,可以将信誉系统引入众包模型中,使用信誉作为衡量工作者能力的指标,确保所选工作者的能力满足任务要求。许多众包研究工作将工作者的信誉评分建立在众包任务请求者的反馈上,然而众包任务请求者作为利益相关方可能会提交不真实的反馈。为了解决该问题,可以考虑招募第三方验证者负责评价众包工作者完成任务的质量,并在此基础上建立针对工作者的信誉系统。为了激励第三方验证者提交诚实的评价报告,众包模型可以将验证者的评价过程建模为同步报告博弈,使用对等预测和对数打分规则衡量验证者报告的诚实性,在此基础上设计针对验证者的信誉系统。

2.2　安全隐私

在传统的基于集中式中央平台的众包系统中,当任务请求者和工作者进行交易时,相关数据由中央平台进行管理,这不仅增加了众包成本,也带来了数据被伪造或篡改的风险。因此,一个安全可信的众包系统必须保证相关数据的一致性,防止被非法用户篡改。同时,众包系统的用户应该可以在任何时间或者地点访问系统,然而传统的基于集中式中央平台的众包系统面对网络攻击或者其他系统故障时,容易造成系统不可用。朱向荣等[10]提出一个基于区块链的众包知识融合系统 FactChain,区块链的去中心化特性和开放性保证了系统的单点崩溃容错性;区块链的开放透明、可溯源和不可篡改性使得知识的贡献者和更新流程可追溯。FactChain 还利用智能合约实现链上的知识融合逻辑,保证了链上知识的一致性访问。

区块链的开放性和透明性保证了事务信息的可溯源和不

可篡改，是区块链技术重要的优点之一，然而这一优点可能会影响对众包用户的隐私保护。虽然区块链系统的匿名机制允许用户使用与真实身份信息无关的公钥标识身份参与众包，但是由于区块链网络环境的开放性，导致恶意用户可以通过分析链上公开的众包交易记录推断目标用户的个人信息（如真实身份、地理位置等），这造成了极大的隐私泄露风险。YANG等[11]提出了一种基于区块链的隐私保护众包系统，可以保护工作者的位置隐私。为了防止通过重识别进行攻击，该系统使用多个私有区块链分散众包参与者的交易记录，不愿透露位置信息的工作者可以从各种私有区块链中选择任务，因此攻击者无法通过观察相应的事务历史推断工作者的身份和位置信息。ZHANG等[12]提出了一种基于区块链的新型隐私保护车辆感知众包方案PRVB，通过保障车辆与感知数据之间的不可链接性达到保护数据隐私的目的。针对众包任务信息的隐私保护，TONG等[13]提出了一个混合区块链众包平台CHChain。CHChain的核心是其混合式区块链结构，它由多个私有任务链和一个公共链组成，将每个任务的私有信息（如任务响应、评估、反馈等）隔离到只有任务参与者可以访问的私有任务链中，将所有任务的公共信息（如任务ID、奖励、截止日期等）记录到一个公共链中。CHChain在实现众包数据分布式透明存储的同时，还在一定程度上保证了任务敏感信息的私密性。

2.3 共识机制

基于集中式中央平台的众包系统缺乏有效的信任机制，无法确保平台的公平性，也无法有效应对任务请求者或工作者可能存在的不诚实行为。在去中心化的众包系统中设计安全、公平的共识机制是解决上述问题的有效方法。共识机制是区块链系统建立信任的基础，比特币使用工作量证明（Proof of Work，PoW）作为底层共识协议，但该协议吞吐量低、资源消耗大等缺点限制了其广泛应用，其他共识算法[14][如基于实用拜占庭容错算法（Practical Byzantine Fault Tolerance，PBFT）、权益证明（Proof of Stake，PoS）等]也存在诸如拓展性差或缺乏公平性等问题。因此，设计共识机制是实现基于区块链的众包系统的关键，通过设计适用于众包场景下的共识算法，可以提升基于区块链的众包系统的可拓展性、公平性和安全性。

ZOU等[15]提出了一种信任证明（Proof of Trust，PoT）共识协议，适用于众包和一般的在线服务行业，基于众包工作者的信任值选择交易验证者，使用Raft算法选取共识领导者，该协议既避免了PoW机制的低吞吐量、高计算消耗问题，也解决了PBFT共识算法的可拓展性问题。但是，该协议使用Raft算法选出的唯一共识领导者可能不可靠，这会增加节点实施恶意行为的概率。ZHU等[16]针对此问题提出了改进的信任证明共识方案，设计了一种适用于众包场景的信誉算法。该共识方案基于节点信誉值选择共识节点参与共识过程，一方面可以保证共识节点是动态变化的，另一方面大大降低恶意节点影响共识的概率。与其他共识协议相比，该协议具有更高的安全性和可扩展性。FENG等[17]提出了MCS-Chain区块链众包系统，该系统使用了一种新的区块生成共识机制，当等待记录到下一个区块中的累计支付金额超过预先定义的阈值时，就会生成一个新区块，并保证即使多个区块同时出现，也可以确定唯一的块，从而大大降低了计算开销。AN等[18]在基于区块链的众包模型中引入了双共识机制，第一个共识保证了工作者与任务请求者之间的雇佣关系，第二个共识保证了工作者和报酬之间的对应关系，双共识方法避免了欺诈行为的发生。

P2P网络的开放性和动态性（节点的加入和退出）大大增加了其安全风险，在这样的网络中，没有集中的权力验证节点和管理网络。因此，有必要设计一套信誉系统约束共识节点的行为，可以基于信誉对节点进行分组，形成包含多个子集群的底层网络和包含主集群的上层网络，将节点共识通信限制在单个集群内，多个子集群独立、并发地执行共识协议，避免了每个节点向全网所有节点广播消息，从而大大降低了通信的复杂度。

2.4 数据存储

使用区块链存储众包数据存在一些问题：首先，区块过大会增加区块链中全节点的压力，这可能会导致全节点的减少和区块链的集中化；其次，向区块链提交事务、达成共识的成本昂贵[19]。区块链提供的优势可以应用于链下存储，例如可以将众包任务数据的哈希值存储在链上，原始数据存储在链下。由于哈希值相对较小，因此相应的存储成本不高。

星际文件系统（Inter Planetary File System，IPFS）是区块链应用中常见的链下存储方案之一。IPFS是一种分布式数据存储协议，支持高容量存储和高并发访问[20]。IPFS是内容可寻址的，为每个存储的文件分配一个唯一的哈希值，它具有良好的重复数据消除机制，没有中央服务器限制，上传到系统中的数据可以永久保存。对于高频查询的数据，IPFS可以沿着查询路径创建重复数据，以便在下一次查询时直接从本地加载。区块链应用中的事务数据的大小可达数万字节，而IPFS哈希值只有几十个字节。因此，可将IPFS作为基于区块链的众包系统的链下存储方案，将众包数据存储在IPFS中，并将IPFS哈希值存储在区块中，从而大大节省了区块链的存储空间。MUGHAL等[21]提出了基于区块链的河流流量数据采集众包系统，将河流流量聚合数据存储到IPFS，实现高效的分布式存储和共享机制。除此之外，还有许多研究将IPFS作为区块链众包系统的链下存储方案[22-24]。将IPFS分布式链下存储与区块链相结合，能在去中心化的同时起到扩展扩容、减轻链上负担的作用。

除了IPFS技术，其他的存储技术（例如中央云服务器、BitTorrent文件系统、Storj、Swarm等）也可用于区块链众包系统的线下存储，例如MA等[25]提出的区块链众包系统将非敏感元数据记录在区块链上，敏感元数据和加密数据被存储在链下的安全云服务器中。

3 未来研究方向（Future research direction）

除了上文梳理的研究角度，基于区块链的众包系统未来的研究方向还包括如下内容。

（1）智能合约的安全性。智能合约是基于区块链的众包系统的关键组成部分，并且由于区块链的不可篡改特性，使智能合约部署到区块链上后便无法修改，所以一旦智能合约出现代码漏洞并被不法分子利用，将会造成巨大的损失。因此，在将智能合约部署上链之前，有必要对合约代码进行检查，分析其是否存在潜在的安全问题。

(2)用户身份认证。恶意攻击者可能会伪造生成多个身份参与共识过程以影响共识结果,即所谓的女巫攻击。为了保证基于区块链的众包系统的安全性,设计更加复杂的节点身份认证机制,增大发动女巫攻击的难度和成本,是未来重要的研究方向。

4　结论(Conclusion)

基于区块链的众包系统作为众包技术的一个研究领域,有许多需要深入研究的问题。本文梳理了近年来的相关研究工作,并从架构设计、安全隐私、共识机制和数据存储四个方面分析了基于区块链的众包系统设计所面临的问题及其对应的解决方案,并对未来的研究方向进行了展望。希望本文的研究可以对未来的研究工作提供有益参考。

参考文献(References)

[1] HOWE J. The rise of crowdsourcing[J]. Wired Magazine, 2006,14(6):1-4.

[2] 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.

[3] 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.

[4] 欧阳丽炜,王帅,袁勇,等. 智能合约:架构及进展[J]. 自动化学报,2019,45(3):445-457.

[5] ZHENG Z B, XIE S A, DAI H N, et al. An overview on smart contracts: challenges, advances and platforms[J]. Future Generation Computer Systems, 2020, 105:475-491.

[6] LI M, WENG J, YANG A J, et al. CrowdBC: a blockchain-based decentralized framework for crowdsourcing[J]. IEEE Transactions on Parallel and Distributed Systems, 2019, 30(6):1251-1266.

[7] GAO L P, CHENG T, GAO L. TSWCrowd: a decentralized task-select-worker framework on blockchain for spatial crowdsourcing[J]. IEEE Access, 2020, 8:220682-220691.

[8] 余春堂,韩志耕,李致远,等. 基于区块链的众包物流分级多层智能服务交易监管架构[J]. 网络与信息安全学报,2020,6(3):50-58.

[9] KADADHA M, MIZOUNI R, SINGH S, et al. ABCrowd an auction mechanism on blockchain for spatial crowdsourcing[J]. IEEE Access, 2020, 8:12745-12757.

[10] 朱向荣,吴鸿祜,胡伟. FactChain:一个基于区块链的众包知识融合系统[J]. 软件学报,2022,33(10):3546-3564.

[11] YANG M M, ZHU T Q, LIANG K T, et al. A blockchain-based location privacy-preserving crowdsensing system[J]. Future Generation Computer Systems, 2019, 94:408-418.

[12] ZHANG C, ZHU L H, XU C, et al. PRVB: achieving privacy-preserving and reliable vehicular crowdsensing via blockchain oracle[J]. IEEE Transactions on Vehicular Technology, 2021, 70(1):831-843.

[13] TONG W, DONG X W, SHEN Y L, et al. CHChain: secure and parallel crowdsourcing driven by hybrid blockchain[J]. Future Generation Computer Systems, 2022, 131:279-291.

[14] 韩璇,刘亚敏. 区块链技术中的共识机制研究[J]. 信息网络安全,2017(9):147-152.

[15] ZOU J, YE B, QU L, et al. A proof-of-trust consensus protocol for enhancing accountability in crowdsourcing services[J]. IEEE Transactions on Services Computing, 2019, 12(3):429-445.

[16] ZHU X Y, LI Y, FANG L, et al. An improved proof-of-trust consensus algorithm for credible crowdsourcing blockchain services[J]. IEEE Access, 2020, 8:102177-102187.

[17] FENG W, YAN Z. MCS-Chain: decentralized and trustworthy mobile crowdsourcing based on blockchain[J]. Future Generation Computer Systems, 2019, 95:649-666.

[18] AN J, LIANG D W, GUI X L, et al. Crowdsensing quality control and grading evaluation based on a two-consensus blockchain[J]. IEEE Internet of Things Journal, 2019, 6(3):4711-4718.

[19] HEPP T, SHARINGHOUSEN M, EHRET P, et al. On-chain vs. off-chain storage for supply- and blockchain integration[J]. It-Information Technology, 2018, 60(5/6):283-291.

[20] 顾龙,王宏伟. 基于IPFS的分布式数据共享系统的研究[J]. 物联网技术,2016,6(6):60-62.

[21] MUGHAL M H, SHAIKH Z A, ALI K, et al. IPFS and blockchain based reliability and availability improvement for integrated rivers' streamflow data[J]. IEEE Access, 2022, 10:61101-61123.

[22] LIN Y F, LI Z J, YUE W B, et al. CrowdIoT: the crowdsourcing test system for IoT devices based on blockchain[J]. Advances in Internet of Things, 2022, 12(2):19-34.

[23] LI C X, QU X D, GUO Y. TFCrowd: a blockchain-based crowdsourcing framework with enhanced trustworthiness and fairness[J]. EURASIP Journal on Wireless Communications and Networking, 2021(1):168.

[24] YANG Q L, WANG T, ZHANG W B, et al. PrivCrowd: a secure blockchain-based crowdsourcing framework with fine-grained worker selection[J]. Wireless Communications and Mobile Computing, 2021, 2021:3758782.

[25] MA H Y, HUANG E X, LAM K Y. Blockchain-based mechanism for fine-grained authorization in data crowdsourcing[J]. Future Generation Computer Systems, 2020, 106:121-134.

作者简介:

吉　原(1999-),男,硕士生。研究领域:区块链。

蒋凌云(1978-),女,博士,副教授。研究领域:区块链,群智感知。