计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.20220493
59(10):2212-2232,2022

区块链群智感知中基于隐私数据真值估计的激励机制

应臣浩[1] 夏福源[1] 李 颉[1,2,3] 斯雪明[1,2,3] 骆 源[1,2,3]
[1](上海交通大学计算机科学与工程系 上海 200240)
[2](上海交通大学区块链研究中心 上海 200240)
[3](无锡市区块链高等研究中心 江苏无锡 214000)
(yingchenhao@sjtu.edu.cn)

Incentive Mechanism Based on Truth Estimation of Private Data for Blockchain-Based Mobile Crowdsensing

Ying Chenhao[1], Xia Fuyuan[1], Li Jie[1,2,3], Si Xueming[1,2,3], and Luo Yuan[1,2,3]
[1](*Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai* 200240)
[2](*Blockchain Research Center, Shanghai Jiao Tong University, Shanghai* 200240)
[3](*Wuxi Blockchain Advanced Research Center, Wuxi, Jiangsu* 214000)

Abstract Recently, building truth estimation mechanism and participant incentive mechanism upon blockchain-based mobile crowd sensing systems attracts more and more attention. Unlike the traditional mobile crowd sensing system that relies on a centralized platform to host the sensing tasks, due to its decentralized structure, transparent operation and immutability nature, such a system built upon the blockchain is more safe and more interactive. However, the existing researches separately focus on building truth estimation mechanism and participant incentive mechanism, which may lead to the performance limitation in practice. Therefore, in this paper, we propose a participant incentive mechanism based on truth estimation of privacy-preserving data for blockchain-based mobile crowd sensing systems. In fact, it consists of two procedures, the privacy-aware truth estimation procedure (PATD) and the privacy-friendly participant incentive procedure (PFPI), both of which are built by applying Cheon, Kim, Kim, and Song's homomorphic encryption mechanism (CKKS). Due to the low accuracy of data collection devices, the collected data usually mixes with some noise. The collectors encrypt their noisy data. Then PATD utilizes the encrypted data submitted by the collectors to do some calculations and regards the corresponding decrypted result as the truth estimation. The privacy of submitted data can be protected since the data for truth estimation is encrypted by utilizing CKKS. It can also guarantee that the decrypted truth estimation has the high accuracy. Additionally, PFPI can attract more participants by satisfying the truthfulness and individual rationality, and also achieve a high social welfare. The privacy of participants' bids is protected by utilizing CKKS. Finally, numerous experiments are conducted to validate the desirable properties of our proposed mechanism, where the results show that compared with the state-of-the-art approaches, it has better performance.

收稿日期:2022-06-10;**修回日期**:2022-08-10
基金项目:上海市大数据试验场研发与转化功能型平台项目(2022-19);国家自然科学基金重点项目(61932014);国家重点研发计划项目(2020YFB1710900)
This work was supported by the Project of Shanghai Research, Development and Transformation Functional Platform on Big Data (2022-19), the Key Program of the National Natural Science Foundation of China (61932014), and the National Key Research and Development Program (2020YFB1710900).
通信作者:骆源(yuanluo@sjtu.edu.cn)

Key words blockchain; mobile crowdsensing; privacy protection; data collection; truth estimation; incentive mechanism

摘 要 在基于区块链的群智感知系统中构建数据真值估计机制和用户激励机制受到了越来越多的关注.与传统的群智感知系统依赖一个集中平台来承载数据感知任务不同,该系统利用区块链分布式结构和操作透明不可抵赖的特性,使其具有更好的安全性和交互性.但是目前的研究总是独立分离设计数据真值估计机制和参与者激励机制,这导致2类机制在实际应用时往往具有局限性.针对这一问题,在综合考虑了数据真值估计精确度与用户激励后,提出了一类基于隐私保护数据真值估计的用户激励机制.该机制由2个模块组成,具有隐私保护的数据真值估计模块PATD和具有隐私保护的用户激励模块PFPI,这2个模块都是通过利用同态加密机制CKKS来构建的.由于数据采集设备精确度不够等原因,用户收集的数据往往具有噪声,因此PATD对用户提交的含有噪声的数据的加密结果进行计算,并将解密后的计算结果作为相应数据真值的估计.因为所用的数据均是加密的,所以可以保护用户数据隐私,同时,该机制还可以保证解密后的估计值具有较高的估计精度.此外,作为一种激励机制,PFPI满足真实性、个体合理性且具有较高的社会福利,同时利用CKKS保证用户在竞标过程中的竞价隐私安全.最后,进行了大量实验来验证所提的基于隐私保护数据真值估计的用户激励机制的各种特性.实验结果表明,该机制与最新方法相比具有更好的性能.

关键词 区块链;群智感知;隐私保护;数据收集;真值估计;激励机制

中图法分类号 TP391

随着大数据、人工智能技术、5G通信技术的高速发展,人们开始进入一个万物互联的物联网时代.根据华为《华为全球产业展望GIV 2025》所述,到2025年,全球将有1 000亿台可以智能互联的设备,同时个人智能移动终端数量将达到400亿,智能家居及其他可穿戴设备数量将达到210亿,其中智能手机数量将达到80亿,平板和个人电脑数量将达到30亿,各类可穿戴设备数量将达到80亿.90%的人群将拥有个人智能助理,12%的家庭将享有智能服务机器人,20%的人将拥有10个以上的智能终端,平均每个人将拥有5个智能终端.各类数据利用率将剧增至80%,全球每年产生的数据将从2015年的8 ZB增长到1 800 ZB,而且全球人均日通信流量将达到4 GB,同时,人均日移动通信流量将达到1 GB.面对如此庞大的数据量,利用可持续和成本低廉的数据感知和收集方案变得越来越重要.群智感知系统[1-4]在此需求下应运而生,该系统利用人们拥有的无处不在的智能移动设备(如智能手机、可穿戴设备、智能车辆等)中嵌入的传感器(如照相机、陀螺仪等)来获取各种各样的数据[5-8],并将收集到的数据传给远程数据中心进行处理[9-11].

目前,群智感知系统已经成为了一种极具吸引力的大规模数据收集与分析的模式,为面积广袤、人流拥挤地区的信息收集提供了有效的解决方案[12-14].该系统强大而有效的数据收集能力使其成为现代智慧城市不可或缺的组成部分[15].此外,当前嵌入在可穿戴设备和智能手机中的各种传感器,使群智感知系统在学术界和工业界获得了相当大的关注,并被应用于环境监控[16]、智能交通[17]、智能车辆网络[18]、社交应用[19]、健康监测[20]以及其他许多方面[21-23].目前主流的群智感知系统有Mechanical Turk,Upwork,BikeNet和Uber等.

尽管最近移动设备的普及推动了群智感知系统的大范围普及,但由于集中式平台和移动用户之间的大量数据处理操作和频繁的通信,此类系统会导致网络拥塞和严重延迟[24].不仅如此,由于移动用户参与群智感知系统会产生花费,所以需要设计激励机制来激励移动用户加入到该系统中[25],但是在该群智感知系统中,激励机制的实施依赖于一个可信的第三方,然而在实际应用中,这样一个可信第三方会带来许多新的问题:1)平台的功能可能会受到参与的移动用户或外部攻击者的损害[26];2)该平台可能不稳定[27];3)增加了大规模隐私泄露的风险[28].

最近,区块链作为一种分布式账本[29-31],以其去中心化、安全、透明、不可篡改等优越功能,被广泛应用于各种领域.同时,区块链中设置的智能合约[32]可以高效、准确、快捷地实现各种复杂的交易和功能[33-34].因此,本文利用区块链的智能合约,建立一

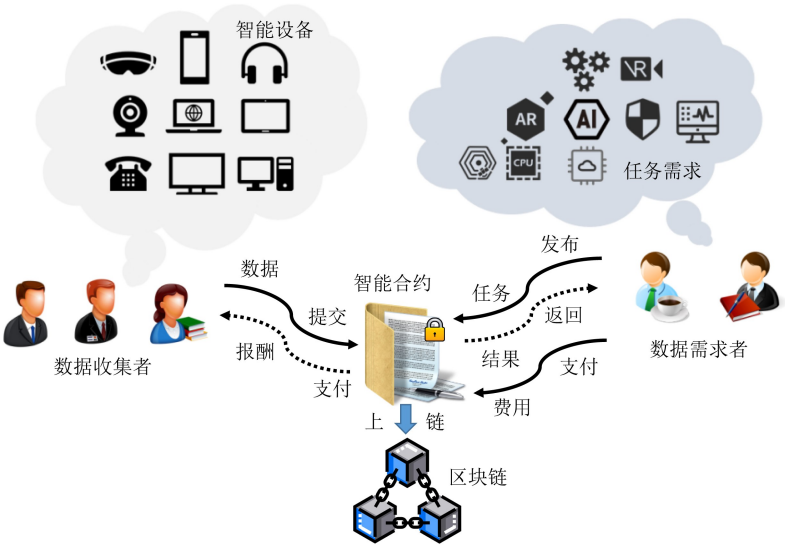

Fig. 1 Blockchain-based mobile crowd sensing systems

图 1 基于区块链的群智感知系统

类如图 1 所示的区块链群智感知系统中基于隐私保护数据真值估计的用户激励机制.该机制由数据真值估计模块和移动用户激励模块共同组成,该机制能够有效提高数据真值估计的准确度,并能够激励更多的移动用户参与到该系统中.

本文的主要贡献有 3 个方面:

1) 机制设计.本文的第 1 个贡献是利用全同态加密算法构建了一类区块链群智感知系统中基于隐私保护数据真值估计的用户激励机制.该机制由数据真值估计模块(PATD)和参与者激励模块(PFPI)组成,能够实现高准确度的数据真值估计和参与者激励.据调研所知,本文尝试在全同态加密状态下实现具有数据真值估计功能的用户激励机制.与传统的机制不同,本文所提机制需要克服移动用户自私属性,提高数据真值估计的准确性,同时要保证全同态加密状态下机制运行的速率.从实验中可以看到,PATD 的数据真值估计准确度相较于已有的方法提高了至少 33%,而 PFPI 达到的社会福利相较于已有的方法提高了至少 21%.同时,当系统具有 128 个数据感知任务时,整个系统的运行时间约为 300 s,且通过不同的参数选择,还可以进一步加快机制运行速度,这表明该机制可被应用于实际的场景.

2) 真值估计.本文的第 2 个贡献是实现了全同态加密状态下的数据真值估计,在保证真值估计的准确度和机制运行速度的同时,保护了用户的数据隐私.由于数据采集设备精确度不够等原因,用户收集的数据往往具有噪声,因此 PATD 对用户提交的含有噪声的数据的加密结果进行计算,并将解密后的计算结果作为相应数据真值的估计.因为所用的数据均是加密的,因此可以保护用户数据隐私,该性质在定理 3 中给出.同时,该机制还可以保证解密后的估计值具有较高的估计精度,可以从定理 1 中看到,估计误差呈指数衰减.就本文作者所知,CKKS 是目前性能最好、计算速度最快的同态加密方案[35],且该方案被 ZAMA 所应用并进行了适当的改进[36].

3) 用户激励.本文的第 3 个贡献是在全同态状态下保证数据真值估计准确性的同时实现了移动用户的激励.PFPI 同样利用 CKKS 方案实现了竞价加密状态下用户激励,同时还从理论上证明了其满足真实性和个体合理性,并可以达到较高的社会福利,可以从定理 2 看到,PFPI 在社会福利上达到了 2 的近似度,同时可以从定理 4 得到,PFPI 可以保护用户竞价的隐私.

1 相关工作

目前,已有基于区块链的群智感知系统,它们利用区块链去中心化、安全、透明的性能构建了数据真值估计机制和移动用户激励机制,分别实现了高效的性能表现,但是这些机制都是分开来建立的,即只建立了数据真值估计机制或者只建立了移动用户激励机制.由这些机制组成的数据收集机制往往很难在实际的系统中实现较好的表现.接下来,我们将

分别介绍基于区块链的数据真值估计机制和用户激励机制.

1.1 基于区块链的数据真值估计机制

Tian 等人[37]提出了一个分布式数据真值估计机制,该机制可以有效地保护用户数据的隐私.该机制将数据融合和处理任务委托给分布式的参与者,通过利用区块链中的智能合约技术来执行和验证其行为.同时,由于区块链缺乏对链上数据保密性的支持,它们利用隐私保护解决方案防止数据泄露.此外,鉴于它们框架的去中心化性质,使得该框架还克服了单点故障的限制,增加了系统的稳定性.但是,该机制对数据隐私的保护是通过对原始数据加入高斯噪声实现的,由于无法消除高斯噪声对真值估计的影响,虽然在理论上可以证明其具有一定的真值估计准确度,但是在实际操作时往往无法达到很好的效果.

Wu 等人[38]提出了一个基于区块链的数据真值估计机制,该机制可以提供可靠的数据真值.为了减轻恶意参与者的影响,该机制集成了一种具有隐私保护的感知验证协议.通过该协议,一些数据参与者可以在不知道任何数据的情况下协作验证真值估计结果.同时,该机制还可以通过经济激励,促使移动用户诚实地参与到群智感知系统中.但是该系统使用的是加法同态加密方案,由于该方案的局限性,使其无法保证数据真值估计的准确度.不仅如此,该机制在移动用户激励中,无法保护用户的隐私信息.

1.2 基于区块链的移动用户激励机制

Huang 等人[39]通过区块链中的智能合约提出了一个基于完整信息动态博弈的激励机制,该机制利用完全信息的 Stackelberg 博弈来对用户数据进行选择,平台和移动用户都是根据对方可能的策略来选择自己的策略以保证自己在对方策略下的利益最大化,从而达到纳什均衡.该博弈具有唯一纳什平衡点,并应用基于区块链的同态水印技术来保护数据版权.但是该机制无法在用户激励过程中保护它们的隐私,所以无法在实际系统中取得很好的效果.

Zhang 等人[40]提出了一种新颖的隐私保护和可靠的车辆移动群智感知系统.该机制包含一种具有隐私保护的车辆数据聚合方案,以保护参与车辆与感知数据之间的数据隐私和不可链接性.此外,还包括 2 个协议来保护数据隐私并为移动用户提供公平的回报.该系统实现了隐私保护、可靠性和公平性,且具有较高的计算和通信效率.但是该系统依据的同样是加法同态加密方案,所以无法实现在用户激励阶段的隐私保护.

Xie 等人[41]提出了一种新颖的基于名誉激励的无人机辅助移动群智感知框架.该框架利用名誉激励方案来选择具有高名誉的无人机来执行数据感知任务,从而保护无人机和任务发布者之间的数据共享免受内部资源不足的无人机攻击.同时,该框架利用一种基于区块链的数据安全传输方案安全地记录无人机的数据交易.此外,由于资源有限的无人机难以执行计算密集型挖矿任务,因此结合边缘计算以增加区块创建的成功概率.无人机与边缘计算提供者之间的交互被建模为 Stackelberg 博弈,以激励无人机参与区块创建过程,同时提供高质量的服务,在该博弈中,无人机和平台根据对方可能的策略来选择策略以保证自己在对方策略下的利益最大化,从而达到纳什均衡.但是与之前的机制类似,该框架只保护了无人机传输数据的隐私安全,但是并没有考虑用户激励过程中的隐私保护.

2 预备知识

本节分别介绍系统概述、数据真值估计、激励模型、同态加密方案和攻击模型.

2.1 系统概述

本文考虑一个基于区块链的群智感知系统,系统组成部分包括:智能合约、一个加密服务中心、一个数据收集者集合 $W=\{w_1,w_2,\cdots,w_m\}$、一个数据需求者集合 $R=\{r_1,r_2,\cdots,r_n\}$ 和感知任务需求集合 $T=\{\tau_1,\tau_2,\cdots,\tau_n\}$,其中每个需求者 r_i 的任务需求为 τ_i.这些感知任务需要一些数据收集者在本地利用智能设备进行数据收集,然后将感知数据通过区块链发送给相应的需求者,其中,用户 w_j 对于任务 τ_i 的感知数据表示为 $m_{i,j}$.为了消除每个移动用户设备差异带来的数据误差,对于每一个任务 τ_i,智能合约会聚合相关收集者的数据以计算一个数据融合结果 m_i,该结果被视为任务 τ_i 的真值 m_i^* 的估计,其中 m_i^* 对参与者和数据收集者而言是未知的.因为智能合约是公开可见的,所以为了防止隐私信息在智能合约进行一些相关操作时泄露,每个数据需求者和数据收集者借助加密服务中心加密他们的信息(包括收集到的数据、上报的竞价),其中加密服务服务中心也是不可信任的,因此也需要防止中心对信息的窃取.图 2 所示为本文所提的基于区块链的群智感知系统中数据收集框架的工作流程.为了方便,表 1 列了一些重要的符号解释.

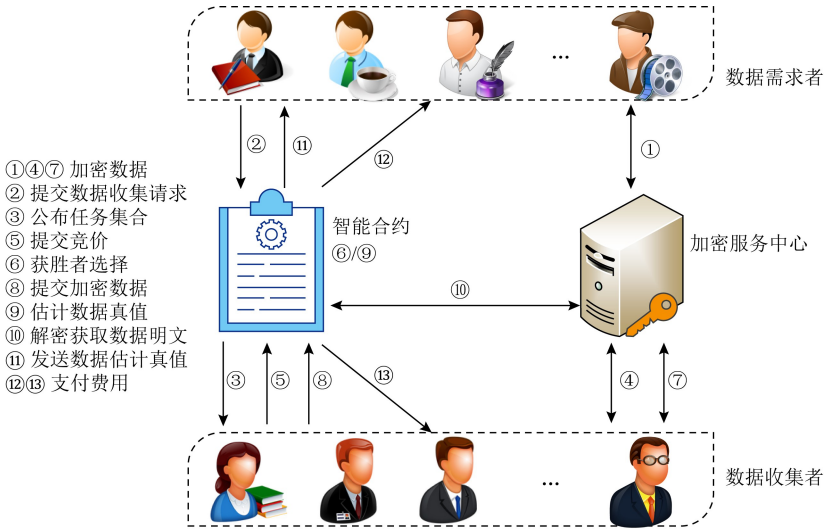

Fig. 2 Workflow of data collection mechanism for blockchain-based mobile crowd sensing systems

图 2 基于区块链的群智感知系统中数据收集机制工作流程

Table 1 Description of Some Important Symbols

表 1 重要符号说明

符号	描述
W,R,T	数据收集者集合、需求者集合、任务集合
S_R,S_{W_j}	获胜的需求者集合、对应于需求者 r_j 的获胜的收集者集合
v_i,c_i	需求者 r_i 的价值、收集者 w_j 的成本
$a_i,b_{i,j}$	需求者 r_i 的竞价、收集者 w_j 的竞价
$\hat{\boldsymbol{a}}_i,\hat{\boldsymbol{b}}_{i,j}$	a_i 的密文,$b_{i,j}$ 的密文
p_{r_i},p_{w_j}	需求者 r_i 支付费用,收集者 w_j 收到费用
$m_i,\hat{\boldsymbol{m}}_i$	任务 τ_i 的数据估计值和相应的密文
u_{r_i},u_{w_j},u_0	需求者 r_i、收集者 w_j 和智能合约的收益
R_j,f,g	随机数
$\alpha_i,\theta_{i,j},\gamma$	选择参数
v,B,N,h,ℓ	加密参数

1) 信息加密.每个数据需求者 r_i 通过加密服务中心的公钥加密她的竞价 a_i,其中竞价 a_i 表示任务 τ_i 被执行时她能够获得的价值(步骤①),然后向智能合约提交一个包含感知任务 τ_i 和加密竞价 $\hat{\boldsymbol{a}}_i$ 的数据收集请求(步骤②).智能合约对外公布数据收集任务集合 T(步骤③).在接收任务后,每个数据收集者 w_j 通过加密服务中心的公钥加密她的竞价 $b_{i,j}$,其中,$b_{i,j}$ 表示执行任务 τ_i 时,收集者 w_j 需要的花费(步骤④),之后收集者 w_j 向智能合约提交她愿意执行的任务集合 $\Gamma_j\subseteq T$ 和每个任务 $\tau_i\in\Gamma_j$ 的相关加密竞价 $\hat{\boldsymbol{b}}_{i,j}$(步骤⑤).

2) 参与者激励.通过加密服务中心的协助,智能合约利用接收到的加密竞价来决定获胜的数据需求者集合 S_R 和执行每个获胜需求者 $r_i\in S_R$ 任务 τ_i 的获胜的数据收集者集合 S_{W_i}、每个获胜需求者 r_i 的需要支付的费用 p_{r_i} 以及支付给每个获胜收集者 w_j 的报酬 p_{w_i}(步骤⑥).实际上,$p_{w_j}=\sum_{i:\tau_i\in\Gamma_j}p_{i,j}$,其中 $p_{i,j}$ 表示对于每个想要执行的任务 τ_i,数据收集者 w_j 从中获取的报酬.此外.注意,因为未获胜需求者的任务不会被执行,所以她们不会支付任何的费用.类似地,未被选中的收集者不会获得任何报酬因为她们不执行任何任务,此时她们各自收益都是 0.

3) 数据加密.每个被选中的数据收集者 $w_j\in S_{W_i}$ 通过加密服务中心的公钥,加密每个任务 $\tau_i\in\Gamma_j$ 收集到的数据 $m_{i,j}$(步骤⑦),然后向智能合约提交加密后的数据 $\hat{\boldsymbol{m}}_{i,j}$(步骤⑧).

4) 数据融合.智能合约对于每个任务 τ_i,计算获胜的收集者所提交加密数据的融合结果 $\hat{\boldsymbol{m}}_i$(步骤⑨),然后借助加密服务中心解密得到明文 m_i(步骤⑩),并将 m_i 发送给相应的获胜的数据需求者 $r_i\in S_R$(步骤⑪)),其中 m_i 便是智能合约计算的对于任务 τ_i 数据真值 m_i^* 的估计值.

5) 报酬收集.最后,智能合约向获胜的需求者 r_i 收取费用 p_{r_i}(步骤⑫)),并向被获胜的收集者 w_j 支付报酬 p_{w_j}(步骤⑬)).

2.2 数据真值估计

为了消除由于数据收集者设备差异带来的数据误差,对于每个任务,智能合约利用数据收集者的数据估计数据真值.因此,与已有的工作类似,本文也

采用了 Jin 等人在文献[42]中提出的数据真值估计机制.为简便起见,在算法 1 中描述了相应的算法.需要注意的是,本文主要关注的是连续任务,而离散任务的真值估计则省略了,但是它可以用类似的方法获得.

算法 1. 数据真值估计算法.

输入:获胜需求者选择集合 S_R 和对应的获胜收集者集合 S_{W_i},其中 $i:r_i\in S_R$,每个任务 τ_i 的阈值参数 α_i,每个获胜收集者 w_j 对于任务 τ_i 的置信度 $\theta_{i,j}$ 和收集的数据 $m_{i,j}$;

输出:每个获胜数据需求者 r_i 的任务 τ_i 的数据真值的估计值 m_i.

/ * 智能合约的操作 * /

for 获胜需求者 $r_i\in S_R$ 的任务 $\tau_i\in\Gamma$ do

① 计算:

$$m_i=\sum_{j:w_j\in S_{W_i},\tau_i\in\Gamma_j}\frac{(\alpha_i-\theta_{i,j})m_{i,j}}{\sum\limits_{k:w_k\in S_{W_i},\tau_i\in\Gamma_k}(\alpha_i-\theta_{i,k})};$$

end for

如算法 1 所示,智能合约利用如下公式计算需求者 r_i 的连续任务 τ_i 真值的估计值:

$$m_i=\sum_{j:w_j\in S_{W_i},\tau_i\in\Gamma_j}\frac{(\alpha_i-\theta_{i,j})m_{i,j}}{\sum\limits_{k:w_k\in S_{W_i},\tau_i\in\Gamma_k}(\alpha_i-\theta_{i,k})},\quad(1)$$

其中,α_i 表示智能合约选取的任务 τ_i 的阈值参数,满足 $\max\limits_{j:\tau_i\in\Gamma_j}\theta_{i,j}<\alpha_i<0.5$,$\theta_{i,j}$ 表示收集者 w_j 对于任务 τ_i 的置信水平,$\theta_{i,j}$ 是智能合约已知的.因为在任务被执行之前,收集者 w_j 关于每个任务 τ_i 的数据可被认作是一个随机变量 $M_{i,j}$,$\theta_{i,j}$ 的定义如定义 1.

定义 1. 收集者 w_j 对于一个连续任务 $\tau_i\in\Gamma$ 的置信水平 $\theta_{i,j}$ 为

$$\theta_{i,j}=E[|M_{i,j}-m_i^*|]\in[0,1],\quad(2)$$

其中数学期望是由于 $M_{i,j}$ 的随机性.

此外,作为一个数据真值的估计需要具有较高的准确度.因此,定义 2 给出了连续任务 (α,γ)-准确度的概念.

定义 2. 对于范围[0,1]内的 2 个随机变量 Y_1 和 Y_2,当 $Pr[|Y_1-Y_2|\geqslant\alpha]\leqslant\gamma$ 时,我们称 Y_1 对 Y_2 是 (α,γ)-准确的,其中 $\alpha,\gamma\in(0,1)$.

这意味着一旦数据估计值和真值满足定义 2,则它们有很高概率是相当接近的.

2.3 激励模型

数据收集者和数据请求者都是自私的,因此都希望最大化她们各自的收益.为了激励更多参与者加入到系统中,本文采用了类似于 Jin 等人在文献[43]中的双边激励模型,其定义如下.

定义 3. 在一个双边激励模型中,每个竞价为 a_i 的数据需求者 r_i 拥有一个感知任务 τ_i,该任务需要被多个数据收集者执行,其中竞价 a_i 表示需求者能够从该任务中获得的价值.该任务被执行时产生一定的实际价值记为 v_i,需求者需向智能合约支付费用 p_{r_i},需求者为了最大化收益可能谎报任务的价值,即 $a_i\neq v_i$.此外,每个数据收集者 w_j 选择执行的任务集合为 $\Gamma_j\subseteq T$,其中,对于每个任务 $\tau_i\in\Gamma_j$,她有一个竞价 $b_{i,j}$,该竞价表示她执行该任务时产生的花费.而收集者 w_j 的实际开销为 $c_{i,j}$,并且将会从智能合约获取报酬 $p_{i,j}$,同样,为了最大化利益,收集者可能会谎报花费,即 $b_{i,j}\neq c_{i,j}$.为了保护隐私,需求者 r_i 和收集者 w_j 将会分别提交加密竞价 $\hat{\boldsymbol{a}}_i$ 和 $\hat{\boldsymbol{b}}_{i,j}$.注意,每个需求者 r_i 的任务 τ_i 所具有的价值 v_i 和收集者 w_j 的花费 $c_{i,j}$ 对于智能合约来说都是未知的.

对于任务 τ_i,当需求者 $r_i\in R$ 能够获得的价值为 v_i,需求者 $w_j\in W$ 的实际开销为 $c_{i,j}$ 时,需求者 r_i 的对应收益为

$$u_{r_i}=\begin{cases}v_i-p_{r_i},&\text{如果 } r_i\in S_R,\\0,&\text{否则.}\end{cases}\quad(3)$$

收集者 w_j 的对应收益为

$$u_{w_j}=\begin{cases}\sum\limits_{i:w_j\in S_{W_i}}(p_{i,j}-c_{i,j}),&\text{如果 } w_j\in S_{W_i},\\0,&\text{否则.}\end{cases}\quad(4)$$

此外,智能合约的收益可以表示为

$$u_0=\sum_{r_i\in S_R}p_{r_i}-\sum_{r_i\in S_R}\sum_{j:w_j\in S_{W_i}}p_{i,j}.\quad(5)$$

如式(3)~(5)所示,收益 u_{r_i},u_{w_j},u_0 实际分别是数据需求者 r_i、数据收集者 w_j 和智能合约的净利润.由于收集者和需求者的自私和策略特性,对于每个任务 τ_i,为了最大化收益,每个需求者 r_i 可能会发送一个加密竞价 $\hat{\boldsymbol{a}}_i$,其明文 a_i 与实际价值 v_i 不同.类似地,每个收集者 w_j 可能发送一个加密竞价 $\hat{\boldsymbol{b}}_{i,j}$,其明文 $b_{i,j}$ 与实际开销 $c_{i,j}$ 不同.这一问题可以通过激励模型的真实性来解决.

定义 4. 如果对于感知任务 τ_i,收集者 w_j 的收益在提交加密竞价 $\hat{\boldsymbol{b}}_{i,j}$ 的明文 $b_{i,j}$ 等于她的实际开销 $c_{i,j}$ 时达到最大,并且需求者 r_i 的收益在提交加密竞价 $\hat{\boldsymbol{a}}_i$ 的明文 a_i 等于获得的实际价值 v_i 时达到

最大,则一个激励模型被称为是双边真实的.

除了真实性,为了激励更多的参与者,设计的激励模块还需要满足个体合理性,具体定义如下.

定义 5. 如果需求者 r_i 和收集者 w_j 的收益分别满足 $u_{r_i}>0$ 和 $u_{w_j}>0$,激励模型满足双边个体合理性.

定义 6. 基于区块链的群智感知系统的社会福利定义为

$$U=u_0+\sum_{r_i\in R}u_{r_i}+\sum_{w_j\in W}u_{w_j}=\sum_{r_i\in S_R}v_i-\sum_{r_i\in S_R}\sum_{j:w_j\in S_{W_i}}c_{i,j}. \tag{6}$$

2.4 同态加密方案

每个数据收集者和数据需求者的信息(例如,收集的数据和提交的竞价)可能会被其他参与者窃听.因此,本文采用 Cheon 等人在文献[44]中提出的同态加密方案 CKKS 来保护它们的隐私.作为同态加密方案,CKKS 支持实数和复数的近似计算.算法如下,其中 $\mathbb{R}$ 为实数域,$\mathbb{C}$ 为复数域.

1) 密钥生成操作 $KeyGen(L,1^\lambda)$

① 给定一个深度参数 L 和一个安全参数 λ,算法选择一个 2 的幂整数 N 并设置密文的模值 $q_\ell=q_0\times p^\ell$,其中 $1\leqslant\ell\leqslant L$,$q_0>0$,$p>0$.$q_0$ 是一个基本模数,p 是一个底数,二者都是整数.

② 密钥分布 χ_{key}、误差分布 χ_{err} 和加密分布 χ_{enc} 都在 $R:=\mathbb{Z}[X]/(X^N-1)$ 多项式环上设置,其中 $\mathbb{Z}$ 是整数环.

③ 算法选取一个密钥参数 $\boldsymbol{s}\leftarrow\chi_{\text{key}}$ 并设置密钥为 $\boldsymbol{sk}\leftarrow(1,\boldsymbol{s})$.

④ 算法同时选取一个公共参数 $\boldsymbol{a}\leftarrow U(R_{q_L})$,并选取一个误差参数 $\boldsymbol{e}\leftarrow\chi_{\text{err}}$,其中 $U(R_{q_L})$ 为 R_{q_L} 上的均匀分布,R_{q_L} 是一个多项式商环 R/q_LR.然后,算法根据选中的参数设置公钥为 $\boldsymbol{pk}\leftarrow(\boldsymbol{b},\boldsymbol{a})\in R_{q_L}^2$,其中 $\boldsymbol{b}$ 是由前边参数确定的另外一个公共参数,其表达式为 $\boldsymbol{b}\leftarrow-\boldsymbol{a}\cdot\boldsymbol{s}+\boldsymbol{e}(\bmod q_L)$.

⑤ 算法采样一个计算参数 $\boldsymbol{a}'\leftarrow U(R_{q_L^2})$ 和另一个误差参数 $\boldsymbol{e}'\leftarrow\chi_{\text{err}}$,然后设置计算密钥为 $\boldsymbol{evk}\leftarrow(\boldsymbol{b}',\boldsymbol{a}')\in R_{q_L^2}^2$,与④设置类似,$\boldsymbol{b}'\leftarrow-\boldsymbol{a}'\cdot\boldsymbol{s}+\boldsymbol{e}'+q_L\cdot\boldsymbol{s}^2(\bmod q_L^2)$ 是一个计算参数.其中 用于 2 个密文的乘法操作,也就是为了进行 2 个密文的乘法运算,必须使用计算密钥 $\boldsymbol{evk}$.

2) 加密操作 $Enc(\boldsymbol{pk},m)$

① $H=\{\boldsymbol{z}\in\mathbb{C}^N:z_j=\bar{z}_{N-j}\}$,其中 $\bar{\boldsymbol{z}}\in\mathbb{C}^N$ 为 $\boldsymbol{z}\in\mathbb{C}^N$ 的共轭.在通过简单地复制明文 m 获取 $\vec{\boldsymbol{m}}\in\mathbb{C}^{N/2}$ 之后,算法的加密操作利用同构映射 $\pi:H\rightarrow\mathbb{C}^{N/2}$ 计算 $\Delta\cdot\pi^{-1}(\vec{\boldsymbol{m}})$,在这之后,该算法进一步计算一个消息多项式 $\tilde{\boldsymbol{m}}=\sigma^{-1}\left(\lfloor\Delta\cdot\pi^{-1}(\vec{\boldsymbol{m}})\rceil_{\sigma(R)}\right)\in R$,其中 $\lfloor y\rceil$ 表示最接近 y 的整数,σ 是后文定义中引入的规范化嵌入.

② 算法选取一个加密参数 $\boldsymbol{v}\leftarrow\chi_{\text{enc}}$ 和 2 个误差参数 $\boldsymbol{e}_0,\boldsymbol{e}_1\leftarrow\chi_{\text{err}}$,随后加密算法输出对应密文 $\hat{\boldsymbol{m}}\leftarrow\boldsymbol{v}\cdot\boldsymbol{pk}+(\tilde{\boldsymbol{m}}+\boldsymbol{e}_0,\boldsymbol{e}_1)(\bmod q_L)$.

3) 解密操作 $Dec(\boldsymbol{sk},\hat{\boldsymbol{m}})$

① 对于一个密文 $\hat{\boldsymbol{m}}=(\hat{\boldsymbol{m}}_0,\hat{\boldsymbol{m}}_1)\in R_{q_\ell}^2$,该解密算法通过进行计算估计得到的消息多项式 $\tilde{\boldsymbol{m}}'\leftarrow\hat{\boldsymbol{m}}_0+\hat{\boldsymbol{m}}_1\cdot\boldsymbol{s}(\bmod q_\ell)$.

② 用 $\vec{\boldsymbol{m}}'\leftarrow\pi(\sigma(\Delta^{-1}\cdot\tilde{\boldsymbol{m}}'))\in\mathbb{C}^{N/2}$ 计算明文 m'.

注意,CKKS 的加密过程引入了一个误差,因此它的解密值与输入值不完全相同.下面将介绍相应的同态运算,包括加法、乘法、标量乘法和缩放.

4) 加法操作 $Add(\hat{\boldsymbol{m}},\hat{\boldsymbol{m}}')$

对于密文 $\hat{\boldsymbol{m}},\hat{\boldsymbol{m}}'\in R_{q_\ell}^2$,算法输出 $\hat{\boldsymbol{m}}_{\text{add}}=\hat{\boldsymbol{m}}+\hat{\boldsymbol{m}}'(\bmod q_\ell)$.

5) 乘法操作 $Mul(\hat{\boldsymbol{m}},\hat{\boldsymbol{m}}')$

对于 2 个相应的密文 $\hat{\boldsymbol{m}}=(\hat{\boldsymbol{m}}_0,\hat{\boldsymbol{m}}_1),\hat{\boldsymbol{m}}'=(\hat{\boldsymbol{m}}_0',\hat{\boldsymbol{m}}_1')\in R_{q_\ell^2}$,乘法操作令 $(d_0,d_1,d_2)=(\hat{\boldsymbol{m}}_0\hat{\boldsymbol{m}}_0',\hat{\boldsymbol{m}}_0\hat{\boldsymbol{m}}_1'+\hat{\boldsymbol{m}}_0'\hat{\boldsymbol{m}}_1,\hat{\boldsymbol{m}}_1\hat{\boldsymbol{m}}_1')$,得到 $\hat{\boldsymbol{m}}_{\text{mul}}=(d_0,d_1)+\lfloor q_L^{-1}\cdot d_2\cdot\boldsymbol{evk}\rceil(\bmod q_\ell)$,从中可以看出,在 CKKS 全同态加密算法中,计算 2 个密文的乘法时,必须使用计算密钥 $\boldsymbol{evk}$,否则乘法操作将无法进行.

6) 常量乘法 $SMul(\bar{a},\hat{\boldsymbol{m}})$

对于常量 $\bar{\boldsymbol{a}}\in R$ 和一个密文 $\hat{\boldsymbol{m}}\in R_{q_\ell}^2$,算法输出 $\hat{\boldsymbol{m}}_{\text{cmul}}=\bar{\boldsymbol{a}}\cdot\hat{\boldsymbol{m}}$.

7) 缩放操作 $Res(\hat{\boldsymbol{m}})$

对于密文 $\hat{\boldsymbol{m}}\in R_{q_\ell}^2$,算法输出 $\hat{\boldsymbol{m}}'\leftarrow\left\lfloor\frac{q_{\ell'}}{q_\ell}\hat{\boldsymbol{m}}\right\rceil(\bmod q_{\ell'})$,说明 $\hat{\boldsymbol{m}}'\in R_{q_{\ell'}}^2$.

定义 7. 对于一个消息多项式 $\tilde{\boldsymbol{m}}\in R$,$\tilde{\boldsymbol{m}}$ 的典范嵌入 $\sigma(\tilde{\boldsymbol{m}})$ 为在第 M 个 N 次割圆多项式 $\Phi_M=X^N+1$ 的根带入计算得到的对应值,也就是说有 $\sigma(\tilde{\boldsymbol{m}})=(\tilde{\boldsymbol{m}}(\zeta_M^j))_{j\in\mathbb{Z}_M^*}\in\mathbb{C}^N$,$\zeta_M=\exp(-2\pi i/M)$ 为一次本原单位根,$\mathbb{Z}_M^*=\{j\in Z_M:gcd(j,M)=1\}$.

定义 8. 对于一个实数消息多项式 $\tilde{\boldsymbol{m}}\in R$,$\tilde{\boldsymbol{m}}$ 的典范嵌入范数 $\|\tilde{\boldsymbol{m}}\|_\infty^{\text{can}}$ 为 $\sigma(\tilde{\boldsymbol{m}})$ 上的 ℓ_∞-范数,表示为 $\|\tilde{\boldsymbol{m}}\|_\infty^{\text{can}}=\max_{j\in\mathbb{Z}_M^*}\{\|\tilde{\boldsymbol{m}}(\zeta_M^j)\|\}$,其中 $\mathbb{Z}_M^*$ 和 ζ_M 的定义在定义 7 中给出.

对于一个消息多项式 $\tilde{\boldsymbol{m}} \in R$,根据推导,加密实数消息多项式 $\tilde{\boldsymbol{m}}'$可以被表示为 $\tilde{\boldsymbol{m}}' = \tilde{\boldsymbol{m}} + \tilde{\boldsymbol{e}}$,其中 $\tilde{\boldsymbol{e}}$ 是一个误差多项式.

定义 9. 如果对于常量 ν 和 B,有 $\|\tilde{\boldsymbol{m}}\|_{\infty}^{\text{can}} \leqslant \nu$, $\|\tilde{\boldsymbol{e}}\|_{\infty}^{\text{can}} \leqslant B$,则称方案$(\hat{\boldsymbol{m}}, \ell, \nu, B)$为 $\tilde{\boldsymbol{m}} \in R$ 的有效加密,其中 $\hat{\boldsymbol{m}}$ 是加密深度为 ℓ 的密文.

2.5 攻击模型

与之前在具有隐私保护的数据真值估计方面的工作类似,我们的安全目标是确保在整个数据真值估计过程中,数据收集者和数据需求者的竞价以及获胜收集者提交的数据受到保护,隐私信息不会被其他参与者获取.事实上,智能合约对于所有参与者是公开透明的,因此其他参与者可能通过窃听智能合约得到需求者和收集者的隐私信息,同时加密服务中心也是诚实但是好奇的,这意味着,加密服务中心能够诚实地遵循预先设计的通信协议,但也试图通过机制执行过程中收到的智能合约发送的信息推断参与者提交的竞价和感知数据.此外,所有其他数据收集者和数据需求者也诚实地遵循预先设计的通信协议,但是,他们也都对其他参与者提交的竞价和数据感到好奇.遵循现有工作中的类似假设,我们假设加密服务中心和其他参与者之间不存在共谋.特别注意,在机制执行过程中,加密服务中心不会主动对智能合约进行窃听并对加密数据进行解密,只会从智能合约发来的信息中推断参与者隐私信息.

需要注意的是,检测那些提交无效数据和竞价来故意破坏系统的恶意数据收集者和数据需求者并不是本文研究的内容.请注意,在本文提出的数据收集框架中,可以使用安全传输协议,例如 SSL 来验证不同参与方之间的通信.

3 具有隐私保护的数据真值估计模块

基于区块链的群智感知系统中数据收集框架由 2 部分组成,本节将详细介绍具有隐私保护的数据真值估计模块.

3.1 设计合理性

由于数据采集设备内部的各个器件本身尺寸具有误差,以及不同数据采集应用之间存在差异,移动用户利用不同设备不同应用对同一个任务采集到的数据往往带有噪声,即彼此数据并不相同.因此,为了能够准确得到数据的真值,需要利用收集到的大量用户数据进行真值估计.实际上,真值估计机制的设计是移动群智感知系统中一个重要的研究方向[1,8,12].

移动用户提交的数据往往含有用户的隐私信息,虽然因为设备和应用自身的差异使得采集到的数据具有噪声,但是这些噪声较为微小,因此这些数据虽然各不相同,但却都在一个较小的范围波动.得到这些数据后,仍然可以推测用户的隐私信息.例如测量移动用户所在地中午 12 点的气温,虽然测得的数据含有噪声,但是温度数据仅在很小的范围波动,因此可以利用这些数据推测用户所在地的一些信息.甚至有时候因为用户设备所带来的噪声,也会导致用户隐私泄露.例如用不同的手机拍摄同一个物品,因为设备的不同,照片的色温等参数可能不同.在得到这些照片后,可以从这些差别中分析用户使用的设备品牌.因此在利用用户采集的数据进行真值估计的同时,还需要保护用户的隐私信息,防止隐私泄露.为此本文利用全同态加密算法设计了数据真值估计机制,该机制在保护用户隐私的同时,具有较高的估计准确度.

3.2 设计细节

为防止隐私泄露,数据真值估计模块利用了 CKKS 同态加密方案对每个数据收集者和数据需求者提交的信息(包括竞价和感知数据)进行加密.利用式(1)对加密数据的真值进行估计,具有隐私保护的数据真值估计模块(PATD)的工作方式如下:

如算法 2 所示,加密服务中心应用 CKKS 的密钥生成操作生成一个公钥 $\boldsymbol{pk}$、一个密钥 $\boldsymbol{sk}$ 和一个计算密钥 $\boldsymbol{evk}$,然后将公钥 $\boldsymbol{pk}$ 发布给所有参与者,每个参与者按照 CKKS 加密操作对相关信息进行加密.注意,如 CKKS 的加密操作所示,消息 $\vec{\boldsymbol{m}} \in \mathbb{C}^{N/2}$.因此,每个获胜的收集者 $w_j \in S_{W_i}$ 将收集到的需求者 r_i 的任务 τ_i 的一维数据 $m_{i,j}$ 通过简单的复制操作扩展为 $N/2$ 维数据,即 $\vec{\boldsymbol{m}}_{i,j} = \underbrace{(m_{i,j}, \cdots, m_{i,j})}_{N/2\text{个}m_{i,j}}$.利用数据向量 $\vec{\boldsymbol{m}}_{i,j}$,一个数据多项式 $\tilde{\boldsymbol{m}}_{i,j} \leftarrow \lfloor \tau^{-1}(\pi^{-1}(\Delta \cdot \vec{\boldsymbol{m}}_{i,j})) \rceil \in R$ 能够通过逆同构映射 τ^{-1}、π^{-1}和预先定义的基底 p 获得.之后,通过公钥 $\boldsymbol{pk}$,每个获胜数据收集者 w_j 根据 CKKS 的加密操作来加密要提交的数据,加密后的数据表示为 $\hat{\boldsymbol{m}}_{i,j}$.

算法 2. 具有隐私保护的数据真值估计算法.

输入:获胜需求者集合 S_R 和对应的获胜收集者集合 S_{W_i},其中 $i: r_i \in S_R$,每个任务 τ_i 的阈值参数 α_i,每个获胜收集者 w_j 对于任务 τ_i 的置信度 $\theta_{i,j}$ 和收集的加密数据 $\hat{\boldsymbol{m}}_{i,j}$;

输出:每个获胜数据需求者 r_i 的任务 τ_i 数据的真值的估计值 m_i.

/* 智能合约的操作 */

for 获胜需求者 $r_i \in S_R$ 的任务 $\tau_i \in T$ do

for 获胜收集者 $w_j \in S_{W_i}$ do

① 计算 $\beta_{i,j}=\dfrac{\alpha_i-\theta_{i,j}}{\sum\limits_{k:w_k\in S_{W_i},\tau_i\in\Gamma_k}(\alpha_i-\theta_{i,k})}$;

② 利用同态映射得到权重多项式 $\tilde{\boldsymbol{\beta}}_{i,j}$;

end for

③ 利用数乘操作 $SMul(\cdot,\cdot)$ 计算 $\tilde{\boldsymbol{\beta}}_{i,j}\cdot\hat{\boldsymbol{m}}_i$;

④ 利用缩放操作 $Res(\cdot,\cdot)$ 计算

$$[\tilde{\boldsymbol{\beta}}_{i,j}\cdot\hat{\boldsymbol{m}}_{i,j}]'\leftarrow\left\lfloor\frac{q_{\ell'}}{q_\ell}\tilde{\boldsymbol{\beta}}_{i,j}\cdot\hat{\boldsymbol{m}}_{i,j}\right\rceil\pmod{q_{\ell'}};$$

for 获胜收集者 $w_j \in S_{W_i}$ do

⑤ 利用加法操作 $Add(\cdot,\cdot)$ 计算

$$\hat{\boldsymbol{m}}_i\leftarrow\sum_{j:w_j\in S_{W_i}}[\tilde{\boldsymbol{\beta}}_{i,j}\cdot\hat{\boldsymbol{m}}_{i,j}]';$$

end for

end for

此外,智能合约收集每个获胜收集者 w_j 的对于任务 τ_i 的置信度 $\theta_{i,j}$ 并为任务 τ_i 设置阈值参数 α_i,满足 $\max\limits_{j:\tau_i\in\Gamma_j}\theta_{i,j}<\alpha_i<0.5$.然后,智能合约为每个任务集 Γ_j 中包含任务 τ_i 的获胜收集者 $w_j\in S_{W_i}$ 计算权重参数 $\beta_{i,j}=\dfrac{\alpha_i-\theta_{i,j}}{\sum\limits_{k:w_k\in S_{W_i},\tau_i\in\Gamma_k}(\alpha_i-\theta_{i,k})}$,并通过简单复制操作获取 $N/2$ 维权重参数向量 $\vec{\boldsymbol{\beta}}_{i,j}=(\beta_{i,j},\cdots,\beta_{i,j})$.利用 $\vec{\boldsymbol{\beta}}_{i,j}=(\beta_{i,j},\cdots,\beta_{i,j})$,智能合约进一步通过逆同构映射 τ^{-1} 和 π^{-1} 获取权重参数多项式 $\tilde{\boldsymbol{\beta}}_{i,j}$.

当从获胜收集者 $w_j\in S_W$ 处接得到为获胜需求者 $r_i\in S_R$ 的任务 τ_i 提交的加密数据 $\hat{\boldsymbol{m}}_{i,j}$,智能合约进行常量乘法操作 $\tilde{\boldsymbol{\beta}}_{i,j}\cdot\hat{\boldsymbol{m}}_{i,j}=SMul(\tilde{\boldsymbol{\beta}}_{i,j},\hat{\boldsymbol{m}}_{i,j})$.在这之后,由于 CKKS 扩散的特性,智能合约需要通过缩放模数 $q_{\ell'}$ 来执行缩放操作 $Res(\cdot,\cdot)$,该缩放值为 $[\tilde{\boldsymbol{\beta}}_{i,j}\cdot\hat{\boldsymbol{m}}_{i,j}]'\leftarrow\left\lfloor\dfrac{q_{\ell'}}{q_\ell}\tilde{\boldsymbol{\beta}}_{i,j}\cdot\hat{\boldsymbol{m}}_{i,j}\right\rceil\pmod{q_{\ell'}}$,这里面有 $\ell'=\ell-1$.最后,执行加法操作 $Add(\cdot,\cdot)$ 得到加密数据估计结果 $\hat{\boldsymbol{m}}_i\leftarrow\sum\limits_{j:w_j\in S_{W_i},\tau_i\in\Gamma_j}[\tilde{\boldsymbol{\beta}}_{i,j}\cdot\hat{\boldsymbol{m}}_{i,j}]'$.

注意整个操作实际是在输入为加密数据时计算式(1).智能合约随后提交任务 τ_i 的加密估计值 $\hat{\boldsymbol{m}}_i$ 到加密服务中心,并在 CKKS 解密操作后得到估计值 m_i.获胜数据需求者集合 S_W 和获胜数据收集者集合 S_R 的选择规则将在 3.3 节中描述.

3.3 理论分析

与其他已有的数据真值估计的工作类似,本节也分析了 PATD 模块的估计准确性.因为所用的 CKKS 同态加密方案是一个近似计算方案,所以需要设计参数来消除近似误差,为此有下面的引理.

引理 1. 对于明文数据为 $m_{i,j}$,若 $(\hat{\boldsymbol{m}}_{i,j},\ell,\nu,B)$ 为数据多项式 $\tilde{\boldsymbol{m}}_{i,j}\in R$ 的有效加密,其中 $\nu>0$,$0<\ell<L$,那么 $(\hat{\boldsymbol{m}}_i,\ell',\nu',B')$ 是对应情况下的数据多项式 $\tilde{\boldsymbol{m}}_i\leftarrow\lfloor\sigma^{-1}(\pi^{-1}(p\cdot\vec{\boldsymbol{m}}_i))\rceil\in R$ 的有效加密,在这其中,$\vec{\boldsymbol{m}}_i=\underbrace{(m_i,\cdots,m_i)}_{N/2\text{个}m_i}$,

$$\begin{aligned}\nu'&=\sum_{j:w_j\in S_{W_i},\tau_i\in\Gamma_j}p^{-1}\cdot\|\tilde{\boldsymbol{\beta}}_{i,j}\|_\infty^{\text{can}}\cdot\nu,\\ B'&=\sum_{j:w_j\in S_{W_i},\tau_i\in\Gamma_j}(p^{-1}\cdot\|\tilde{\boldsymbol{\beta}}_{i,j}\|_\infty^{\text{can}}\cdot B+B_{\text{scale}}),\\ B_{\text{scale}}&=\sqrt{N/3}\cdot(3+8\sqrt{h}),\\ B&=8\sqrt[2]{2}\sigma N+6\sigma\sqrt[2]{N}+16\sqrt[2]{hN},\end{aligned}\tag{7}$$

而 h 为密钥参数 s 的方差,σ^2 为每个 $\tilde{\boldsymbol{m}}_{i,j}$ 的系数的方差,N 为 2 的幂次整数.整数 $p>0$ 为基底,$\tilde{\boldsymbol{\beta}}_{i,j}\leftarrow\lfloor\sigma^{-1}(\pi^{-1}(p\cdot\beta_{i,j}))\rceil\in R$ 为由逆同构 σ^{-1} 和 π^{-1} 计算的权重参数多项式,其中,$\vec{\boldsymbol{\beta}}_{i,j}=\underbrace{(\beta_{i,j},\cdots,\beta_{i,j})}_{N/2\text{个}\beta_{i,j}}$,$\beta_{i,j}$ 为式(1)中相应权重系数.

证明. 如算法 2 所示,智能合约对数据收集者 w_j 为数据需求者 r_i 的任务 τ_i 收集的加密数据 $\hat{\boldsymbol{m}}_{i,j}$ 进行了一次乘法操作和一次缩放操作,然后进行了一些加法操作.因此,借助 Cheon 等人在文献[44]的引理 1~4 可知,结论成立. 证毕.

虽然如引理 1 所示,乘法操作、缩放操作和加法操作引入了一些噪声,但一些工作证明了解密后的值仍然确保较高的精度.为此,下面的定理分析 PATD 模块的准确性.

定理 1. 对于准确度参数为 α_i 的获胜需求者 $r_i\in S_R$ 的每个任务 $\tau_i\in T$,和获胜收集者 $w_j\in S_{W_i}$ 的置信度 $\theta_{i,j}$,用引理 1 设置相应参数,基底 p 满足 $p>(2|S_{W_i}|\cdot B_{\text{scale}}+N)+\sum\limits_{j:w_j\in S_{W_i}}16\sqrt{2}\sigma\cdot\|\tilde{\boldsymbol{\beta}}_{i,j}\|_\infty^{\text{can}}$ 时,PATD 模块的 $Pr[|M_i-\varepsilon_i-m_i^*|\geqslant\alpha_i]$ 满足不等式

$$Pr[|M_i-\varepsilon_i-m_i^*|\geqslant\alpha_i]\leqslant \exp\Big(-\sum_{j:w_j\in S_{W_i}}(\alpha_i-\theta_{i,j})^2\Big),$$

此外,当

$$\exp\Big(-\sum_{j:w_j\in S_{W_i}}(\alpha_i-\theta_{i,j})^2\Big)\leqslant\gamma_i \tag{8}$$

时,所提模块是(α_i,γ_i)-准确的,也就是说对于误差满足 $Pr[|M_i-\varepsilon_i-m_i^*|\geqslant\alpha_i]\leqslant\gamma_i$,其中 ε_i 是由CKKS加密方案引起的误差,m_i^* 是任务 τ_i 的数据真值,M_i 是随机变量,表示数据的估计值.

证明. 为方便起见,需求者 r_i 的任务 τ_i 的融合结果可以表示为

$$m_i=\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}m_{i,j}, \tag{9}$$

其中 $\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}=1$.

由于收集者 w_j 关于每个任务 τ_i 的数据在该任务被执行之前可以被认作是随机变量 $M_{i,j}$,这些数据的融合结果也能够被认作是随机变量.因此,有

$$|M_i-\varepsilon_i-m_i^*|=\Big|\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}M_{i,j}-\varepsilon_i-m_i^*\Big|\leqslant |\varepsilon_i|+\sum_{j:w_j\in S_{W_i}}|\lambda_{i,j}(M_{i,j}-m_i^*)|, \tag{10}$$

其中,$\varepsilon_{i,j}$ 为加解密过程中CKKS引入的误差,因为CKKS的计算为近似结果.

该证明首先分析$|\varepsilon_i|$,由定理1可得

$$\|\langle\hat{\boldsymbol{m}}_i,\boldsymbol{sk}\rangle-\tilde{m}_i(\mathrm{mod}\ q_{\ell'})\|_\infty^{\mathrm{can}}=\|\tilde{e}_i\|_\infty^{\mathrm{can}}\leqslant B'=\sum_{j:w_j\in S_{W_i}}(p^{-1}\cdot\|\tilde{\boldsymbol{\beta}}_{i,j}\|_\infty^{\mathrm{can}}\cdot B+B_{\mathrm{scale}}), \tag{11}$$

其中

$$B_{\mathrm{scale}}=\sqrt{N/3}\cdot(3+8\sqrt{h}),\quad B=8\sqrt{2}\sigma N+6\sigma\sqrt{N}+16\sqrt{hN}. \tag{12}$$

$\langle\hat{\boldsymbol{m}}_i,\boldsymbol{sk}\rangle$ 是密文 $\hat{\boldsymbol{m}}_i$ 的解密输出.此外,根据Cheon等人在文献[44]中的引理1,对于一个消息向量$\vec{\boldsymbol{m}}_i\in\mathbb{C}^{N/2}$,对于 $\tilde{\boldsymbol{m}}_i=Ecd(\vec{\boldsymbol{m}}_i,p)$的一个有效加密操作都是对的.$p\cdot\sigma^{-1}(\pi^{-1}(\vec{\boldsymbol{m}}_i))$的一个有效加密,其相应的误差界为 $B''=B'+N/2$.因此,如果 $p^{-1}\cdot B''<1/2$,在CKKS解码算法中,可以通过舍入运算去除相应的误差多项式,即

$$\sum_{j:w_j\in S_{W_i}}p^{-2}\cdot\|\tilde{\boldsymbol{\beta}}_{i,j}\|_\infty^{\mathrm{can}}\cdot B+p^{-1}\cdot(|S_{W_i}|\cdot B_{\mathrm{scale}}+N/2)\leqslant 1/2\Rightarrow\sum_{j:w_j\in S_{W_i}}2\cdot\|\tilde{\boldsymbol{\beta}}_{i,j}\|_\infty^{\mathrm{can}}\cdot B+p(2|S_{W_i}|\cdot B_{\mathrm{scale}}+N)\leqslant p^2. \tag{13}$$

为了表示简便起见,这里令 $b'=2|S_{W_i}|\cdot B_{\mathrm{scale}}+N$,$c'=\sum_{j:w_j\in S_{W_i}}2\cdot\|\tilde{\boldsymbol{\beta}}_{i,j}\|_\infty^{\mathrm{can}}$. 因此,式(13)能够被改写为

$$c'+p\cdot b'\leqslant p^2\Rightarrow p>\frac{b'+\sqrt{b'^2+4c}}{2}\overset{(a)}{\approx}b'+\frac{c'}{b'}, \tag{14}$$

其中,根据式(12),又因为对于极小 x,利用无穷小代换有$\sqrt{1+x}\approx1+\frac{x}{2}$,所以$\overset{(a)}{\approx}$成立.此外,将式(13)中的表达式 b'和c'代入式(14)中,得

$$b'+\frac{c'}{b'}=(2|S_{W_i}|\cdot B_{\mathrm{scale}}+N)+\frac{\sum_{j:w_j\in S_{W_i}}2\cdot\|\tilde{\boldsymbol{\beta}}_{i,j}\|_\infty^{\mathrm{can}}\cdot B}{2|S_{W_i}|\cdot B_{\mathrm{scale}}+N}\approx(2|S_{W_i}|\cdot B_{\mathrm{scale}}+N)+\sum_{j:w_j\in S_{W_i}}16\sqrt{2}\sigma\cdot\|\tilde{\boldsymbol{\beta}}_{i,j}\|_\infty^{\mathrm{can}}. \tag{15}$$

然后分析式(10)中不等式右边的第2项.为了分析误差,我们为每个任务 τ_i 定义一个随机变量 $V_i=\sum_{j:w_j\in S_{W_i}}|\lambda_{i,j}(M_{i,j}-m_i^*)|$,该变量为随机变量 $V_{i,j}=|\lambda_{i,j}(M_{i,j}-m_i^*)|$的和,可得

$$E[V_i]=\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}E[|M_{i,j}-m_i^*|]=\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}\theta_{i,j}. \tag{16}$$

因此,利用霍夫丁不等式,有

$$Pr[|M_i-m_i^*|\geqslant\alpha_i]\leqslant Pr\Big[\sum_{j:w_j\in S_{W_i}}|\lambda_{i,j}(M_{i,j}-m_i^*)|\geqslant\alpha_i\Big]\Rightarrow Pr[V_i\geqslant\alpha_i]\leqslant\exp\left(-\frac{2(\alpha_i-E[V_i])^2}{\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}^2}\right)=\exp\left(-\frac{2\Big(\alpha_i-\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}^2\theta_{i,j}\Big)^2}{\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}^2}\right)=\exp\left(-\frac{2\Big(\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}(\alpha_i-\theta_{i,j})\Big)^2}{\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}^2}\right). \tag{17}$$

为了最小化式(17)最后一个等号后的式子,我们等价地最大化函数 $\phi(\lambda_i)$,该函数定义为

$$\phi(\lambda_i)=\frac{2\Big(\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}(\alpha_i-\theta_{i,j})\Big)^2}{\sum_{j:w_j\in S_{W_i}}\lambda_{i,j}^2}. \tag{18}$$

根据柯西-施瓦茨不等式,可得

$$\phi(\lambda_i) \leqslant \frac{\left(\sum_{j:w_j \in S_{W_i}} \lambda_{i,j}^2\right)\left(\sum_{j:w_j \in S_{W_i}} (\alpha_i - \theta_{i,j})^2\right)}{\sum_{j:w_j \in S_{W_i}} \lambda_{i,j}^2} = \sum_{j:w_j \in S_{W_i}} (\alpha_i - \theta_{i,j})^2, \tag{19}$$

当 $\lambda_{i,j} \propto (\alpha_i - \theta_{i,j})$, 等式成立. 又由 $\sum_{j:w_j \in S_{W_i}} \lambda_{i,j} = 1$,

$$\lambda_{i,j} = \frac{\alpha_i - \theta_{i,j}}{\sum_{k:w_k \in S_{W_i}, \tau_i \in \Gamma_k} (\alpha_i - \theta_{i,k})}, \tag{20}$$

因此, 当 $\lambda_{i,j}$ 满足式(20), 有

$$Pr[|M_i - \varepsilon_i - m_i^*| \geqslant \alpha_i] \leqslant \exp\left(-\sum_{j:w_j \in S_{W_i}} (\alpha_i - \theta_{i,j})^2\right).$$

根据相应的定义, PATD 模块是 (α_i, γ_i)-准确的, 如果

$$\exp\left(-\sum_{j:w_j \in S_{W_i}} (\alpha_i - \theta_{i,j})^2\right) \leqslant \gamma_i, \tag{21}$$

即结论成立. 证毕.

4 具有隐私保护的用户激励模块

在介绍了 PATD 模块后, 本节将接着介绍具有隐私保护的参与者激励(PFPI)模块, 将分别介绍激励模块的数学优化模型、提出算法, 并进行理论分析.

4.1 数学优化模型

数据真值的估计准确度与所收集的数据量有关, 同时, 所提框架希望能够最大化系统的社会福利, 为此将建立式(22)数学优化模型.

优化问题. 社会福利最大化:

$$\max \sum_{i:\tau_i \in T} \left(a_i - \sum_{j:w_j \in W_i} b_{i,j}\right) x_i,$$
$$\text{s.t.} \sum_{i:\tau_i \in T} \exp\left(-\sum_{j:w_j \in W_i} (\alpha_i - \theta_{i,j})^2\right) x_i \leqslant \gamma, \tag{22}$$
$$x_i \in \{0,1\}.$$

1) 优化常数. 社会福利最大化时, 每一个任务 τ_i 所对应的数据收集者集合 W_i, 数据需求者 r_i 的竞价 a_i, 数据收集者 w_j 的竞价 $b_{i,j}$, 任务集合 T, 每个数据收集者 w_j 的信任水平 $\theta_{i,j}$, 准度参数 α_i 和 γ 是常数, 且 $\gamma = \min\{\gamma_i | i: r_i \in S_R\}$.

2) 目标函数. 为了能够保证 PATD 模块的准确度, 对于任意一个任务 τ_i, 算法将该任务的所有数据都收集上来, 并在此前提下最大化社会福利.

3) 优化约束. 根据式(8)和参数 γ 的选中, 可知当约束不等式成立时, 对于每一个任务 τ_i, 其数据真值估计都满足 (α_i, γ_i)-准确度.

4) 优化变量. 该优化问题中的优化变量为 x_i, 当 $x_i = 1$ 表示任务 τ_i 被智能合约选中并执行, 当 $x_i = 0$ 表示该任务没有被选中.

通过上述数学模型, 可以得到该优化问题是 NP 难问题.

引理 2. 上述优化问题是 NP 难的.

证明. 为了简化问题, 令

$$\delta_i = a_i - \sum_{j:w_j \in W_i} b_{i,j},$$
$$\bar{\omega}_i = \exp\left(-\sum_{j:w_j \in W_i} (\alpha_i - \theta_{i,j})^2\right) \tag{23}$$

上述问题可简化为

$$\max \sum_{i:\tau_i \in T} \delta_i x_i,$$
$$\text{s.t.} \sum_{i:\tau_i \in T} \bar{\omega}_i x_i \leqslant \gamma, \tag{24}$$
$$x_i \in \{0,1\}.$$

可以知道这是一个 0-1 背包问题, 已知 0-1 背包问题为一个 NP 完全问题, 所以可知, 上述优化问题是一个 NP 难问题. 证毕.

由于该问题的 NP 困难属性, 因此, 4.2 节将提出一个近似算法, 高效地解决该问题.

4.2 算法设计

为了解决该问题, 本节提出一个近似算法. 该算法的伪代码如算法 3 所示, 其具体流程如下.

具有隐私保护的用户激励算法由 2 部分组成, 即参与者选择部分和报酬确定部分.

4.2.1 参与者选择部分

算法输入: 收集者集合 W、需求者集合 R、每个数据需求者 r_i 的加密竞价 $\hat{\boldsymbol{a}}_i$、每个收集者 w_j 对每个任务 τ_i 的加密竞价 $\hat{\boldsymbol{b}}_{i,j}$ 和置信度 $\theta_{i,j}$、每个任务 τ_i 的阈值参数 α_i、在 CKKS 同态加密方案中的编码多项式.

算法主体: 对于每一个任务 τ_i, 智能合约计算权重

$$\bar{\omega}_i = \exp\left(-\sum_{j:w_j \in W_i} (\alpha_i - \theta_{i,j})^2\right) \tag{25}$$

以及权重编码多项式 $\hat{\boldsymbol{w}}_i$, 并计算加密状态下的收益

$$\hat{\boldsymbol{\delta}}_i = \hat{\boldsymbol{a}}_i - \sum_{j:w_j \in W_i} \hat{\boldsymbol{b}}_{i,j} \tag{26}$$

和加密状态下的加权收益 $\hat{\boldsymbol{\delta}}_i / \hat{\boldsymbol{w}}_i$. 之后, 智能合约利用 Hong 等人在文献[45]中提出的 K-加密排序算

法将 $\hat{\boldsymbol{\delta}}_i/\hat{\boldsymbol{w}}_i$ 按照从大到小的顺序排序,我们假设 K-加密排序算法输出左边第 ℓ 个密文为 $\hat{\boldsymbol{\varphi}}_\ell$,即对应的明文 φ_ℓ 有第 ℓ 大的值,注意序号 ℓ 是它在排序输出的位置,与 $\hat{\boldsymbol{\varphi}}_\ell$ 对应的算法输入 $\hat{\boldsymbol{\delta}}_i/\hat{\boldsymbol{w}}_i$ 的序号智能合约是不知道的.得到排序结果后,对于每一个任务 τ_i,智能合约选择一个随机数 $R_j\in[0,1]$,该随机数只有智能合约自己知道,并利用密文 $\hat{\boldsymbol{\varphi}}_j$ 计算修正密文 $\hat{\boldsymbol{\zeta}}_{i,j}$

$$\hat{\boldsymbol{\zeta}}_{i,j}=\hat{\boldsymbol{R}}_j\cdot(\hat{\boldsymbol{\delta}}_i/\hat{\boldsymbol{w}}_i-\hat{\boldsymbol{\varphi}}_j),\tag{27}$$

其中 $\hat{\boldsymbol{R}}_j$ 是 R_j 的密文.并将 $\hat{\boldsymbol{\zeta}}_{i,j}$ 发送给加密服务中心,加密服务中心将其解密得到明文 $\zeta_{i,j}$,若 $\zeta_{i,j}<\eta$,则将序号 i 发送给智能合约,其中 η 是一个接近于 0 的随机数,智能合约将 $\hat{\boldsymbol{\delta}}_i/\hat{\boldsymbol{w}}_i$ 排在第 j 个位置.为了方便,我们假设密文 $\hat{\boldsymbol{\delta}}_i/\hat{\boldsymbol{w}}_i$ 的对应明文 $\delta_i/\bar{\omega}_i$ 是第 i 大的值(注意,实际操作过程中往往不是).之后,智能合约计算按照序号依次计算 $\bar{\omega}_i+C$.若 $\bar{\omega}_i+C\leqslant\beta$,则将数据需求者 r_i 加入到需求者备选集合 A_R 中,并得到收集者备选集合 $A_{W_i}=W_i$.当 $\bar{\omega}_i+C>\beta$ 时,智能合约利用 K-加密排序算法比较 $\sum_{i:\tau_i\in S_R}\hat{\boldsymbol{\delta}}_i$ 和 $\hat{\boldsymbol{\delta}}_\ell$ 的大小,其中 $\hat{\boldsymbol{\delta}}_\ell$ 是所有未获胜的数据需求者中其任务具有最大加权收益的,等价地可得 $\ell:\hat{\boldsymbol{\delta}}_\ell=\arg\max\{\hat{\boldsymbol{\delta}}_j/\hat{\boldsymbol{w}}_j\mid j:r_j\notin S_R\}$. $\sum_{i:\tau_i\in S_R}\hat{\boldsymbol{\delta}}_i\geqslant\hat{\boldsymbol{\delta}}_\ell$ 时,则输出 $S_R=A_R$ 和 $S_{W_i}=A_{W_i}$,其中 $i:r_i\in S_R$,我们称这种情况下获胜的参与者分别为第 1 类获胜需求者和第 1 类获胜收集者.否则,得到 $i^*=\arg\max\{\hat{\boldsymbol{\delta}}_i\mid i:r_i\in R\}$.并输出获胜需求者集合 $S_R=\{r_{i^*}\}$ 和获胜收集者集合 $S_{W_{i^*}}=W_{i^*}$,我们称这种情况下的获胜参与者分别为第 2 类获胜需求者和第 2 类获胜收集者.

算法输出:获胜需求者集合 S_R 和每个获胜需求者 r_i 对应的获胜收集者集合 S_{W_i},其中 $i:r_i\in S_R$.

算法 3. 参与者选择部分.

输入:收集者集合 W、需求者集合 R、每个数据需求者 r_i 的加密竞价 $\hat{\boldsymbol{a}}_i$、每个收集者 w_j 对每个任务 τ_i 的加密竞价 $\hat{\boldsymbol{b}}_{i,j}$ 和置信度 $\theta_{i,j}$、每个任务 τ_i 的阈值参数 α_i;

输出:获胜需求者集合 S_R 和对应的获胜收集者集合 S_{W_i},其中 $i:r_i\in S_R$.

/* 智能合约的操作 */

① 定义需求者选择集合 $S_R=\varnothing$ 和对应的收集者选择集合 $S_{W_i}=\varnothing$,其中 $i:r_i\in S_R$,并定义常数 $C=0$;

for 每一个 $\tau_i\in T$ do

② 计算 $\bar{\omega}_i=\exp\left(-\sum_{j:w_j\in W_i}(\alpha_i-\theta_{i,j})^2\right)$ 和权重编码多项式 $\tilde{\boldsymbol{w}}_i$;

③ 计算 $\hat{\boldsymbol{\delta}}_i=\hat{\boldsymbol{a}}_i-\sum_{j:w_j\in W_i}\hat{\boldsymbol{b}}_{i,j}$ 和 $\hat{\boldsymbol{\delta}}_i/\tilde{\boldsymbol{w}}_i$;

end for

④ 利用 K-加密排序算法将 $\hat{\boldsymbol{\delta}}_i/\tilde{\boldsymbol{w}}_i$ 从大到小排序,记算法排序后第 j 个输出值为 $\hat{\boldsymbol{\varphi}}_j$;

for 每一个 $\tau_i\in T$ do

for 每一个 $\hat{\boldsymbol{\varphi}}_j$ do

⑤ 选择一个随机数 R_j;

⑥ 计算 $\hat{\boldsymbol{\zeta}}_{i,j}=\hat{\boldsymbol{R}}_j\cdot(\hat{\boldsymbol{\delta}}_i/\tilde{\boldsymbol{w}}_i-\hat{\boldsymbol{\varphi}}_j)$;

⑦ 发送 $\hat{\boldsymbol{\zeta}}_{i,j}$ 给加密服务中心;

end for

end for

⑧ go to 步骤⑮;

⑨ 得到相应的排序数列 $\hat{\boldsymbol{\delta}}_1/\tilde{\boldsymbol{w}}_1,\cdots,\hat{\boldsymbol{\delta}}_n/\tilde{\boldsymbol{w}}_n$,其中 $\hat{\boldsymbol{\delta}}_i/\tilde{\boldsymbol{w}}_i$ 的明文 $\delta_i/\bar{\omega}_i$ 有第 i 大的值;

while $\bar{\omega}_i+C\leqslant\beta$ do

⑩ $A_R=A_R\cup\{r_i\}$, $A_{W_i}=W_i$;

⑪ $C=C+\bar{\omega}_i$;

end while

⑫ 比较 $\sum_{i:\tau_i\in S_R}\hat{\boldsymbol{\delta}}_i$ 和 $\hat{\boldsymbol{\delta}}_\ell$ 大小;

if $\sum_{i:\tau_i\in S_R}\hat{\boldsymbol{\delta}}_i>\hat{\boldsymbol{\delta}}_\ell$ do

⑬ return $S_R=A_R$ 和所有 $S_{W_i}=A_{W_i}$, 其中 $i:r_i\in S_R$;

else

⑬ 计算 $i^*=\arg\max\{\hat{\boldsymbol{\delta}}_i\mid i:r_i\in R\}$;

⑭ $S_R=\{r_{i^*}\}$, $S_{W_{i^*}}=W_{i^*}$;

⑮ return S_R 和所有 $S_{W_{i^*}}$;

end if

/* 加密服务中心的操作 */

for 每一个 $\tau_i\in T$ do

for 每一个 $\hat{\boldsymbol{\varphi}}_j$ do

⑯ 解密 $\hat{\boldsymbol{\zeta}}_{i,j}$ 得到明文 $\zeta_{i,j}$;

if $\zeta_{i,j}<\eta$ do

⑰ 将 $\zeta_{i,j}$ 的序号 i 发送给智能合约;

end if

end for

end for

⑱ go to 步骤⑦;

在进行完参与者选择部分后,智能合约进入报酬确定部分.

4.2.2 报酬确定部分

算法输入:所有收集者 w_j 的加密竞价 $\hat{\boldsymbol{b}}_{i,j}$ 和所有需求者的加密竞价 $\hat{\boldsymbol{a}}_i$,权重多项式 $\hat{\boldsymbol{w}}_i$,获胜需求者集合 S_R 和对应的获胜收集者集合 S_{W_i},其中 $i:r_i\in S_R$.

算法主体:对于每一个获胜需求者 $r_i\in S_R$,智能合约计算

$$\hat{\boldsymbol{\pi}}_i=\min\left\{\frac{\hat{\boldsymbol{w}}_i}{\hat{\boldsymbol{w}}_\ell}\hat{\boldsymbol{\delta}}_\ell+\sum_{j:w_j\in W_i}\hat{\boldsymbol{b}}_{i,j},\hat{\boldsymbol{\delta}}_{i'}+\sum_{j:w_j\in W_i}\hat{\boldsymbol{b}}_{i,j}\right\},\quad(28)$$

其中 ℓ 是未获胜需求者中加权收益明文值 $\delta_\ell/\bar{\omega}_\ell$ 最大的,即 $\ell=\arg\max\{\delta_j/\bar{\omega}_j\mid j:r_j\notin S_R\}$,而式(28)中相应的序号 $i'=\arg\max\{\delta_j\mid j:r_j\in R_i\}$,其中集合 $R_i=\{r_k\mid r_k\in R,\delta_k<\delta_i\}$.智能合约计算 $\hat{\boldsymbol{\rho}}_i=\hat{\boldsymbol{g}}\cdot\hat{\boldsymbol{\pi}}_{r_i}$,并将 $\hat{\boldsymbol{\rho}}_i$ 发送给加密服务中心,其中 $\hat{\boldsymbol{g}}$ 是 g 的密文而 $g<1$ 是智能合约选择的一个随机数,该随机数只有智能合约自己知道.

对于相应的获胜收集者 $w_j\in S_{W_i}$,智能合约计算

$$\hat{\boldsymbol{\pi}}_{i,j}=\max\left\{\hat{\boldsymbol{\delta}}_i+\hat{\boldsymbol{b}}_{i,j}-\frac{\hat{\boldsymbol{w}}_i}{\hat{\boldsymbol{w}}_\ell}\hat{\boldsymbol{\delta}}_\ell,\hat{\boldsymbol{\delta}}_i+\hat{\boldsymbol{b}}_{i,j}-\hat{\boldsymbol{\delta}}_{i'}\right\},\quad(29)$$

并计算 $\hat{\boldsymbol{\rho}}_{i,j}=\hat{\boldsymbol{h}}\cdot\hat{\boldsymbol{\pi}}_{i,j}$,并将 $\hat{\boldsymbol{\rho}}_{i,j}$ 发送给加密服务中心,其中 $\hat{\boldsymbol{h}}$ 是 h 的密文,而 $h>1$ 是智能合约选择的一个随机数.

加密服务中心收到 $\hat{\boldsymbol{\rho}}_i$ 和 $\hat{\boldsymbol{\rho}}_{i,j}$ 后,进行解密得到 ρ_i 和 $\rho_{i,j}$,之后选择 2 个随机数 f 和 d,其中 $f<1$ 而 $d>1$,该随机数只有中心自己知道,并计算 $p_i=f\cdot\rho_i$ 和 $p_{i,j}=d\cdot\rho_{i,j}$,然后将 p_i 和 $p_{i,j}$ 发送给智能合约.

算法输出:每一个获胜需求者 $r_i\in S_R$ 需要支付的费用 p_i 和对应的获胜收集者 $w_j\in S_{W_i}$ 能够收到的报酬 $p_{i,j}$,其中 $i:r_i\in S_R$.

算法 4. 报酬确定部分.

输入:所有收集者 w_j 的加密竞价 $\hat{\boldsymbol{b}}_{i,j}$ 和所有需求者的加密竞价 $\hat{\boldsymbol{a}}_i$,权重多项式 $\hat{\boldsymbol{w}}_i$,获胜需求者集合 S_R 和对应的获胜收集者集合 S_{W_i},其中 $i:r_i\in S_R$.

输出:获胜需求者 $r_i\in S_R$ 的支付费用 p_i 和对应获胜收集者 $w_j\in S_{W_i}$ 收到的报酬 $p_{i,j}$,其中 $i:r_i\in S_R$.

① 智能合约选择 2 个随机数 $g<1$ 和 $h>1$,对应密文分别为 $\hat{\boldsymbol{g}}$ 和 $\hat{\boldsymbol{h}}$;

② 加密服务中心选择 2 个随机数 $f<1$ 和 $d>1$,其密文分别记为 $\hat{\boldsymbol{f}}$ 和 $\hat{\boldsymbol{d}}$;

/* 智能合约的操作 */

for 每一个 $r_i\in S_R$ do

③ 计算式(28);

④ 计算 $\hat{\boldsymbol{\rho}}_i=\hat{\boldsymbol{g}}\cdot\hat{\boldsymbol{\pi}}_i$,并将 $\hat{\boldsymbol{\rho}}_i$ 发送给加密服务中心;

⑤ go to 步骤⑪;

for 每一个 $w_j\in S_{W_i}$ do

⑥ 计算式(29);

⑦ 计算 $\hat{\boldsymbol{\rho}}_{i,j}=\hat{\boldsymbol{h}}\cdot\hat{\boldsymbol{\pi}}_{i,j}$,并发送 $\hat{\boldsymbol{\rho}}_{i,j}$ 给加密服务中心;

⑧ go to 步骤⑭;

⑨ return $p_{i,j}$;

end for

⑩ return p_i;

end for

/* 加密服务中心的操作 */

for 每一个 $\hat{\boldsymbol{\rho}}_i$ do

⑪ 解密得到明文 ρ_i;

⑫ 计算 $p_i=f\cdot\rho_i$,并发送 p_i 给智能合约;

⑬ go to 步骤⑥;

end for

for 每一个 $\hat{\boldsymbol{\rho}}_{i,j}$ do

⑭ 解密得到明文 $\rho_{i,j}$;

⑮ 计算 $p_{i,j}=d\cdot\rho_{i,j}$,发送 $p_{i,j}$ 给智能合约;

⑯ go to 步骤⑨;

end for

在结束本节之前,我们给出一个例子来具体说明该算法的工作流程.

示例 1. 假设有 3 个数据需求者 $R=\{r_1,r_2,r_3\}$,他们的任务集合为 $T=\{\tau_1,\tau_2,\tau_3\}$,与之相对应的竞价为 $\{a_1=0.5,a_2=0.2,a_3=0.1\}$,以及 3 个数据收集者 $W=\{w_1,w_2,w_3\}$,各自感兴趣的任务为 $\Gamma_1=\{\tau_1,\tau_2,\tau_3\}$,每个任务的竞价为 $b_{1,1}=0.2$,$b_{2,1}=0.1$,$b_{3,1}=0.02$,$\Gamma_2=\{\tau_1,\tau_3\}$,每个任务的竞价为 $b_{1,2}=0.16$,$b_{3,2}=0.024$,$\Gamma_3=\{\tau_2\}$,每个任务的竞价为 $b_{2,3}=0.04$.可以得到 $\delta_1=a_1-b_{1,1}-b_{1,2}=0.14$,$\delta_2=0.06$,$\delta_3=0.056$,为了方便,假设 $\bar{\omega}_1=0.2$,$\bar{\omega}_2=0.3$,$\bar{\omega}_3=0.56$,同时假设 $\gamma=0.2$.因此,可以得到 $\frac{\delta_1}{\bar{\omega}_1}=\frac{0.14}{0.2}=0.7$,$\frac{\delta_2}{\bar{\omega}_2}=0.2$,$\frac{\delta_3}{\bar{\omega}_3}=0.1$.假设 K-加密排

序算法输出的值为 $\varphi_1=0.699$,$\varphi_2=0.198$,$\varphi_3=0.101$(注意,这些值在算法中都是密文).对于任务 τ_1,智能合约选择一个随机数 $R_1=1$,则可以得到 $\zeta_{1,1}=R_1\left(\frac{\delta_1}{\bar{\omega}_1}-\varphi_1\right)=0.001$,$\zeta_{1,2}=-0.499$,$\zeta_{1,3}=-0.599$.加密服务中心解密将序号 1 发送给智能合约,对于任务 τ_2 和 τ_3 的操作类似.智能合约根据序号将密文位置进行相应调整,根据参与者选择部分的操作,可以得到获胜需求者集合为 $S_R=\{r_1\}$,而获胜收集者集合 $S_{W_1}=\{w_1,w_2\}$,因为 $\bar{\omega}_1=\gamma$,且 $\delta_1>\delta_2$.之后进入报酬确定部分,根据相应的操作可得 $\pi_1=\min\left\{\frac{\bar{\omega}_1}{\bar{\omega}_2}\delta_2+(b_{1,1}+b_{1,2}),\delta_2+(b_{1,1},b_{1,2})\right\}$,因此 $\pi_1=\min\{0.40,0.42\}=0.40$,同样地,可以得到 $\pi_{1,1}=\max\left\{a_1-b_{1,2}-\frac{\bar{\omega}_1}{\bar{\omega}_2}\delta_2,a_1-b_{1,2}-\delta_2\right\}$,类似地,可以得到 $\pi_{1,1}=\max\{0.3,0.28\}=0.3$,同理 $\pi_{1,2}=0.26$.智能合约选择随机数 $g=0.9$,$h=1.1$,加密服务中心选择随机数 $f=0.8$,$d=1.2$,则最后可得数据需求者 r_1 需要支付的费用为 $p_1=0.284$,数据收集者 w_1 和 w_2 得到的报酬为 $p_{1,1}=0.396$ 和 $p_{1,2}=0.343$.

4.3 激励属性

本节将证明所提出的 PFPI 模块满足双边真实性、双边个体合理性,同时在社会福利最大化中达到了 2 的近似比.

引理 3. PFPI 对于每个需求者和收集者都满足真实性,这意味着每个需求者和收集者都诚实的报告竞价.

证明. 我们通过证明 PFPI 满足单调性和临界报酬 2 个性质来说明参与的需求者具有真实性,而对于参与的收集者,可以用相似的方法得到结论,因此将省略相关的证明过程.注意,算法使用的 CKKS 是近似算法,但是由于算法的设计,使其在密文状态下的各种计算并没有改变相应的数值,因此在证明过程中所有变量均为明文.

单调性:假设需求者 r_i 是获胜者,则相应的获胜数据收集者集合为 $S_{W_i}=W_i$,如算法 3 所示,当满足 $\sum_{i:r_i\in S_R}\delta_i>\delta_\ell$ 时,可知如果需求者 r_i 获胜了,则有 $r_i\in S_R$ 且 $\frac{\delta_i}{\bar{\omega}_i}>\frac{\delta_\ell}{\bar{\omega}_\ell}$,其中 $\delta_i=a_i-\sum_{j:w_j\in W_i}b_{i,j}$,而 $\bar{\omega}_i=\exp\left(-\sum_{j:w_j\in W_i}(\alpha_i-\theta_{i,j})^2\right)$,相似地,而其中与之相对应的变量 δ_ℓ 和 $\bar{\omega}_\ell$ 分别为 $\delta_\ell=a_\ell-\sum_{j:w_j\in W_\ell}b_{i,\ell}$ 和 $\bar{\omega}_\ell=\exp\left(-\sum_{j:w_j\in W_\ell}(\alpha_\ell-\theta_{\ell,j})^2\right)$.因为 $\bar{\omega}_i$ 和 $\bar{\omega}_\ell$ 不变,所以如果需求者 r_i 以竞价 a_i 获胜,则当她以 $a_i'>a_i$ 竞价时,仍然会获胜.同时,当 $\sum_{r_k\in S_R}\delta_k<\delta_\ell$ 时,如果需求者 r_i 获胜,因为 $\delta_i=a_i-\sum_{j:w_j\in W_i}b_{i,j}$ 是需求者集合 R 中收益最大的,所以如果需求者 r_i 以竞价 a_i 获胜,则当她以 $a_i'>a_i$ 竞价时,仍然会获胜.因此该算法对于需求者而言,满足单调性.

临界报酬:需求者 r_i 支付费用 $p_i=g\cdot f\cdot\pi_i$,其中 $\pi_i=\min\left\{\frac{\bar{\omega}_i}{\bar{\omega}_\ell}\delta_\ell+\sum_{j:w_j\in W_i}b_{i,j},\delta_{i'}+\sum_{j:w_j\in W_i}b_{i,j}\right\}$,序号 $\ell=\arg\max\{\delta_j/\bar{\omega}_j\mid j:r_j\notin S_R\}$,同时,其中的序号 $i'=\arg\max\{\delta_j\mid j:r_j\in R_i\}$,其中集合 $R_i=\{r_k\mid r_k\in R,\delta_k<\delta_i\}$.而 $g<1$ 且 $f<1$.这里我们仅考虑 $\pi_i=\frac{\bar{\omega}_i}{\bar{\omega}_\ell}\delta_\ell+\sum_{j:w_j\in W_i}b_{i,j}$ 的情况,对于 $\pi_i=\delta_{i'}+\sum_{j:w_j\in W_i}b_{i,j}$ 的情况可以用类似的方法得到结论.在下面的证明中,为了方便表示,我们简单地令 $\pi_i'=\delta_{i'}+\sum_{j:w_j\in W_i}b_{i,j}$.考虑需求者 r_i 以 $a_i'<a_i$ 的竞价参与竞争时的情况.

当 $a_i'\geqslant p_i$ 时,获胜的需求者需要支付的费用仍然是 p_i,非获胜需求者则需要支付的费用为 0.因此,下面的证明只考虑 $a_i'<p_i$ 的情况.在这种情况下,可以知道 $a_i'<p_i<\pi_i<\pi_i'$ 成立.根据算法 3 可知,如果需求者 r_i 获胜,则她一定是第 1 类获胜者或第 2 类获胜者.但是当需求者 r_i 以竞价 a_i' 进行竞争时,有 $a_i'<\pi_i$,即 $a_i'<\pi_i=\frac{\bar{\omega}_i}{\bar{\omega}_\ell}\delta_\ell+\sum_{j:w_j\in W_i}b_{i,j}$,可以得到 $\frac{\delta_i'}{\bar{\omega}_i}<\frac{\delta_\ell}{\bar{\omega}_\ell}$,其中 $\delta_i'=a_i'-\sum_{j:w_j\in W_i}b_{i,j}$,根据算法 3 可知,需求者 r_i 将不是第 1 类获胜者.同理可知,有 $a_i'<\pi_i'$,也就是说 $a_i'<\pi_i'=\delta_{i'}+\sum_{j:w_j\in W_i}b_{i,j}$,可以得到 $\delta_i'<\delta_{i'}$,其中 $\delta_i'=a_i'-\sum_{j:w_j\in W_i}b_{i,j}$,因此,根据算法 3 可知,需求者 r_i 将不是第 2 类获胜者.综上可知当 $a'_i<p_i$ 时,需求者 r_i 不会获胜,因此,她的收益为 0. 证毕.

除了真实性,还需要保证 PFPI 模块具有个体合理性.

引理 4. PFPI 模块满足每个需求者和收集者的个体合理性,这意味着每个需求者和收集者都可以得到非负的收益.

证明. 这里只证明对于每个需求者,PFPI 模块满足个体合理性,对于收集者的情况可以利用相同

的方法得到结论.证明中所有变量都是明文.

当需求者 r_i 获胜,则她需要支付的费用为 $p_i=f\cdot g\cdot\pi_i$,其中 $f<1,g<1$ 是 2 个随机数,而且 $\pi_i=\min\left\{\frac{\bar{\omega}_i}{\bar{\omega}_\ell}\delta_\ell+\sum_{j:w_j\in W_i}b_{i,j},\delta_{i'}+\sum_{j:w_j\in W_i}b_{i,j}\right\}$,序号 $\ell=\arg\max\{\delta_j/\bar{\omega}_j\mid j:r_j\notin S_R\}$,同时,其中的序号 $i'=\arg\max\{\delta_j\mid j:r_j\in R_i\}$,其中集合 $R_i=\{r_k\mid r_k\in R,\delta_k<\delta_i\}$.这里我们仅证明情况为 $\pi_i=\frac{\bar{\omega}_i}{\bar{\omega}_\ell}\delta_\ell+\sum_{j:w_j\in W_i}b_{i,j}$ 的结论,$\pi_i=\delta_{i'}+\sum_{j:w_j\in W_i}b_{i,j}$ 时的情况可以用类似的方法证明.为了方便,在该证明中令 $\pi_i'=\delta_{i'}+\sum_{j:w_j\in W_i}b_{i,j}$. 从引理 2 可知,每个数据需求者都诚实地报告她的竞价,即 $a_i=v_i$,因此,她被获胜时的收益 $u_i=v_i-p_i=a_i-f\cdot g\cdot\pi_i$.如果她是第 1 类获胜者,则可以得到 $\frac{\delta_i}{\bar{\omega}_i}>\frac{\delta_\ell}{\bar{\omega}_\ell}$,化简后可得 $a_i>\frac{\bar{\omega}_i}{\bar{\omega}_\ell}\delta_\ell+\sum_{j:w_j\in W_i}b_{i,j}$,即 $a_i>\pi_\ell$,因此 $u_i>0$.如果需求者 r_i 是第 2 类获胜者,则根据算法 3 可知,$\delta_i=\max\{\delta_j\mid j:\tau_j\in T\}$,因此 $\delta_i>\delta_{i'}$,化简可得 $a_i>\delta_{i'}+\sum_{j:w_j\in W_i}b_{i,j}$,即 $a_i>\pi_i'$,可知 $\pi_i>\pi_i'$,因此,$u_i>0$.综上可知结论成立. 证毕.

接下来将证明 PFPI 模块在系统的社会福利上达到了 2 的近似度.为了证明该结论,将利用分数背包问题,其定义如下.

问题. 分数背包问题:

$$\begin{aligned}&\max\sum_{i:\tau_i\in T}\delta_i x_i,\\&\text{s.t.}\sum_{i:\tau_i\in T}\bar{\omega}_i x_i\leqslant\gamma,\\&x_i\in[0,1],\end{aligned}\tag{30}$$

其中 $\delta_i=a_i-\sum_{j:w_j\in W_i}b_{i,j}$,而与之项对应的变量则为 $\bar{\omega}_i=\exp\left(-\sum_{j:w_j\in W_i}(\alpha_i-\theta_{i,j})^2\right)$. 与 0-1 背包问题中变量 x_i 只能取 0 或 1 不同,分数背包问题中,变量 x_i 可以取分数,即 $x_i\in[0,1]$.对于该问题有下述引理.

引理 5. 假设当上述分数背包问题达到最大值 OPT_F 时,x_{i^*} 和 x_{j^*} 属于相应的最优解,那么如果有 $\frac{\delta_{i^*}}{\bar{\omega}_{i^*}}>\frac{\delta_{j^*}}{\bar{\omega}_{j^*}}$,且 $x_{j^*}>0$,则 $x_{i^*}=1$.

证明. 利用反证法证明该结论.假设当上述分数背包问题达到最大值 OPT_F 时,x_{i^*} 和 x_{j^*} 属于相应的最优解,那么如果有 $\frac{\delta_{i^*}}{\bar{\omega}_{i^*}}>\frac{\delta_{j^*}}{\bar{\omega}_{j^*}}$,且 $x_{j^*}>0$,但是此时 $x_{i^*}<1$.令 $\varepsilon=\min\left\{(1-x_{i^*})\frac{\bar{\omega}_{i^*}}{\bar{\omega}_{j^*}},x_{j^*}\right\}$,构造 2 个新的变量取值 $\bar{x}_{j^*}=x_{j^*}-\varepsilon$ 和 $\bar{x}_{i^*}=x_{i^*}+\varepsilon\cdot\frac{\bar{\omega}_{j^*}}{\bar{\omega}_{i^*}}$,其余变量的取值均与最优算法得到的理想解完全相同.因为有 $\varepsilon\cdot\frac{\bar{\omega}_{j^*}}{\bar{\omega}_{i^*}}\cdot\bar{\omega}_{i^*}-\varepsilon\cdot\bar{\omega}_{j^*}=0$,可知新变量取值同样满足约束.根据 ε 定义可知 $\bar{x}_{j^*}=x_{j^*}-\varepsilon\geqslant x_{j^*}-x_{j^*}=0$,且 $\bar{x}_{i^*}=x_{i^*}+\varepsilon\cdot\frac{\bar{\omega}_{j^*}}{\bar{\omega}_{i^*}}\leqslant x_{i^*}+(1-x_{i^*})\frac{\bar{\omega}_{i^*}}{\bar{\omega}_{j^*}}\cdot\frac{\bar{\omega}_{j^*}}{\bar{\omega}_{i^*}}=1$.因此可知新变量取值满足条件.而优化函数的取值变化量为 $\varepsilon\cdot\frac{\bar{\omega}_{j^*}}{\bar{\omega}_{i^*}}\cdot\delta_{i^*}-\varepsilon\cdot\delta_{j^*}>0$.这意味着我们得到了另一组解,且达到了更高的收益,这与 x_{i^*} 和 x_{j^*} 属于相应的最优解矛盾,因此结论成立. 证毕.

利用引理 5,可以得到如下定理.

定理 2. PFPI 模块在系统的社会福利上达到了 2 的近似比.

证明. 根据引理 4 可知,令 $x_\ell=\frac{\beta-C}{\bar{\omega}_\ell}<1$,利用算法 3 可以得到相应的分数背包问题的最优解.因此可知当 $r_j\notin A_R\cup\{r_\ell\}$,该需求者 r_i 也不在相应的分数背包问题的最优解中,即 $x_\ell=0$.由此可得

$$\sum_{k:r_k\in A_R}\delta_k+\delta_\ell\geqslant OPT_F\geqslant OPT,\tag{31}$$

其中 OPT 是原始优化问题的最大值.由此可知

$$\max\left\{\sum_{k:r_k\in A_R}\delta_k,\delta_\ell\right\}\geqslant\frac{OPT}{2}.\tag{32}$$

由算法 3 可知,选中需求者集合为 $S_R=A_R$ 或 $S_R=\{r_{i^*}\}$,其中 $i^*=\arg\max\{\delta_i\mid i:r_i\in R\}$,因此 $\delta_{i^*}\geqslant\delta_\ell$.由此可知该算法满足

$$\max\left\{\sum_{k:r_k\in A_R}\delta_k,\delta_{i^*}\right\}\geqslant\frac{OPT}{2},\tag{33}$$

所以结论成立. 证毕.

5 安全分析

如第 3 节和第 4 节所示,所提出的基于区块链的群智感知系统中数据收集框架利用 CKKS 保护数据需求者和数据收集者的数据以及竞价隐私.本节将详细分析所提框架的安全性.

定理 3. PATD 模块保证了收集者提交的数据对诚实但是好奇的窃听者的安全性,这意味着在执

行算法过程中,加密服务中心和其他参与者除了最后的数据真值的估计值外无法获得其他关于数据的任何信息.

证明. 如 PATD 模块所示,在提交数据之前,每个收集者使用加密服务中心生成的密钥对各自数据进行加密,然后智能合约根据加密数据估计数据真值.最后,加密状态的数据估计值由加密服务中心进行解密.在此过程中,CKKS 的安全性保证了数据隐私.与现有的工作类似,假设加密服务中心和其他参与者(数据需求者和数据收集者)之间没有勾结,且加密服务中心不会在模块执行过程中主动解密收集者提交的加密数据,因此,除了最后公开的数据真值的估计值,没有人可以获得数据的任何信息.综上分析,PATD 模块可以保证数据的安全性. 证毕.

值得注意的是,现有的许多具有隐私保护的数据真值估计算法为了得到最终的估计值需要在计算的过程中解密一些中间结果,而这些解密操作会导致数据隐私泄露,与这些工作不同,PATD 模块不需要解密中间结果来得到最终的数据真值估计值,因此进一步保证了数据的安全性.此外,现在的一些已有工作在保护数据隐私的同时,还保护用户权重的隐私,因为研究者认为在估计真值过程中使用的用户权重也是隐私信息.与这些工作不同,在本文所提算法中,用户权重是他们的置信度,而这些置信度在区块链中是公开可见的,实际上,在区块链中,往往要根据每个参与者的置信度来决定谁有资格当矿工.

定理 4. 具有隐私保护的激励模块在面对半诚实但是好奇的攻击者时,可以保护数据需求者和数据收集者竞价的隐私安全,这意味着在执行算法过程中,智能合约、加密服务提供方和其他参与者除了最后的报酬外无法获得其他关于竞价的任何信息.

证明. 根据算法 3 和算法 4,数据需求者和数据收集者提交的是加密竞价,这意味着智能合约无法直接获取他们的竞价.该激励模块由 2 个子模块组成,参与者选择模块和报酬确定模块,因此下面的证明将分别从对 2 个子模块的安全性进行分析.

对于参与者选择模块,智能合约在收到数据需求者和数据收集者的加密竞价后,计算了每个任务的密态权重收益 δ_i/ω_i,之后利用 K-加密排序算法得到了相应的从小到大排序的输出 $\hat{\boldsymbol{\varphi}}_j$,并选择了一个只有自己知道的随机数 R_j,其密文为 $\hat{\boldsymbol{R}}_j$,同时对于每一个密态加权收益 $\hat{\boldsymbol{\delta}}_i/\hat{\boldsymbol{w}}_i$ 计算 $\tilde{\boldsymbol{\zeta}}_{i,j}=\hat{\boldsymbol{R}}_j\cdot(\hat{\boldsymbol{\delta}}_i/\hat{\boldsymbol{w}}_i-\hat{\boldsymbol{\varphi}}_j)$,其中 $\hat{\boldsymbol{\varphi}}_j$ 是 K-加密排序算法输出结果中排在第 j 个位置的密文,即 $\hat{\boldsymbol{\varphi}}_j$ 对应的明文 φ_j 在所有输出 $\hat{\boldsymbol{\varphi}}_i$ 对应的明文 φ_i 有第 j 大的值,其中 $i:\tau_i\in T$.在计算 $\tilde{\boldsymbol{\zeta}}_{i,j}$ 的过程中,所有变量均为加密状态,所以其他参与者无法得到竞价.之后智能合约将 $\tilde{\boldsymbol{\zeta}}_{i,j}$ 发送给加密服务中心,中心解密后得到每个 $\tilde{\boldsymbol{\zeta}}_{i,j}$ 的明文 $\zeta_{i,j}$,因为随机数 R_j 只有智能合约知道,因此服务中心无法从 $\zeta_{i,j}$ 中得到竞价信息.服务中心将每个 $\zeta_{i,j}$ 对应的任务序号 i 发送给智能合约,智能合约只能得到每个密态加权收益的相对大小,无法知道每个竞价的信息.之后智能合约的操作同样是对密文进行的,所以无法获知竞价信息,用类似的方法对其他涉及比较操作的步骤进行分析可知同样不会泄露竞价.因此参与者选择部分能够保证竞价的隐私安全.

报酬确定部分涉及到数据需求者的费用确定和数据收集者的报酬确定,这里只分析需求者的费用确定部分,后者可以用类似的方法分析.智能合约在加密状态下计算每一个获胜需求者 r_i 的 $\hat{\boldsymbol{\pi}}_i$ 时,涉及到一次比较操作,该操作与参与者选择阶段的比较操作流程一样,因此可知不会泄露竞价隐私.之后智能合约对每一个 $\hat{\boldsymbol{\pi}}_i$ 计算 $\hat{\boldsymbol{\rho}}_i=\hat{\boldsymbol{g}}\cdot\hat{\boldsymbol{\pi}}_i$,其中 $\hat{\boldsymbol{g}}$ 是随机数 g 的密文,且 g 只有智能合约知道.由于 $\hat{\boldsymbol{g}}$ 的存在,加密服务中心无法在解密得到 $\hat{\boldsymbol{\rho}}_i$ 的明文 ρ_i 后得到竞价信息,之后中心选择一个只有自己知道的随机数 f,并将 $p_i=f\cdot g\cdot\pi_i$ 发给智能合约,同样由于 f 的存在,其他参与者无法从智能合约收到的 p_i 得知竞价信息.因此报酬确定部分也可以保证竞价的隐私安全.综合上述分析可知,PFPI 模块可以保证数据需求者和数据收集者竞价的隐私安全.

证毕.

6 实验仿真

本节将从实验仿真的角度验证所提机制的性能,该节首先介绍比较基线,之后将说明实验仿真中的参数设置,最后将给出实验结果.

6.1 比较基线

本文所提的区块链群智感知系统中基于隐私保护数据的用户激励机制包含 PATD 模块和 PFPI 模块 2 部分,因此本节将验证这 2 个机制的性能.

1) 具有隐私保护的数据真值估计机制. 实验仿真考虑 3 种比较基线:第 1 种是均值估计基线(MEAN),该基线将数据的均值作为最后的数据真值的估计值

输出;第 2 种是中位数估计基线(MEDIAN),该基线将数据的中位数作为最后的数据真值估计的输出;第 3 种是迭代估计基线(IBTD),该基线由 Zhang 等人在文献[46]中提出,在每次迭代时,更新每个数据的权重,并利用更新后的权重计算数据真值估计,在达到迭代次数后,机制输出估计结果.注意,实验仿真过程中,均使用 CKKS 同态加密方案对数据进行加密.

2) 具有隐私保护的参与者激励机制. 因为目前没有基于同态加密方案的激励机制,因此实验仿真考虑 2 种基线:第 1 种基于收益的激励机制(BFI),该机制将每个任务 τ_i 的 δ_i 从大到小进行排序,然后从最大的开始选择获胜的数据需求者和收集者;第 2 种基于权重的激励机制(WFI),该机制将每个任务 τ_i 的权重 $\bar{\omega}_i$ 按照从小到大排序,然后每次选择权重最小的任务确定相应的获胜数据需求者和收集者.

6.2 参数设置

为了方便,本节将实验的参数设置列在了表 2 中.与 Jin 等人的文献[43]类似,一些参数都是在某个区间均匀随机选择.具体而言,对于每一个任务 τ_i,阈值参数 α_i 和精度参数 γ_i 在[0.2,0.3]均匀随机选取,每个数据需求者 r_i 能够获得的价值 v_i 在[20,25]上均匀随机选取.同时,每个数据收集者 w_j 的数据 $x_{i,j}$ 都是从均值为 $\mu_{i,j}$、方差为 $\sigma_{i,j}$ 的截断高斯分布中采样得到的,即 $x_{i,j}\in[0,1]$,且 $\mu_{i,j}$ 取值为[0,1],而 $\sigma_{i,j}$ 取值为[1,2],数据置信度 $\theta_{i,j}$ 在[0,0.1]均匀选取,花费 $c_{i,j}$ 取值为[0,1],同时愿意执行的任务集合 Γ_j 中任务的数量 $|\Gamma_j|$ 取值区间为[1,5].在仿真集合 Ⅰ 中,数据收集者的数量 M 从 2 变到 512,而数据需求者的数量 N 保持 100 不变;在仿真集合 Ⅱ 中,数据需求者的数量 N 从 2 变到 512,而数据收集者的数量 M 保持 100 不变.

Table 2 The Simulation Parameter Settings

表 2 仿真参数设置

集合	α_i	γ_i	$c_{i,j}$	v_i	$\mu_{i,j}$	$\sigma_{i,j}$	$\lvert\Gamma_j\rvert$	$\theta_{i,j}$	M	N
Ⅰ	[0.2,0.3]	[0.2,0.3]	[0,1]	[20,25]	[0,1]	[1,2]	[1,5]	[0,0.1]	[2,512]	100
Ⅱ	[0.2,0.3]	[0.2,0.3]	[0,1]	[20,25]	[0,1]	[1,2]	[1,5]	[0,0.1]	100	[2,512]

此外,本文采用 CKKS 作为同态加密方案,因此,相关参数设置为:

维度参数 $N=2^{13}$,缩放因子 $\Delta=2^{40}$,密文模数 $q_L=2^{1503}$.同时,如 Cheon 等人在参考文献[47]中采用的多项式组合方法,本文也利用相同的方法,其中多项式 f_d 和 g_d 分别选取 $d=3$,而相应的表达式为

$$f_3(x)=(35x-35x^3+21x^5-5x^7)/2^4,$$

$$g_3(x)=(4\,589x-16\,577x^3+25\,614x^5-12\,860x^7)/2^{10}.$$

相应的组合参数分别为 d_f 和 d_g,其中 f^{d_f} 表示 $f\circ f\circ\cdots\circ f$ 做 d_f 次组合,同样 g^{d_g} 表示 $g\circ g\circ\cdots\circ g$ 做 d_g 次组合.

本实验环境为 Ubuntu 18.04.2 LTS, AMD Ryzen 7 5800H CPU,16 GB 内存,16 线程.

6.3 仿真结果

本文仿真实验中的考察指标——社会福利,为整个机制运行完后对 PFPI 的性能评价指标,另一个考察指标——真值估计准确度,为整个机制运行完后对 PATD 的性能评价指标.最后一个评价指标——运行时间,为整个机制运行完成后所需的时间,但是因为 PATD 模块只涉及同态加法操作,运行时间较短,而 PFPI 涉及到排序操作和大量乘法操作,运行时间较长.

图 3、图 4 所示为 PATD 的仿真结果.图 3 为数据收集者数量不同时,各数据真值估计机制的 *MAE*;而图 4 为数据需求者数量不同时,各数据真值估计机制的 *MAE*.其中,*MAE* 表示各数据的估计值与真值之间的平均距离,计算公式为

$$MAE=\frac{1}{N}\sum_{i,\tau_i\in T}|x_i-x_i^*|. \tag{34}$$

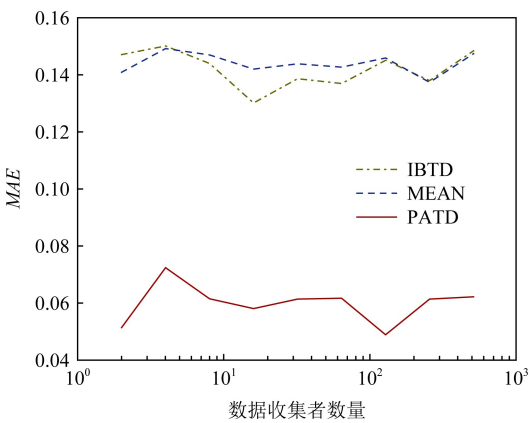

Fig. 3 *MAE* under different number of data collectors

图 3 不同数量的数据收集者时的 *MAE*

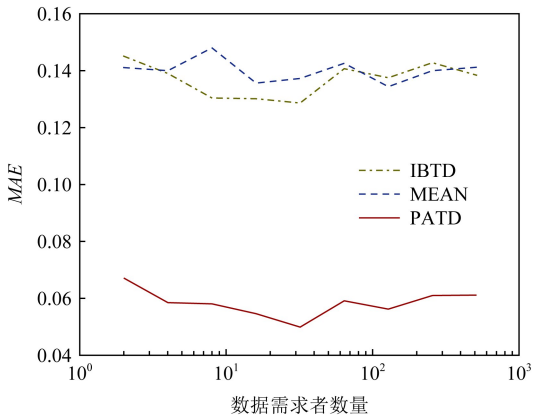

Fig. 4 *MAE* under different number of data requesters

图 4 不同数量的数据需求者时的 *MAE*

从式(34)可以知道,*MAE* 的值越小,表示算法的准确性越高.从图 3 和图 4 中可以看出,本文所提的 PATD 算法具有最小的 *MAE*,因此,在所有比较的数据真值估计机制中具有最好的性能表现.在对 IBTD 机制进行仿真时,其迭代次数被设置为 1 000 次.

本文所提的区块链群智感知系统中基于隐私保护数据的用户激励机制包括 2 部分,即 PATD 和 PFPI,在验证了 PATD 的性能后,接着对 PFPI 的性能进行验证.

图 5、图 6 所示为 PFPI 的仿真结果.图 5 为数据收集者数量不同时,各具有隐私保护的激励机制达到的社会福利;而图 6 为数据需求者数量不同时,各具有隐私保护的激励机制达到的社会福利.需要指出的是,其中 BFI 机制可以满足真实性和个体合理性,而 WFI 机制只满足个体合理性但不满足真实性,在进行仿真的过程中,BFI 和 WFI 中数据需求者和收集者提交的竞价都利用 CKKS 同态加密方案进行加密.从仿真结果可以看出 PFPI 达到的社会福利最高,这说明 PFPI 与其他基准激励机制相比具有更好的性能.同时,在图 5 和图 6 中可以看到,BFI 的性能相较于 WFI 的性能更好.

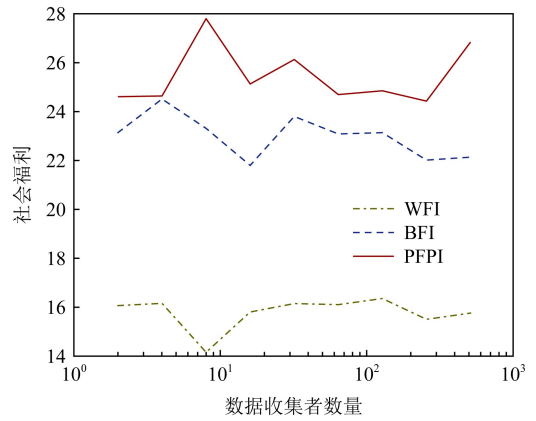

Fig. 5 Social welfare under different number of data collectors

图 5 不同数量的数据收集者时的社会福利

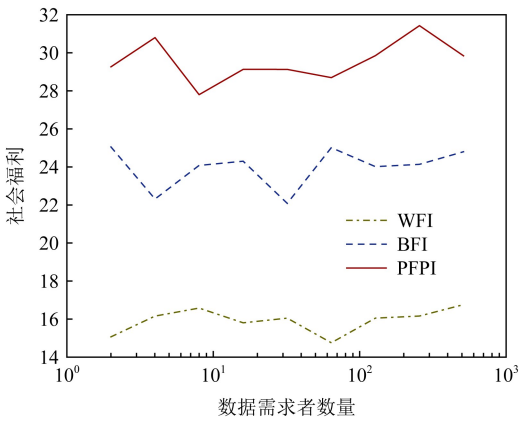

Fig. 6 Social welfare under different number of data requesters

图 6 不同数量的数据需求者时的社会福利

由于区块链每时每刻发生大量的交易,因此对于各种基于区块链的应用的运行速度有较高的要求,在性能仿真的最后,本节验证了利用本文所提区块链群智感知系统中基于隐私保护数据的用户激励机制的运行时间,并将仿真结果列于表 3 中.在表 3 中分别让(d_f,d_g)取不同的组合,并统计任务数从 2 增加到 512 时,不同任务数所需的运行时间.从表 3 中可以看出,随着任务数的增加,所需要的运行时间在增加,这是因为需要的乘法操作在增加.同时可以看出不同的(d_f,d_g)组合所需的时间不同,这是因为不同的组合需要计算的乘法操作次数是不同的,而且可以看出 d_f 的影响要比 d_g 的影响大.

Table 3 Running Time of Different Parameters

表 3 不同参数时的运行时间

任务数	不同(d_f,d_g)的运行时间/s		
	(4,7)	(3,6)	(3,7)
2	10.205	6.394	6.771
4	52.703	30.837	32.138
8	113.721	64.291	70.634
16	197.763	111.064	124.612
32	302.928	168.708	191.114
64	316.230	242.823	266.805
128	420.559	323.569	359.140
256	552.770	423.096	483.633
512	694.591	519.388	582.218

表3为PFPI的运行速度,为了整个实验的完整性,本文最后对所提机制整体进行运行时间的仿真,以此来论证机制的合理性与连续性.

从表4可以看出,PFPI的运行时间相较于PATD要长得多,这是因为PATD模块仅涉及同态加法操作以及少量的明文与密文的乘法操作,这些操作所需时间较短.与之相反,PFPI在用户选择阶段和报酬确定阶段均涉及到排序操作,而排序操作需要大量的密文之间的同态乘法计算来完成,而乘法操作需要时间较长.从表4中可以看到,所提机制对时间敏感度不高的在线系统或者是完全线下的系统操作性较强.

Table 4 Running Time of the Whole Scheme

表4 机制整体的运行时间

任务数	PATD运行时间/s	PFPI运行时间/s ($d_f=3, d_g=5$)	总时间/s
2	0.012	3.187	3.199
4	0.018	26.513	26.531
8	0.026	60.865	60.891
16	0.037	103.944	103.981
32	0.049	161.881	161.930
64	0.062	226.06	226.122
128	0.087	308.02	308.107
256	0.106	396.442	396.548
512	0.181	489.292	489.473

7 总结与展望

本文针对现有基于区块链群智感知系统的数据收集框架总是独立分离设计数据真值估计机制和参与者激励机制而导致其无法达到最佳性能的问题,提出一类新的具有隐私保护的数据收集机制.该机制由数据真值模块PATD和参与者激励模块PFPI共同构成.相较于分离设计的框架,该机制具有更好的性能.为了保护参与者的隐私信息,该机制利用CKKS同态加密方案.本文在提出数据收集机制后,从理论角度证明了PATD具有较高的真值估计准确度,同时证明了PFPI不仅满足真实性和个体合理性,而且具有较高的社会福利;从实验仿真的角度验证了PATD和PFPI的安全性能.从仿真结果可知,它们与各自的基准方案相比具有更好的安全性能.

作者贡献声明:应臣浩提出了算法思路和实验方案;夏福源负责完成实验仿真;李颉提出理论分析指导意见;斯雪明提出实验仿真指导意见;骆源提供论文修改意见.

参 考 文 献

[1] Peng Dan, Wu Fan, Chen Guihai. Data quality guided incentive mechanism design for crowdsensing [J]. IEEE Transactions on Mobile Computing, 2017, 17(2): 307-319

[2] Ying Chenhao, Jin Haiming, Wang Xudong, et al. Double insurance: Incentivized federated learning with differential privacy in mobile crowdsensing [C] //Proc of 2020 IEEE Int Symp on Reliable Distributed Systems. Piscataway, NJ: IEEE, 2020: 81-90

[3] Ying Chenhao, Jin Haiming, Wang Xudong, et al. CHASTE: Incentive mechanism in edge-assisted mobile crowdsensing [C] //Proc of 2020 17th Annual IEEE Int Conf on Sensing, Communication, and Networking. Piscataway, NJ: IEEE, 2020: 1-9

[4] Jin Haiming, Guo Hongpeng, Su Lu, et al. Dynamic task pricing in multi-requester mobile crowd sensing with Markov correlated equilibrium [C] //Proc of 2019 IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2019: 1063-1071

[5] Wang Xiumin, Wu Weiwei, Qi Deyu. Mobility-aware participant recruitment for vehicle-based mobile crowdsensing [J]. IEEE Transactions on Vehicular Technology, 2017, 67(5): 4415-4426

[6] Tian Feng, Liu Bo, Sun Xiao, et al. Movement-based incentive for crowdsourcing [J]. IEEE Transactions on Vehicular Technology, 2017, 66(8): 7223-7233

[7] Qu Yuben, Tang Shaojie, Dong Chao, et al. Posted pricing for chance constrained robust crowdsensing [J]. IEEE Transactions on Mobile Computing, 2020, 19(1): 188-199

[8] Han Kai, Huang He, Luo Jun. Quality-aware pricing for mobile crowdsensing [J]. IEEE/ACM Transactions on Networking, 2018, 26(4): 1728-1741

[9] Restuccia F, Ferraro P, Silvestri S, et al. IncentMe: Effective mechanism design to stimulate crowdsensing participants with uncertain mobility [J]. IEEE Transactions on Mobile Computing, 2019, 18(7): 1571-1584

[10] Jin Wenqiang, Xiao Mingyan, Li Ming, et al. If you do not care about it, sell it: Trading location privacy in mobile crowd sensing [C] //Proc of 2019 IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2019: 1045-1053

[11] Wang Liang, Yu Zhiwen, Han Qi, et al. Multi-objective optimization based allocation of heterogeneous spatial crowdsourcing tasks [J]. IEEE Transactions on Mobile Computing, 2018, 17(7): 1637-1650

[12] Jin Haiming, Su Lu, Chen Danyang, et al. Thanos: Incentive mechanism with quality awareness for mobile crowd sensing [J]. IEEE Transactions on Mobile Computing, 2018, 18(8): 1951-1964

[13] Karaliopoulos M, Koutsopoulos I, Spiliopoulos L. Optimal user choice engineering in mobile crowdsensing with bounded rational users [C] //Proc of 2019 IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2019: 1054-1062

[14] Zhang Yanru, Gu Yunan, Pan Miao, et al. Multi-dimensional incentive mechanism in mobile crowdsourcing with moral hazard [J]. IEEE Transactions on Mobile Computing, 2017, 17(3): 604-616

[15] Wang Jiangtao, Wang Yasha, Zhang Daqing, et al. Multi-task allocation in mobile crowd sensing with individual task quality assurance [J]. IEEE Transactions on Mobile Computing, 2018, 17(9): 2101-2113

[16] Gao Yi, Dong Wei, Guo Kai, et al. Mosaic: A low-cost mobile sensing system for urban air quality monitoring [C] // Proc of IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2016: 1-9

[17] Ganti R K, Pham N, Ahmadi H, et al. GreenGPS: A participatory sensing fuel-efficient maps application [C] // Proc of the 8th Int Conf on Mobile Systems, Applications, and Services. Piscataway, NJ: IEEE, 2010: 151-164

[18] Thiagarajan A, Ravindranath L, LaCurts K, et al. Vtrack: Accurate, energy-aware road traffic delay estimation using mobile phones [C] //Proc of the 7th ACM Conf on Embedded Networked Sensor Systems. New York: ACM, 2009: 85-98

[19] Eisenman S B, Miluzzo E, Lane N D, et al. BikeNet: A mobile sensing system for cyclist experience mapping [J]. ACM Transactions on Sensor Networks, 2010, 6(1): 1-39

[20] Gao Chunming, Kong Fanyu, Tan Jindong. Healthaware: Tackling obesity with health aware smart phone systems [C] //Proc of 2009 IEEE Int Conf on Robotics and Biomimetics. Piscataway, NJ: IEEE, 2009: 1549-1554

[21] Hu Yidan, Zhang Rui. Differentially-private incentive mechanism for crowdsourced radio environment map construction [C] //Proc of 2019 IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2019: 1594-1602

[22] Bhattacharjee S, Ghosh N, Shah V K, et al. QnQ: Quality and quantity based unified approach for secure and trustworthy mobile crowdsensing [J]. IEEE Transactions on Mobile Computing, 2018, 19(1): 200-216

[23] Han Kai, Liu Huan, Tang Shaojie, et al. Differentially private mechanisms for budget limited mobile crowdsourcing [J]. IEEE Transactions on Mobile Computing, 2018, 18(4): 934-946

[24] Gong Xiaowen, Shroff N B. Truthful mobile crowdsensing for strategic users with private data quality [J]. IEEE/ACM Transactions on Networking, 2019, 27(5): 1959-1972

[25] Jin Haiming, Su Lu, Xiao Houping, et al. Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems [J]. IEEE/ACM Transactions on Networking, 2018, 26(5): 2019-2032

[26] Lin Jian, Yang Dejun, Wu Kun, et al. A sybil-resistant truth discovery framework for mobile crowdsensing [C] //Proc of 2019 IEEE 39th Int Conf on Distributed Computing Systems. Piscataway, NJ: IEEE, 2019: 871-880

[27] Zhang Zhikun, He Shibo, Chen Jiming, et al. REAP: An efficient incentive mechanism for reconciling aggregation accuracy and individual privacy in crowdsensing [J]. IEEE Transactions on Information Forensics and Security, 2018, 13(12): 2995-3007

[28] Zhou Jun, Shen Huajie, Lin Zhongyun, et al. Research advances on privacy preserving in edge computing [J]. Journal of Computer Research and Development, 2020, 57 (10): 2027-2051 (in Chinese)
(周俊, 沈华杰, 林中允, 等. 边缘计算隐私保护研究进展 [J]. 计算机研究与发展, 2020, 57(10): 2027-2051)

[29] Luu L, Narayanan V, Zheng Chaodong, et al. A secure sharding protocol for open blockchains [C] //Proc of the 2016 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2016: 17-30

[30] Wang Chenxu, Cheng Jiacheng, Sang Xinxin, et al. Data privacy-preserving for blockchain: State of the art and trends [J]. Journal of Computer Research and Development, 2021, 58(10): 2099-2119 (in Chinese)
(王晨旭, 程加成, 桑新欣, 等. 区块链数据隐私保护:研究现状与展望[J]. 计算机研究与发展, 2021, 58(10): 2099-2119)

[31] Das P, Eckey L, Frassetto T, et al. FastKitten: Practical smart contracts on bitcoin [C] //Proc of the 28th USENIX Security Symp. New York: ACM, 2019: 801-818

[32] Zhang Peiyun, Zhou Mengchu. Security and trust in blockchains: Architecture, key technologies, and open issues [J]. IEEE Transactions on Computational Social Systems, 2020, 7(3): 790-801

[33] Zheng Peiling, Xu Quanqing, Zheng Zibin, et al. Meepo: Sharded consortium blockchain [C] //Proc of 2021 IEEE 37th Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2021: 1847-1852

[34] Huang Huawei, Yue Zhengyu, Peng Xiaowen, et al. Elastic resource allocation against imbalanced transaction assignments in sharding-based permissioned blockchains [J]. IEEE Transactions on Parallel and Distributed Systems, 2022, 33 (10): 2372-2385

[35] Aloufi A, Hu Peizhao, Song Y, et al. Computing blindfolded on data homomorphically encrypted under multiple keys: A survey [J]. ACM Computing Surveys, 2022, 54(9): 1-37

[36] Chillotti I, Gama N, Georgieva M, et al. Tfhe: Fast fully homomorphic encryption over the torus [J]. Journal of Cryptology, 2020, 33(1): 34-91

[37] Tian Yifan, Yuan Jiawei, Song Houbing. Secure and reliable decentralized truth discovery using blockchain [C] //Proc of 2019 IEEE Conf Communications and Network Security. Piscataway, NJ: IEEE, 2019: 1-8

[38] Wu Haiqin, Düdder B, Wang Liangmin, et al. Blockchain-based reliable and privacy-aware crowdsourcing with truth and fairness assurance [J]. IEEE Internet of Things Journal, 2022, 9(5): 3586-3598

[39] Huang Zhiyuan, Zheng Jun, Xiao Mingjun. Privacy-enhanced crowdsourcing data trading based on blockchain and stackelberg game [C] //Proc of 2021 IEEE 18th Int Conf on Mobile Ad Hoc and Smart Systems. Piscataway, NJ: IEEE, 2021: 621-626

[40] Zhang Can, Zhu Liehuang, Xu Chang, et al. PRVB: Achieving privacy-preserving and reliable vehicular crowdsensing via blockchain oracle [J]. IEEE Transactions on Vehicular Technology, 2021, 70(1): 831-843

[41] Xie Liang, Su Zhou, Chen Nan, et al. Secure data sharing in UAV-assisted crowdsensing: Integration of blockchain and reputation incentive [C] //Proc of 2021 IEEE Global Communications Conf. Piscataway, NJ: IEEE, 2021: 1-6

[42] Jin Haiming, Su Lu, Xiao Houping, et al. Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems [J]. IEEE/ACM Transactions on Networking, 2018, 26(5): 2019-2032

[43] Jin Haiming, Su Lu, Nahrstedt K. CENTURION: Incentivizing multi-requester mobile crowd sensing [C] // Proc of 2017 IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2017: 1-9

[44] Cheon J H, Kim A, Kim M et al. Homomorphic encryption for arithmetic of approximate numbers [C] //Proc of Int Conf on the Theory and Application of Cryptology and Information Security. Berlin: Springer, 2017: 409-437

[45] Hong S, Kim S, Choi J et al. Efficient sorting of homomorphic encrypted data with k-way sorting network [J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 4389-4404

[46] Zhang Chuan, Zhu Liehuang, Xu Chang et al. Reliable and privacy-preserving truth discovery for mobile crowdsensing systems [J]. IEEE Transactions on Dependable and Secure Computing, 2021, 18(3): 1245-1260

[47] Cheon J H, Kim D, Kim D. Efficient homomorphic comparison methods with optimal complexity [C] //Proc of Int Conf on the Theory and Application of Cryptology and Information Security. Berlin: Springer, 2020: 221-256

Ying Chenhao, born in 1993. PhD, assistant researcher. His main research interests include information security, privacy computing and blockchain.

应臣浩,1993 年生.博士,助理研究员.主要研究方向为信息安全、隐私计算和区块链.

Xia Fuyuan, born in 2000. PhD candidate. His main research interests include blockchain as well as data privacy security.

夏福源,2000 年生.博士研究生.主要研究方向为区块链和数据隐私安全.

Li Jie, born in 1959. PhD, endowed chair professor. Member of CCF. His main research interests include big data and AI, blockchain, network system and security, as well as smart city.

李　颉,1959 年生.博士,讲席教授.CCF 会员.主要研究方向为大数据和 AI、区块链、网络系统和安全、智慧城市.

Si Xueming, born in 1966. Master, professor. Member of CCF. His main research interests include blockchain and network security.

斯雪明,1966 年生.硕士,教授.CCF 会员.主要研究方向为区块链和网络安全.

Luo Yuan, born in 1971. PhD, professor. Member of CCF. His main research interests include information theory and blockchain.

骆　源,1971 年生.博士,教授.CCF 会员.主要研究方向为信息论和区块链.