群智感知中支持隐私保护的激励机制研究

梁 艳[1),2)] 安 健[2),3)] 胡先智[4)] 杨 倩[1)] 司海峰[1)]

[1)](西安思源学院理工学院 西安 710038)

[2)](西安交通大学计算机科学与技术学院 西安 710049)

[3)](西安交通大学陕西省计算机网络重点实验室 西安 710049)

[4)](西安理工大学网络信息管理中心 西安 710048)

摘 要 针对已有大多数研究在设计激励机制时未考虑用户的隐私泄露问题,本文提出一种支持隐私保护的激励机制综合方案 IMPP(Incentive Mechanism with Privacy-Preserving in mobile crowd sensing). 首先,基于轻量级隐私保护思想,采用单向安全哈希函数生成 256 位哈希值作为参与者的匿名身份标识,以此来保护参与者的身份隐私;其次,依据参与者的数据效用值、期望回报及感知任务预算实现面向数据质量的补偿激励,选择性价比最高的胜出者;接着,借助分布式压缩感知理论,对胜出者的原始感知数据压缩处理,得到剔除冗余的观测值,并在观测值中添加哈希函数值等噪扰数据后传送于服务器端聚合,以增强感知数据的隐私性保护,之后对隐私数据集进行完整性校验并重构;最后,利用真实数据集,通过仿真实验对 IMPP 的有效性进行对比分析. 实验结果表明,IMPP 机制在隐私保护水平、数据完整性、数据精确性、时间效率、评估准确率、重构匹配度及激励效果等方面是高效的.

关键词 群智感知;隐私保护;激励机制;哈希函数;分布式压缩感知

中图法分类号 TP309 **DOI 号** 10.11897/SP.J.1016.2020.02414

Research on Incentive Mechanism with Privacy-Preserving in Mobile Crowd Sensing

LIANG Yan[1),2)] AN Jian[2),3)] HU Xian-Zhi[4)] YANG Qian[1)] SI Hai-Feng[1)]

[1)](*School of Technology, Xi'an Siyuan University, Xi'an* 710038)

[2)](*School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an* 710049)

[3)](*Shaanxi Province Key Laboratory of Computer Network, Xi'an Jiaotong University, Xi'an* 710049)

[4)](*Center of Network Information & Management, Xi'an University of Technology, Xi'an* 710048)

Abstract Mobile crowd sensing (MCS) applications are becoming more and more widespread with the popularity of mobile smart devices. And it is a current research hotspot. MCS uses the mobile terminal equipment of ordinary users to carry out conscious or unconscious collaboration through the mobile Internet, so as to realize sensing tasks distribution and sensed data collection, and complete large-scale social sensing tasks. Accordingly, MCS needs a large number of users to participate in sensing tasks, and hopes that users can collect high-quality sensed data and provide better service quality to the public. However, during the user performing sensing tasks, the cost is relatively high because it not only consumes the user's time, but also consumes the user's equipment resources. In addition, sensed data collected by the user may imply sensitive information related to the user, e.g., identity information, living habits, health status and social relations, which results in the user to face the risk of personal privacy leakage. But privacy leakage

收稿日期:2019-11-01;在线发布日期:2020-09-30. 本课题得到国家重点研发计划课题(2018YFB1800304)、国家自然科学基金(61502380)、陕西省重点研发项目(2020GY-033,2019GY-005,2017ZDXM-GY-011)、陕西省教育厅专项科学研究计划项目(19JK0686)资助.**梁 艳**,硕士,讲师,主要研究方向为群智感知. E-mail:21943748@qq.com. **安 健**(通信作者),博士,高级工程师,中国计算机学会(CCF)会员,主要研究方向为群智感知、物联网. E-mail:anjian@mail.xjtu.edu.cn. **胡先智**,硕士,工程师,主要研究方向为物联网、信息安全. **杨 倩**,硕士研究生,主要研究方向为群智感知. **司海峰**,硕士,高级工程师,主要研究方向为群智感知.

will hinder the user' enthusiasm to participate. Therefore, we studied the comprehensive mechanism of privacy-preserving and compensation incentive to MCS. On the one hand, it can reduce users' concerns about their own privacy leakage, and on the other hand, it can compensate users for the cost of participating in sensing tasks, so as to mobilize the enthusiasm of users and attract more users to participate in the sensing tasks for a long time. It is of great significance to further expand MCS application. In the existing research, the research on the comprehensive scheme or the integrated system framework that combines privacy-preserving and incentive mechanism is still a minority. Many studies usually focus on the user's unilateral privacy-preserving or compensation incentive, and these studies don't fully consider the communication consumption caused by the transmission of a large number of redundant sensed data, and the risk of the user's privacy leakage under the incomplete trusted sensing platform. Yet these problems have a direct impact on the long-term development of MCS system, and should be paid close attention to. Aiming at these problems, we proposed a comprehensive incentive mechanism with privacy-preserving (IMPP) in MCS. we carried out our research work from three stages: participant registration, winner selection, and sensed data aggregation. First, based on the idea of lightweight privacy-preserving, IMPP uses a one-way hash function to generate a 256-bit hash value as an anonymous identity of the participant, so as to protect the identity privacy of the participant; Secondly, according to the participant's data utility value, expected reward and the sensing task total budget, it has achieved the compensation incentives for data quality and chose those winners with the best buys; Next, with the help of distributed compressed sensing theory it has compressed the winner's original sensed data and removed redundant data to obtain the observed value. By adding noise data such as hash values to the observed value, the privacy-preserving of sensed data can be enhanced. After the winner's privacy sensed data is aggregated at the application server (AS), AS verifies the integrity of privacy data sets and reconstructs the original sensed data; Finally, using the real data sets, the effectiveness of IMPP is compared and analyzed through simulation experiments. The experimental results show that the IMPP mechanism is efficient in terms of privacy-preserving level, data integrity, data accuracy, time efficiency, evaluation accuracy, reconstruction matching degree and incentive effect.

Keywords　mobile crowd sensing; privacy-preserving; incentive mechanism; hash function; distributed compressed sensing

1　引　言

随着移动智能设备的爆炸式普及，群智感知应用越来越广泛，已成为当前的研究热点[1-3]. 移动群智感知(Mobile Crowd Sensing，MCS)是借助用户的移动终端等通信设施作为认知单元，通过移动网络直接或间接地协作，来完成大规模的各类感知任务的数据采集[4]. MCS 系统需要大量用户主动参与感知任务，并希望用户能够采集高质量的感知数据，为公众提供更好的服务质量，然而用户在参与感知任务时所花费的代价较为高昂，因为它不仅消耗用户的时间，还消耗用户设备的电量、运算、存储、通信等系统资源[5-7]. 此外，用户在采集数据时，可能不经意间将自身的姓名、性别、身份证号等敏感隐私潜入感知数据中，导致用户隐私信息的泄露[8-9]. 然而隐私泄露又会消弱用户主动参与感知任务的热情[10]. 因此，面向群智感知系统，研究在保护用户隐私的前提下激励用户采集高质量数据的机制，一方面减少用户对自身隐私泄露的顾虑，另一方面补偿用户执行感知任务所付出的高昂成本，以激发用户长期参与的热性和积极性，对于进一步拓展群智感知应用规模具有重要意义.

现有研究中，将群智感知系统的隐私保护或激励机制问题单独研究的成果较多. 如文献[5-7]是研究激发用户参与感知任务的激励机制，并未对用户

数据隐私保护问题加以考虑;文献[10-12]是对用户的数据或位置等隐私信息保护展开研究,也不涉及补偿激励问题.

也有少数学者对兼顾隐私保护和激励机制的综合方案或一体化系统框架展开了研究,如文献[13-14].文献[13]提出了一种基于用户偏好的诚实激励机制,在保护隐私的同时根据用户对整体服务准确性的个人贡献对其支付,以激励用户提供真实数据.文献[14]提出了一种新型的移动群智感知系统框架,它融合了激励、数据聚合和扰动机制等.其激励机制选择那些更有可能提供可靠数据的用户,并对他们的感知成本给予补偿;其数据聚合机制结合了用户的可靠性来生成高精度的聚合结果;其数据扰动机制减弱了用户对隐私泄露的顾虑,提高了用户满意度和最终扰动结果的理想精度.这些研究成果对于本文研究基于隐私保护下的激励机制问题有着重要的指导意义.

前人的研究成果虽然在隐私保护或激励机制问题上取得了较好的效果,但仍然存在被普遍忽略的问题:

(1)资源受限移动终端上的计算效率不容忽视.通常情况下,群智感知平台是半可信的,而在已有的一些隐私保护策略和方法中,常用同态加密等重量级密码算法,实现对匿名用户的隐私数据聚合,导致计算能力有限的感知端面对庞大的数据量,计算效率与实际要求存在差距[15].

(2)冗余感知数据不应随原始感知数据一起传输.群智感知系统中存在海量的用户感知数据,而应用中通常只关注有效数据信息,并不需要大量的冗余数据.若每个用户都将未预处理的原始数据传输到系统服务器,那么大量的冗余数据传输将会极大地增加感知网络的通信开销[16].

(3)激励机制不应忽视感知数据的质量.群智感知系统中非常重要的一个方面就是高质量数据服务[1],感知数据质量对于群智感知技术走向大规模应用起着关键作用.若忽略了数据质量,可能导致提供低质量数据的用户仍会获得奖励报酬,从而影响采集高质量数据用户的奖励,长期以来造成诚信用户数量的减少.

以上这些方面都是本文的研究重点并加以解决的问题.

针对已有大多数研究在设计激励机制时未考虑用户的隐私泄露问题,本文提出群智感知中支持隐私保护的激励机制 IMPP(Incentive Mechanism with Privacy-Preserving in mobile crowd sensing).先基于轻量级加密哈希函数技术,对用户的身份进行匿名处理,以支持用户匿名参与感知任务;任务执行过程中,利用面向数据质量的激励措施来激发用户提高数据质量;竞价获胜的用户需要上传数据到感知平台,为了保护感知数据隐私,借助分布式压缩感知理论及方法,得到剔除冗余数据的压缩观测值,并在观测值中加入哈希函数值等噪扰数据以增强隐私性保护.

基于以上基本思想,本文从参与者注册、胜出者选择、感知数据聚合三个阶段展开具体研究.首先,基于轻量级隐私保护思想,采用单向安全哈希函数生成 256 位哈希值作为参与者感知任务的匿名身份标识,以此来保护参与者的身份隐私;然后,感知平台依据参与者的数据效用值、期望回报及感知任务预算进行参与者最优选择,且数据效用值由感知数据的时间因子、地点因子和数据精确率多种因素决定,并借助消错决策理论的评估方法实现面向数据质量的感知激励,且平台对竞价失败者也给予适当补偿,以此达到感知平台长期维持大量参与者的目的;接着,利用分布式压缩感知方法构造观测矩阵,胜出者通过观测矩阵对原始感知数据合理表示和压缩,实现冗余数据的剔除,并在所得的压缩观测值中添加哈希函数值等噪扰数据,之后传输于服务器端聚合,以增强胜出者感知数据的隐私性保护;最后,服务器对聚合数据进行完整一致性校验,并对校验合格的数据通过最优范数算法二次恢复(重构).

本文的主要贡献有 4 个方面:

(1)对用户身份和感知数据的隐私保护进行研究.针对资源受限的移动终端上使用重量级等密码导致计算开销过大问题,提出一种基于单向安全哈希函数的轻量级隐私保护方法.该方法既增强了数据的隐私安全,又提高了感知端的计算效率,同时,利用哈希函数的单向性解决了数据完整性校验问题.

(2)为了降低隐私保护中感知数据的通信开销,剔除冗余数据,对分布式压缩感知技术进行研究.基于正交对称托普利兹矩阵(Orthogonal Symmetric Toeplitz matrices,OST),提出一种优化型确定性观测矩阵,通过阈值迭代收缩法来降低观测矩阵与稀疏变换基之间的互相关性,从而提高了感知数据的重构准确度.

(3)针对激励机制往往忽视感知数据质量控制问题,对感知数据质量的评估方法进行研究.引入消

错决策理论评估方法，设计了由时间因子、地点因子及数据精确率多因素决定的数据效用值评估指标，并引入权值因子，以适应不同情形下的数据质量需求. 为了实现有限任务预算下合理分配资源，提出了基于数据效用值的参与者最优选择策略，保障面向数据质量的补偿激励.

(4) 针对现有研究大多仅关注群智感知的隐私保护或补偿激励单方面问题，本文提出一种群智感知中支持隐私保护的激励机制 IMPP，旨在解决用户的隐私泄露问题，同时促进用户主动且高质量地完成感知任务. 通过安全性分析及实验验证，IMPP 机制有效可行.

2　相关工作

通过分析近十年来领域代表性论文[17-25]可以看出，对群智感知系统中隐私保护问题的研究主要关注三类方法[17]：

(1) 基于匿名方法的隐私保护. 该类方法的基本思想是采用匿名方式保护用户的身份等信息，用户的真实身份和他(她)的敏感信息不能通过其它手段进行关联. 以此实现用户信息的隐私保护[18]. 此类方法最具代表性的是 Sweeney 提出的 K-匿名方法[19]. 近年来匿名方法受到广大学者的关注和研究. Wang 等人[10]通过构造 n 个互不链接的等价类，使用迭代法在等价类中对数据进行迭代，依次确保每个用户的数据都被上传到平台，实现了感知数据的隐私性保护. Zhou 等人[20]在传统模型中增加实时的数据领导者得到改进模型，并提出匿名化数据收集协议，确保无可信第三方前提下，用户利用 K-匿名技术可达到效用最大化的数据采集，使得用户的隐私泄露概率得到有效降低. Zhang 等人[21]提出针对多敏感属性的个性化 K-匿名算法，通过在等价类中抑制敏感属性值使其分布更加均衡来满足多样性的要求，达到保护敏感数据的隐私. 这类方法虽然可以通过泛化或抑制的方式保护个人敏感信息，但是无法抵抗背景知识攻击，而且隐私数据的信息损失量较大，影响了数据的效用.

(2) 基于数据扰动的隐私保护. 数据扰动的基本原理是通过扰动技术(如随机噪声、变换技术等)来隐藏真实的原始数据，使原始数据只呈现出其统计学特征. Ganti 等人[22]设计了一个群智感知应用系统 PoolView. 该系统利用一个近似模型来生成与真实数据集特性类似的噪声模型，用户根据该噪声模型通过调整参数在本地生成噪声并加入到感知数据中，使得攻击者重构用户感知数据变得困难. 但是当所有用户都采用相同的噪声分布时，攻击者可以从扰乱数据中获取隐私信息. 对此，Zhang 等人[23]提出一种增强隐私的状态相关扰动方法，为每个用户分配不同的噪声分布，并且噪声随真实数据的状态而变化，该方法能够在恶意者获取扰动数据情况下为用户提供更好的隐私保护. 这类方法虽然通过加入随机噪声实现了原始数据的有效隐藏，但实际情况中很难合理控制噪声量，并且噪声的添加往往使得聚合后的数据可用性降低.

(3) 基于数据加密的隐私保护. 数据加密隐私保护是借助公钥加密、同态加密、安全多方计算等技术将感知数据进行处理后再上报给感知平台. Cristofaro 等人[24]采用了对用户身份加密的隐私保护方法. 该方法不必绑定固定的用户公钥，而是将与用户有关的身份信息直接作为用户的公钥，诸如身份证号码、手机号码、邮箱地址等. 私钥则由可信的私钥生成器管理和颁发，从而实现用户和感知平台之间安全有效地通信. Xiong 等人[25]首次提出一种新的角色对称加密算法(rse)和一种基于 rse 的所有权证明方案(rse-pow)，能够防止多媒体数据隐私泄露和抵御侧信道攻击. 这类方法能够对数据进行安全验证，具有高强度的隐私保护性能，但不足是重量级密码运算会产生较大的计算和通信开销，导致资源有限的移动终端负载过大. 此外，还需生成和维护多个密钥，缺乏灵活性.

隐私保护方法不仅要注重安全性，还应兼顾感知数据聚合时的数据完整性、可用性和通信开销. 目前围绕群智感知中数据聚合问题的研究成果相对较少[10]. 如何在用户隐私保护的前提下实现感知数据的有效聚合成为当前的研究热点. Zhang 等人[26]提出基于保密混淆的节能隐私保护数据聚合算法 CESPT，通过逐跳方式，将随机伪对与真实数据混淆，并对混淆数据分组加密，在设计的时间片内传送给邻居节点，实现了数据机密性保护，减少了通信能耗和消息冲突. 但未关注数据完整性保护. Zhou 等人[27]提出一种隐私数据融合方法 ICKPDA，可进行数据完整性检测. 该方法兼顾了数据的隐私性和完整性，但由于使用了密钥噪声和数据分片重组，导致产生了较大的通信开销. Wu 等人[28]借助分布式压缩感知思想，通过分析不同样本数据间的时空关联性，设计了一种具有良好网络性能，且保持较低通信开销和计算复杂度的隐私保护方法 DCSPDA，但未

考虑数据的完整性检测.

若要实现移动群智感知技术在各领域中长期大规模应用,参与者的感知数据质量是关键.因此,群智感知系统在保护用户隐私的同时,如何设计激励机制以激发更多用户积极采集高质量的感知数据,这也是当前值得学术界迫切关注的研究课题.Lee等人[29]在群智感知激励机制研究中最先引入逆向拍卖思想,提出一种参与式感知的动态价格逆向拍卖激励机制 RADP.该机制以竞价最低者作为支付报酬的获胜者,并引入虚拟积分来避免竞价频繁失败的用户退出,实现了任务预算和参与率之间的平衡,但未约束用户采集的数据质量.Han 等人[30]研究的拍卖激励机制,是由用户结合个人所付出的感知成本来竞争参与任务,任务完成后由感知平台在预算范围内向用户支付报酬.该机制同样未涉及用户感知数据质量的评估.Nan 等人[31]为了提高任务感知用户的参与积极性,提出一种动态激励模型 CSII,在群智系统、任务需求者和用户之间通过跨空间数据分析进行多元交互,以激励用户积极参与,从而提高任务完成率和数据质量.以上这些研究成果通过不同的方式对用户进行了补偿激励,提高了用户参与度,然而,他们并未关注用户的隐私泄露问题.

综上所述,现有研究大多集中在群智感知的隐私保护或补偿激励单方面问题,对于在激励机制中兼顾隐私保护的综合方案或整体系统框架的研究还非常少.为此,本文与前人不同,提出支持隐私保护的综合激励方案,以期减少用户的隐私泄露顾虑,促进用户积极参与感知任务,并对提交高质量感知数据的用户优先奖励补偿,从而吸引更多优质用户,使得群智感知系统应用能够进一步稳固发展,也为现有领域的研究成果作出贡献.

3 系统模型、威胁模型与问题分析

3.1 系统模型

本文面向通用的群智感知系统,该系统主要由参与者、胜出者、应用服务器部分组成.

(1)参与者(Workers)

参与者是具有感知能力的移动群体,其组成了群智感知的基本单元,主要完成环境、事件、健康等方面的多种类型数据的采集.

(2)胜出者(Winners)

胜出者是由应用服务器从注册成功的参与者集合中,根据平台的激励机制选择出的参与者子集,这些人群需要上传感知数据到应用服务器.

(3)应用服务器(Application Server,AS)

应用服务器通过感知平台的消息窗口与参与者实现消息交互,并聚合胜出者上传的感知数据,为任务需求者提供服务响应.需要注意的是,应用服务器半可信.也就是说,一方面,应用服务器会根据需求者的请求为其提供及时的服务响应;另一方面,应用服务器可能会有意或无意泄露参与者的身份等敏感信息.

图 1 描述了群智感知系统的一般工作流程.首先,Workers 需主动向 AS 申请匿名身份注册,注册后选择其感兴趣的任务并根据任务属性信息到达目的地去执行(①).其次,完成感知任务的 Workers 可凭借注册的匿名身份标识向 AS 提交他们的竞价 *Bid*(即效用回报)来参与竞争(②).然后,AS 根据参与者最优选择策略筛选出“性价比”最高的 Winners(③).接着,Winners 将经过隐私保护处理后的感知数据上传给 AS,AS 对接收到的感知数据进行聚合、验证并重构(④).最后,AS 结合任务预算为 Winners 分配相应的回报,并对在竞价中失败的 Workers 给予一定补偿(⑤).

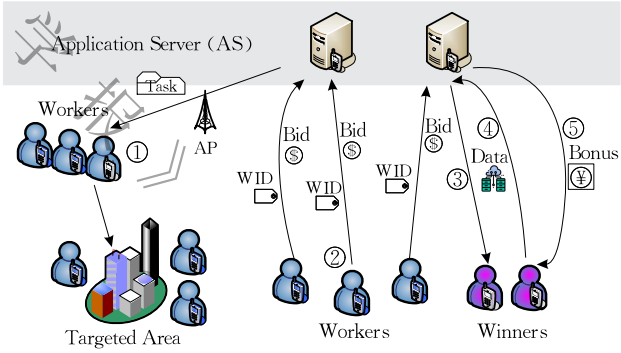

图 1 群智感知系统的工作流程

3.2 威胁模型

群智感知系统可能会面临内部和外部的攻击.外部攻击者可以监视通信,以窃听参与者的相关活动信息.他们还可以通过提交未经授权的数据或重放善意参与者的数据来操纵平台的信息收集.通常,可以使用传统的加密机制来削弱这些攻击,以保证数据通信的安全性,例如在通信实体之间建立端到端的安全传输层协议(TLS).外部攻击者也可以通过发起 DoS 攻击来瞄准系统的可用性,但本文假设这些攻击是由网络运营商处理的,因此不列入本研究范围之内.此外,为了防止基于网络标识符的客户端去匿名化,本文假设存在 TOR 匿名网络.

另一方面，内部攻击者可能是瞄准竞价或奖励过程的恶意参与者，也可能是群智感知系统实体. 例如，恶意参与者为了达成个人利益去获取其他参与者的隐私，他们可通过对目标参与者进行侧面分析并与系统中的其他实体合谋来获取目标参与者的身份或数据. 针对这种攻击，可采用安全加密哈希函数来切断参与者真实身份与其感知数据之间的链接性. 另外，恶意参与者还可能通过伪造感知数据或数据效用值的方式来骗取报酬，对此，感知平台可通过验证数据质量的方式加以解决. 本文假设平台对重构的感知数据验证数据质量，以校验其数据效用值是否与实际数据质量相匹配.

3.3　问题分析

面对参与者身份、感知数据隐私泄露和内外攻击的情况，如何抵御攻击以减少隐私泄漏，保护参与者身份及数据的安全，并激励更多参与者积极参与任务并能提高感知数据质量，IMPP应满足以下特性需求.

3.3.1　隐私安全需求

(1) 参与者身份隐私

参与者的真实身份信息关系到他们的切身利益，为了减少参与者对身份隐私泄露的顾虑，群智感知系统应注重身份隐私保护. 对身份进行匿名化处理是一种有效的解决方法，可实现参与者匿名参与感知任务，阻断其身份隐私与竞价等相关活动的联系. 本文对参与者身份信息采用安全加密哈希函数加以隐匿处理，并借助公钥加密的数字签名技术实现 AS 对匿名身份的认证. 相较前人的研究成果(如同态加密、安全多方计算等重量密码方案)有所不同，这是一个安全的轻量级隐私保护方法.

(2) 感知数据隐私

在群智感知网络中，由于参与者通常是利用移动终端设备通过无线网络来上传感知数据，而开放式的通信链路很容易被攻击者窃听并获取到数据. 另外，在应用服务器半可信的情况下，显式的感知数据会使参与者面临数据隐私泄露的巨大风险，从而关联的身份隐私信息也被泄露. 这就要求感知数据必须加以隐秘. 本文基于分布式压缩感知理论及方法，对原始感知数据压缩观测并添加噪扰数据后加密传输，从而保护感知数据在链路中的传输以及 AS 端的隐秘性.

3.3.2　激励需求

在群智感知系统中建立激励机制，一方面是为了激发参与者的兴趣度，使其自愿持续参与感知任务；另一方面是为了在有限任务预算下收集到足够的高质量数据，解决资源合理分配问题. 因此，激励机制与高质量数据收集密不可分. 这就要求激励机制应能够对数据质量进行评估. 本文采用数据效用值来衡量参与者的感知数据质量. 为了控制预算，最大化平台与参与者的利益，参与者选择策略尤为重要，以便筛选出“性价比”最高的参与者. 激励机制还应考虑平台需长期维持一定数量的参与者，所以需对竞价失败者也要给予一定补偿，以此来避免竞价多次失败的参与者在后续感知任务中流失.

3.3.3　性能需求

除了上述安全需求外，方案还应考虑群智感知系统的通信开销等性能. 因此，还需考虑以下问题：

(1) 存储：应最小化移动设备中存储的数据量.

(2) 通信：应最小化参与者和 AS 之间交换的消息量.

(3) 可扩展性：AS 应能够处理大量请求.

4　IMPP 机制的设计

基于第 3 节的分析，本节详细阐述 IMPP 的设计过程.

4.1　参与者注册

参与者执行感知任务前需主动向 AS 申请身份注册. 为了保护参与者的身份隐私，IMPP 支持匿名注册，即先由参与者向 AS 发送一个虚拟身份标识 $BidderID$，经过 AS 验证后，最终 AS 向参与者返回一个经它签名的虚拟身份标识 $\overline{BidderID}$，以此替代真实的参与者身份，参与者可用 $\overline{BidderID}$ 参与感知任务，并与 AS 进行信息交互. 对于 AS，它并不知道真实的参与者身份信息；对于攻击者，也无法由 $\overline{BidderID}$ 推断出真实身份. $\overline{BidderID}$ 是经过 AS 签名认证的，确保了 AS 承认其有效性，以便参与者在后续的胜出者竞争、感知数据提交及奖励分配阶段能够以合法的匿名身份参与相应活动.

假设感知平台上存在“消息窗口”能够支持 AS 与参与者之间进行信息交互. AS 具有在“消息窗口”上发布、修改或删除消息的权限，其发布的消息即刻生效，对于任何参与者均可见，参与者对消息具有监督权. AS 的公钥 e_{AS} 通过“消息窗口”发布，各参与者获知 e_{AS} 后，可以用 e_{AS} 将消息加密发送给 AS，并用

它验证 AS 签名返回的消息. 而 AS 的私钥 d_{AS} 则由 AS 严密保管. 由于需要为每个参与者生成唯一的虚拟身份标识,且该标识具有隐匿参与者真实身份的高度安全性,所以 IMPP 采用安全加密哈希函数 SHA-256($Hash()$)来实现. 该函数具有单向性,即由函数值(哈希值)很难反向推导出自变量(输入),但很容易实现将任意长度的自变量变换成定长的哈希值(输出). 它具有以下 4 个特性:

(1) 对于任意给定的消息,很容易计算出哈希值.

(2) 难以由一个已知的哈希值推导出原始消息.

(3) 对于两个不相同的消息,得出相同哈希值的可能性极小.

(4) 由它计算得到的哈希值和原始消息之间没有任何联系.

换言之,对于 $Hash()$,要找两个不同的 x_1 和 x_2,使得 $Hash(x_1)=Hash(x_2)$ 在计算上是困难的. 因此,用 $Hash()$ 生成 256 位哈希值作为参与者的虚拟身份标识是非常安全的. AS 最终将签名认证的 $\overline{BidderID}$ 发布于"消息窗口",来宣告各参与者的 $\overline{BidderID}$ 已经生效,即参与者在 AS 端匿名注册成功.

对于参与者的公钥问题,这里采用临时公钥. 由于公钥加密体制中通信双方使用数字证书(绑定了用户身份)获取对方公钥的方式会暴露用户的身份. 为了避免参与者身份泄漏,IMPP 允许参与者在每次感知任务时随机选择一个与身份不相关的一次性临时公钥 k,用 k 作为 $Hash()$ 的输入来得到 $BidderID$. 参与者的私钥 k^{-1} 则由私钥生成器颁发和管理,由参与者个人秘密保存. 本文假设参与者的临时公钥 k 通过 TLS 信道传输给 AS.

定义 1. 如果实体 X 和 Y 是通信双方,X 是发送方,Y 是接收方,则 $E_X(m)$ 表示 X 用 Y 的公钥对发送消息 m 的加密算法;$D_Y(m)$ 表示 Y 用自身私钥对接收到的消息 m 的解密算法.

定义 2. 如果实体 X 和 Y 是通信双方,X 是发送方,Y 是接收方,(e,d) 是 X 的公私密钥对,则 $Sig_{X,d}(m)$ 表示 X 用私钥 d 对消息 m 创建的数字签名 σ,即 $\sigma=Sig_{X,d}(m)$;$Sig.verify_{Y,e}(m,\sigma)$ 表示 Y 用 X 的公钥 e 对消息 m 的数字签名 σ 解密验证的结果(真或假).

基于轻量级隐私保护思想及安全加密哈希函数技术,本文提出参与者身份匿名保护算法如算法 1 所示.

算法 1. Identity Anonymity Protection.

输入:参与者 w 的一次性临时公钥 k

输出:AS 签名认证的 $\overline{BidderID}$

1. w: $h=Hash(k)$,即 $h=BidderID$;
2. w: $\sigma=Sig_{w,k^{-1}}(h)$;
3. w: $h^*=E_w(h,\sigma)$,其中,h^* 为 h 的密文;
4. $w\rightarrow$AS: h^*;
5. AS: $h=D_{AS}(h^*)$;
6. AS: IF $Sig.verify_{AS,k}(h,\sigma)=$true THEN
7. AS$\rightarrow w$: $E_{AS}(\bar{h},\bar{\sigma})$,其中,$\bar{\sigma}=Sig_{AS,d_{AS}}(h)$;
8. ELSE
9. AS: Discard h;
10. ENDIF
11. w: $D_w(E_{AS}(\bar{h},\bar{\sigma}))$ and $Sig.verify_{w,e_{AS}}(\bar{h},\bar{\sigma})$;
12. AS broadcasts $\bar{h}$,即 $\bar{h}=\overline{BidderID}$.

算法 1 中,参与者首先通过 $Hash(k)$ 计算出哈希值 h,即 $BidderID$,然后对其数字签名并加密发送给 AS(步骤 1～4);AS 对收到的密文 h^* 进行解密得到明文 h(步骤 5);接下来再对 h 的数字签名 σ 进行验证(步骤 6);如果验证正确,则 AS 再次对 h 数字签名,签名后的 h 记作 $\bar{h}$,并向参与者返回加密数据包(步骤 7);否则,AS 丢弃 h(步骤 8～10);参与者收到 AS 返回的消息后,同样经过解密、验证数字签名(步骤 11);最后,AS 把它签名认证的 $\overline{BidderID}$ 发布于"消息窗口",宣告参与者注册成功(步骤 12).

综上所述,算法 1 实现了参与者的匿名注册,保护了参与者的身份隐私,并为其授权一个经过 AS 签名的虚拟标识 $\overline{BidderID}$,为参与者后续参与竞价等活动提供了有效凭证.

4.2 胜出者选择

当已注册的参与者完成感知任务后,他们凭借虚拟身份标识可在平台规定的时间周期内向 AS 发起竞价. 假设大多数参与者是诚实的,他们能够依据自身的感知数据质量提出合理竞价. 基于各参与者的竞价,AS 根据面向数据质量的参与者最优选择策略,筛选出"性价比"最高的前 n 个胜出者. 本文采用数据效用值代表感知数据的质量.

4.2.1 数据质量评估指标

影响数据质量的因素较为复杂,目前数据质量的评估仍是具有挑战性的问题. 不同类型的数据衡量其数据质量的方式也不同[31]. 本文假设感知数据类型是数值型数据,考虑到感知数据的质量会受时

效性、地域性及精确性等因素的影响，为此，IMPP从时间因子、地点因子和数据精确率三个指标来衡量感知数据质量.

(1) 时间因子(T_w)

假设任务开始时间为s_task，结束时间为e_task.则对于参与者，其采集数据的有效时间为$t\in[s_task, e_task]$，那么影响数据质量的时间因子T_w可以用函数$k(x)$[31]来刻画，函数表达式如式(1)所示，函数模型如图2所示.T_w的值会随着有效时间t值的变化而取值于不同的数值范围.如果$t<s_task$，则$x=s_task-t$，T_w为函数$k(s_task-t)$，且$k(s_task-t)$的值随着t值减小逐渐递减并趋近于0；如果$s_task\leq t\leq e_task$，则$x=t-e_task$，T_w为函数$k(t-e_task)$，且$k(t-e_task)$的值为1；如果$t>e_task$，则$x=t-e_task$，T_w为函数$k(t-e_task)$，且$k(t-e_task)$的值随着t值增大逐渐递减并趋近于0.由此可见，随着参与者采集数据时间的提前或延后，时间因子呈现负增长，从而影响数据质量.

$$k(x)=2\times sgn(x)\times S(-x)+sgn(-x) \quad (1)$$

其中，$sgn(x)$为符号函数，函数表达式为

$$sgn(x)=\begin{cases}1, & x>0\\ 1/2, & x=0,\\ 0, & x<0\end{cases}$$

$S(x)$为Sigmoid函数，函数表达式为

$$S(x)=\frac{1}{1+e^{-x}}.$$

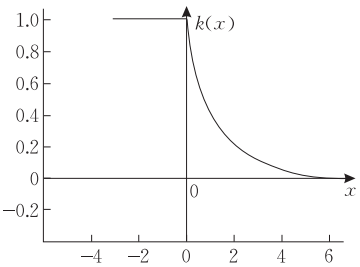

图2　$k(x)$函数模型

(2) 地点因子(P_w)

地点因子P_w对数据质量的影响也可以用$k(x)$函数来反映.感知数据采集区域如图3所示，图中每个方格代表一个区域，用区块数sl表示参与者采集数据的区域和任务目标区域的最短间距[31].例如，灰色区域为任务目标区域，三角形区域和五角星区域都是参与者选择的数据采集区域，则三角形区域的区块数为1，五角星区域的区块数为2.地点因子P_w对应的$k(sl)$函数如式(2)所示.显然，随着区块数sl的增大，数据质量受地点因子P_w的负影响也越大.

$$k(sl)=2\times sgn(sl)\times S(-sl)+sgn(-sl) \quad (2)$$

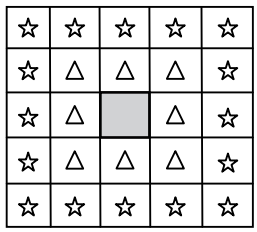

图3　感知数据采集区域

(3) 数据精确率(D_w)

假设一个参与者的所有数据集合为$\Gamma=\{d_1, d_2, \cdots, d_n\}$.对于每个数据组$I$可以计算任意成员$d_i$到其所在组中心$I(d_i).center$的距离$\Delta d_i$，即$\Delta d_i=d_i-I(d_i).center$.则每个数据$d_i$的精确率可由$1-\frac{\Delta d_i}{MaxDis(I(d_i))}$得到，进而可得出一个参与者的总体数据精确率$D_w$为该参与者所有数据精确率的平均值.计算如式(3)所示，$D_w\in[0,1]$.

$$D_w=\frac{\sum_{i=1}^{|\Gamma|}\left(1-\frac{d_i-I(d_i).center}{MaxDis(I(d_i))}\right)}{|\Gamma|} \quad (3)$$

4.2.2　数据效用值

基于时间因子、地点因子和数据精确率三个评估指标，借助消错决策理论[32]，对低质量感知数据依据误差损失值加以剔除，并对符合任务需求的数据排序，以更好地识别与分类数据质量.

假设有m个参与者，其感知数据集合设为$U=\{u_1, u_2, \cdots, u_m\}$且$|U|=m$，数据质量测评参数集合设为$Z=\{T_w, P_w, D_w\}$且$|Z|=3$.决策矩阵为$\boldsymbol{G}=[g_{i,j}](i=1,2,\cdots,m; j=1,2,3)$，$g_{i,j}$是参与者感知数据$u_i$对应各质量测评参数的属性值.用$v_j^{\min}$表示质量测评参数的下限值，用$v_j^{\max}$表示质量测评参数的上限值.判断参与者感知数据是否存在错误，可利用评估规则在各质量测评参数下进行衡量.根据误差值，来统筹排序参与者数据.误差值$l_{i,j}(l_{i,j}\in[0,1])$的计算如式(4)所示.

$$l_{i,j}=\begin{cases}\rho, & g_{i,j}<v_j^{\min}\text{或}\ g_{i,j}>v_j^{\max}\\ \dfrac{v_j^{\max}-g_{i,j}}{v_j^{\max}-v_j^{\min}}, & v_j^{\min}\leq g_{i,j}\leq v_j^{\max}\end{cases} \quad (4)$$

当属性值$g_{i,j}$低于下限值$v_j^{\min}$或高于上限值$v_j^{\max}$时，说明该参与者数据不符合感知需求，存在错误，其错误值为常数$\rho=1+\varrho$，ϱ为正无穷小数；当属

性值 $g_{i,j}$ 介于 $v_j^{\min}$ 和 $v_j^{\max}$ 之间时，说明感知数据存在误差，且误差值随着 $g_{i,j}$ 值的减小而变大.

由式(4)可得出各数据质量测评参数下的每个 u_i 的误差序列值，记为 $\{l_{i,1},l_{i,2},l_{i,3}\}$.

接着计算参与者数据 u_i 的最大误差值 $l_i^{\max}$，并判断合理性，$l_i^{\max}$ 的计算如式(5)所示. 当 $l_i^{\max}=\rho$ 时，则该参与者的感知数据不满足需求，应剔除此错误数据.

$$l_i^{\max}=\max\{l_{i,j}\} \tag{5}$$

剔除了低质量和异常数据后，先对数据质量测评参数用正规化方法进行处理，如式(6)所示，然后对符合任务需求的感知数据进行排序.

$$\begin{cases} V_j^{\max}=\dfrac{v_j^{\max}}{v_j^{\max}}=1 \\ V_j^{\min}=\dfrac{v_j^{\min}}{v_j^{\max}} \end{cases},\ j=1,2,3 \tag{6}$$

基于式(6)，对数据质量测评参数极限损失值 E_j^*($E_j^*\in[0,1]$)进行计算，如式(7)所示.

$$E_j^*=\frac{(V_j^{\max}-V_j^{\min})^2}{\sum_{k=1}^{3}(V_k^{\max}-V_k^{\min})^2} \tag{7}$$

为实现感知数据的排序，还需计算误差损失序列 s_{u_i}. 由于各群智感知任务对实时性、区域性及数据准确性等属性关注度有所不同，因此，引入权值因子 ω_j，以满足不同情形下的数据质量需求，如式(8)所示.

$$\begin{cases} s_{u_1}=\{\omega_1 l_{1,1}E_1^*,\omega_2 l_{1,2}E_2^*,\omega_3 l_{1,3}E_3^*\} \\ s_{u_2}=\{\omega_1 l_{2,1}E_1^*,\omega_2 l_{2,2}E_2^*,\omega_3 l_{2,3}E_3^*\} \\ \cdots \\ s_{u_i}=\{\omega_1 l_{i,1}E_1^*,\omega_2 l_{i,2}E_2^*,\omega_3 l_{i,3}E_3^*\} \end{cases} \tag{8}$$

其中，$\sum_{j=1}^{3}\omega_j=1$，$i=1,2,\cdots,m$.

通过式(8)的计算得到各参与者数据的误差损失序列 s_{u_i}，将其转换成三维空间上的坐标点，进一步计算得到数据效用值 DE_i，如式(9)所示. 坐标点越接近原点 O，则参与者的数据效用值越大，说明数据质量越高，反之，数据质量越低.

$$DE_i=\frac{1}{\sqrt{\sum_{j=1}^{3}(\omega_j l_{i,j}E_j^*)^2}},\ i=1,2,\cdots,m \tag{9}$$

最终 AS 依据各参与者的数据效用值 DE_i，可完成对感知数据质量的排序.

考虑到参与者在执行感知任务时花费了很大代价，他们希望能够获得满意的回报. 因此，IMPP 除了关注数据质量外，还支持参与者提出期望回报 b_i，参与者要依据自身的数据效用值 DE_i 并结合感知任务预算 $Budget$ 提出合理的回报. 记 $Bid_i=(b_i,DE_i)$ 为效用回报，参与者将封装了 Bid_i 和 $\overline{BidderID_i}$ 的加密数据包 $E_w(Sig_{w,k^{-1}}(Bid_i\|\overline{BidderID_i}))$ 发送于 AS 参与竞价，$\overline{BidderID_i}$ 能够证明参与者已经是注册用户.

4.2.3 参与者最优选择策略

AS 收到各参与者的效用回报 Bid_i 后，可依据其中的数据效用值 DE_i 对感知数据质量加以优劣区分. 为了控制任务预算，IMPP 设计了参与者最优选择策略(规则 1)，以支持 AS 挑选出“性价比”最高的胜出者.

规则 1(参与者最优选择). 假设有两个参与者 u_i 和 u_j，其效用回报如果满足条件

$$\begin{cases} DE_i>DE_j \\ b_i<b_j<Budget \end{cases} \text{且}$$

则 u_i 为“性价比”更高的参与者，会被选择为胜出者.

AS 依据规则 1 对参与者进行“性价比”排序，结合任务需求选择前 n 个“性价比”最高的参与者成为胜出者，并将各胜出者的 $\overline{BidderID_i}$ 公布于“消息窗口”，至此完成胜出者选择环节. 特此说明，对于胜出者的具体奖励机制不列入本文研究范围，将作为本文的后续研究.

总之，IMPP 是基于数据效用值来测评参与者的感知数据质量，依据参与者最优选择策略控制有限任务预算下的资源合理分配. 因此，IMPP 与前人的研究[29-31]不同，是一种更科学合理的激励机制.

4.3 感知数据聚合

胜出者需要将感知数据提交给 AS 后才能获得相应回报. 为了保护其感知数据的隐私性，IMPP 利用分布式压缩感知方法，并结合安全哈希函数技术对感知数据加以隐匿. 首先，对原始感知数据通过设计的观测矩阵进行稀疏表示和压缩观测，以剔除冗余数据，得到观测值；然后，将观测值作为 $Hash()$ 的输入所得的哈希值(即噪扰数据)和随机函数生成的随机噪扰数据联合添加到原观测值中加以混淆，达到隐匿真实感知数据的目的；接着，胜出者把混淆的观测值及 $\overline{BidderID_i}$ 封装加密形成隐私数据包发送给 AS，AS 对收到的隐私数据包进行聚合并校验数据的完整性；最后，借助最优范数重构通过完整性

校验的感知数据.

4.3.1　分布式压缩感知理论及方法

近年来新兴的一种感知数据采集、聚合解决方案为分布式压缩感知 DCS(Distributed Compressed Sensing),它是对传统奈奎斯特采样定理的突破创新[28].其利用新颖的采样编码技术,分析不同样本数据间的时间、空间关联性,将数据通过非自适应线性投影到观测矩阵,从而获得一组远小于原始数据长度的观测值,并通过高效的最优范数算法从观测值中重构原始数据.DCS 理论主要涉及数据的稀疏表示、观测矩阵构建和重构 3 个基本核心问题.研究已证明,DCS 压缩技术能够实现源端的压缩编码加密传输后,到达接收端经解码仍然能够获取相同的数据量[33].采用 DCS 方法对感知数据压缩处理,需要建立能够描述、压缩感知数据的稀疏模型.DCS 基本模型如式(10)所示.

$$\min\|\boldsymbol{\vartheta}\|_{l_1}\ \text{s.t.}\ \boldsymbol{y}=\boldsymbol{\Phi x}=\boldsymbol{\Phi\Psi\vartheta}=\boldsymbol{A}^{\text{DCS}}\boldsymbol{\vartheta} \tag{10}$$

其中,$\boldsymbol{x}$ 为原始数据,$\boldsymbol{x}\in R^{N\times 1}$;$\boldsymbol{\Psi}=[\psi_1,\psi_2,\cdots,\psi_N]$ 为 $N\times N$ 的稀疏变换矩阵(稀疏变换基);$\boldsymbol{\vartheta}$ 为 $\boldsymbol{x}$ 在 $\boldsymbol{\Psi}$ 域的变换向量,若向量 $\boldsymbol{\vartheta}$ 中非零元素的数目 $\zeta\ll N$(远小于原始数据的长度 N),则 $\boldsymbol{x}$ 在 $\boldsymbol{\Psi}$ 域是 ζ-稀疏的;$\boldsymbol{y}$ 为压缩观测值,$\boldsymbol{y}\in R^{M\times 1}$;$\boldsymbol{\Phi}$ 为 $M\times N$ 的观测矩阵;$\boldsymbol{A}^{\text{DCS}}=\boldsymbol{\Phi\Psi}$ 为 $M\times N$ 的感知矩阵.

为了确保精确重构原始数据,要求观测矩阵 $\boldsymbol{\Phi}$ 满足式(11)所示的限制等距性质(Restricted Isometry Property,RIP).

$$(1-\varpi_\zeta)\|\boldsymbol{x}\|_2^2\le\|\boldsymbol{\Phi x}\|_2^2\le(1+\varpi_\zeta)\|\boldsymbol{x}\|_2^2 \tag{11}$$

其中,$0<\varpi_\zeta<1$.然而对于构造的观测矩阵要证明其满足 RIP 性质是很困难的,因此,学术界通常依据等价于 RIP 性质的非相关特性理论来构造与优化观测矩阵.

根据稀疏分解理论,N 维原始数据 $\boldsymbol{x}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)$可以表示为一组标准正交变换基的线性组合:$\boldsymbol{x}=\boldsymbol{\Psi\vartheta}$.目前流行的稀疏变换基有:离散余弦变换基(DCT)、小波变换基(DWT)和傅里叶变换基(DFT)等[34].对原始数据的重构可由式(10)反推 $\boldsymbol{x}$,并通过求解 l_1 最优范数问题实现.

本文采用 DCT 对原始数据进行稀疏表示,用 ζ 代表稀疏度.接下来重点对观测矩阵进行研究设计.观测矩阵对压缩感知起着至关重要的作用,其性能好坏直接影响原始数据的压缩与重构效果.目前观测矩阵一般分为随机矩阵和确定性矩阵[34-35].研究表明,确定性观测矩阵比随机观测矩阵在实际应用中更易于硬件实现.另外,对于观测矩阵的优化,学者们常常从非相关性出发,寻求如何增大观测矩阵和稀疏变换基二者之间非相关性的策略.因此,本文对确定性观测矩阵进行研究,以正交对称托普利兹矩阵(OST)[35]为基础,从非相关性考虑,对其改进优化,构造优化型 OST 观测矩阵(Optimized OST,OP-OST).

OST 观测矩阵的原始模型[35]如下:

$$\boldsymbol{\Phi}=\begin{bmatrix}
a & b & c & \cdots & g & h & g & \cdots & c & b\\
b & a & b & c & \cdots & g & h & g & \cdots & c\\
c & b & a & b & c & \cdots & g & h & g & \cdots\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
g & f & e & d & c & b & a & \cdots & g & h\\
h & g & f & e & d & c & b & a & \cdots & g\\
g & h & g & f & e & d & c & b & a & \cdots\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
b & c & d & e & f & g & h & \cdots & b & a
\end{bmatrix}.$$

该观测矩阵是通过特定的随机序列循环移位,并对任意的 M 行施加标准化运算得到.

首先,OST 矩阵的第一列元素由 $N\times N$ 循环矩阵的第一列元素$(\delta_1,\delta_2,\cdots,\delta_N)^{\text{T}}$通过式(12)所示的逆傅里叶变换得到.

$$N\begin{bmatrix}\chi_1\\ \chi_2\\ \vdots\\ \chi_N\end{bmatrix}=\boldsymbol{F}_N^*\begin{bmatrix}\delta_1\\ \delta_2\\ \vdots\\ \delta_N\end{bmatrix} \tag{12}$$

其中,$\boldsymbol{F}_N^*$ 为 $N\times N$ 的逆傅里叶变换矩阵;$(\delta_1,\delta_2,\cdots,\delta_N)=(\gamma,\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_{N/2-1},\beta,\varepsilon_{N/2-1},\cdots,\varepsilon_1)$,且$(\gamma,\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_{N/2-1},\beta)\in\{-1,1\}^{N/2+1}$.

然后,将第一列元素通过逐列循环移位得到其余列,并定义一个对角矩阵 $\mathcal{M}=\text{diag}(\delta_1,\delta_2,\cdots,\delta_N)$,计算 $\ddot{\boldsymbol{\phi}}_N=\frac{1}{N}\boldsymbol{F}_N^*\mathcal{M}\boldsymbol{F}_N$,从而得到正交对称形式的托普利兹方阵 $\ddot{\boldsymbol{\phi}}_N$.

最后,任意选取其中的M行,并乘以系数$(N/M)^{1/2}$加以标准化,最终得到 $M\times N$ 的 OST 观测矩阵 $\boldsymbol{\Phi}$,即 $\boldsymbol{\Phi}=(N/M)^{1/2}\ddot{\boldsymbol{\phi}}_N$.

本文基于 OST 观测矩阵 $\boldsymbol{\Phi}$ 的原型,引入阈值迭代收缩算法[36]对 OST 进行改进.首先,计算稀疏变换基 $\boldsymbol{\Psi}$ 和观测矩阵 $\boldsymbol{\Phi}$ 两者列向量的数量积,将数量积中最大值定义为相关性系数 η.非相关特性与 η 值有关,并随着 η 值的减小其性能越好.

然后,将感知矩阵 $\boldsymbol{A}^{\mathrm{DCS}}=\boldsymbol{\Phi\Psi}$ 作初等行和列变换,得到单位矩阵$\widetilde{\boldsymbol{A}^{\mathrm{DCS}}}$,则定义 $\boldsymbol{A}^{\mathrm{DCS}}$的列向量相关性系数 $\eta(\boldsymbol{A}^{\mathrm{DCS}})$为式(13).

$$\eta(\boldsymbol{A}^{\mathrm{DCS}})=\max_{i\neq j}\left\{\frac{|\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{A}_j|}{\|\boldsymbol{A}_i\|\|\boldsymbol{A}_j\|}\right\}=\max_{i\neq j}\{|\widetilde{\boldsymbol{A}}_i^{\mathrm{T}}\widetilde{\boldsymbol{A}}_j|\} \tag{13}$$

其中,$\boldsymbol{A}_i$和$\widetilde{\boldsymbol{A}}_i$分别是$\boldsymbol{A}^{\mathrm{DCS}}$和$\widetilde{\boldsymbol{A}^{\mathrm{DCS}}}$的列向量.

接着,生成格拉姆矩阵 $\boldsymbol{GR}=\widetilde{\boldsymbol{A}^{\mathrm{DCS}}}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{DCS}}$,其相关性系数可定义为:$\eta(\boldsymbol{GR})=\max_{i\neq j}|cu_{ij}|=\max_{i\neq j}|\widetilde{\boldsymbol{A}}_i^{\mathrm{T}}\widetilde{\boldsymbol{A}}_j|$,$cu_{ij}$是$\boldsymbol{GR}$的因子,为$\boldsymbol{\Phi}$和$\boldsymbol{\Psi}$不同列的数量积.$\eta(\boldsymbol{GR})$和$\eta(\boldsymbol{A}^{\mathrm{DCS}})$等价,但不能刻画全局相关性.为进一步降低$\boldsymbol{\Phi}$和$\boldsymbol{\Psi}$的平均相关性,采用基于阈值$\xi$的平均相关系数$\eta_\xi$,对$\boldsymbol{GR}$中非对角线上所有大于或等于阈值$\xi$的元素取模并求取平均程度,如式(14)所示.

$$\eta_\xi=\frac{\sum_{1\leq i,j\leq n,i\neq j}(|cu_{ij}|\geq\xi)\cdot|cu_{ij}|}{\sum_{1\leq i,j\leq n,i\neq j}(|cu_{ij}|\geq\xi)} \tag{14}$$

总之,上述观测矩阵设计过程中,先由感知矩阵$\boldsymbol{A}^{\mathrm{DCS}}$生成格拉姆矩阵,然后对其进行阈值迭代,最后再反推出感知矩阵,进而得到具有较好重构性能的观测矩阵$\boldsymbol{\Phi}'$.

4.3.2 感知数据的隐私性保护

为便于阐述,把每个胜出者当作一个节点i.节点i的原始感知数据经过 4.3.1 节中所改进的观测矩阵压缩观测后得到观测值$\boldsymbol{q}_i$,再对$\boldsymbol{q}_i$利用$Hash()$求取哈希值λ_i,如式(15)所示.

$$\lambda_i=Hash(\boldsymbol{q}_i) \tag{15}$$

将λ_i当作噪扰数据,加入到$\boldsymbol{q}_i$中以混淆真实的$\boldsymbol{q}_i$,把λ_i在$\boldsymbol{q}_i$中的位置标记为$ZVPS$.为进一步增强隐匿强度,根据实际需求,可由随机函数生成适量任意值作为随机噪扰数据,一并填充到$\boldsymbol{q}_i$中形成联合混淆.同时,把随机噪扰数据在$\boldsymbol{q}_i$中的位置标记为$\overline{ZVPS}$.

记两种噪扰数据的填充位置集合为$GVPS=ZVPS+\overline{ZVPS}$,原观测值集合为$OVPS$.胜出者节点$i$需要把填充了噪扰数据的观测值经过签名并加密后上传给AS,同时在加密数据包里附加$\overline{BidderID_i}$,以证明该隐私数据包是来源于经 AS 授权的胜出者i.隐私数据包PD_i如式(16)所示.

$$PD_i=E_w(Sig_{w,k^{-1}}(\lambda_i),(GVPS\|OVPS\|\overline{BidderID_i})) \tag{16}$$

4.3.3 感知数据的完整性校验

AS 收到节点i上传的隐私数据包,需要对其进行完整性校验,仍然采用单向安全哈希函数来完成.校验过程如图 4 所示.

(1)节点i对感知数据压缩观测得到观测值,利用$Hash()$对观测值求取哈希值λ_i.

(2)AS 收到节点i的PD_i后,经过验证数字签名可判断PD_i的信息内容是否被伪造,再根据其中的噪扰数据填充信息剔除噪扰数据,获取原观测值信息,对观测值再次求取哈希值$\overline{\lambda}_i$.

(3)AS 比较两个哈希值.如果$\lambda_i=\overline{\lambda}_i$,说明观测值在传输过程中未被篡改,可通过数据完整性校验,观测值能够被重构;否则,AS 丢弃PD_i.

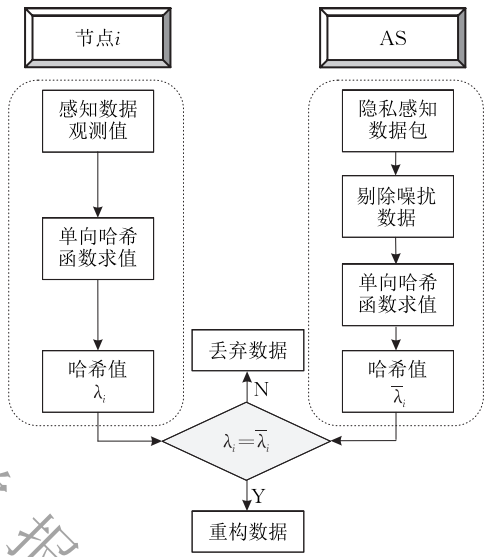

图 4 感知数据完整性校验

4.3.4 感知数据的重构

AS 对通过完整性校验的观测值进行重构,恢复出原始感知数据.分布式压缩感知重构算法一般包括凸优化类和迭代贪婪类两类算法[37].由于迭代贪婪重构算法具有计算复杂度低且重构精确度高的优势,被广泛应用于压缩感知领域.因此,本文采用经典的迭代贪婪类算法——正交匹配跟踪算法(Orthogonal MP,OMP)实现原始数据的重构.OMP 算法能够对所选原子做正交化处理,有效克服匹配跟踪算法 MP 的缺陷,在相同精度要求下,较 MP 算法具有更快的收敛速度.

依据 DCS 理论,如果$\boldsymbol{A}^{\mathrm{DCS}}$满足观测矩阵$\boldsymbol{\Phi}$与稀疏变换基$\boldsymbol{\Psi}$非相关,那么节点$i$的稀疏变换向量$\boldsymbol{\vartheta}$可通过求解最优$l_1$范数问题精确重构.重构计算如式(17)所示.

$$\begin{aligned}&\hat{\boldsymbol{\vartheta}}_i=\arg\min\|\boldsymbol{\vartheta}_i\|_{l_1}\\&\text{s.t. }\bar{\boldsymbol{q}}_i=\boldsymbol{A}^{\mathrm{DCS}}\boldsymbol{\vartheta}_i\end{aligned} \tag{17}$$

其中，$\bar{\boldsymbol{q}}_i$ 为经 AS 完整性校验合格的观测值；$\hat{\boldsymbol{\vartheta}}_i$ 为通过最优范数算法求解得到的估计值.

最后，节点 i 的原始感知数据 $\hat{\boldsymbol{p}}_i$ 可由式(18)重构.

$$\hat{\boldsymbol{p}}_i=\boldsymbol{\psi}\hat{\boldsymbol{\vartheta}}_i \tag{18}$$

4.3.5　基于分布式压缩感知的隐私保护算法

在对 4.3.1～4.3.4 节 DCS 理论及方法研究的基础上，本文提出感知数据隐私保护算法如算法 2 所示.

算法 2.　Privacy-preserving of Sensory Data.

输入：原始感知数据 $\boldsymbol{p}_i$

输出：重构的感知数据 $\hat{\boldsymbol{p}}_i$

1. FOR each $i, i=1,\cdots,n$ DO
2. $\quad \boldsymbol{q}_i=\boldsymbol{\Phi}'\boldsymbol{p}_i=\boldsymbol{\Phi}'\boldsymbol{\Psi}\boldsymbol{\vartheta}_i=\boldsymbol{A}^{\mathrm{DCS}'}\boldsymbol{\vartheta}_i$;
3. $\quad \lambda_i=Hash(\boldsymbol{q}_i)$;
4. $\quad PD_i=E_w(Sig_{w,k^{-1}}(\lambda_i),$
$\qquad (GVPS\|OVPS\|\overline{BidderID}_i))$;
5. $\quad i\rightarrow \mathrm{AS}: PD_i$;
6. $\quad \mathrm{AS}: Sig.verify_{\mathrm{AS},k}(D_{\mathrm{AS}}(PD_i))$;
7. $\quad \bar{\lambda}_i=Hash(OVPS)$;
8. $\quad$ IF $\lambda_i=\bar{\lambda}_i$ THEN
9. $\qquad \mathrm{AS}: \hat{\boldsymbol{p}}_i=\boldsymbol{\psi}\hat{\boldsymbol{\vartheta}}_i$;
10. $\quad$ ELSE
11. $\qquad \mathrm{AS}:$ Discard PD_i;
12. $\quad$ ENDIF
13. ENDFOR
14. AS: return $(\hat{\boldsymbol{p}}_1\cup\hat{\boldsymbol{p}}_2\cup\cdots\cup\hat{\boldsymbol{p}}_n)$.

IMPP 采用算法 2 实现了剔除冗余感知数据，通过填充噪扰数据增强了数据的隐私性保护，借助单向安全加密哈希函数完成了数据完整性校验，为减小网络传输通信开销以及提高数据聚合效率提供了解决方案.

5　性能分析及实验

为评估 IMPP 的有效性，本节先对 IMPP 的安全性及性能进行分析，然后再从隐私保护水平、数据完整性、数据精确性、时间效率、评估准确率、重构匹配度及激励效果等方面进行实验验证.

5.1　安全及性能分析

(1) 隐私安全性分析

对于参与者的身份隐私保护，如果安全加密哈希函数 $Hash()$ 是安全的，则参与者身份匿名保护算法(算法 1)能够保护参与者的身份不泄露. 因为参与者每次参与任务时选择一个随机数作为临时公钥 k，且 k 与参与者身份信息没有任何联系，所以从公钥层面就隐藏了参与者的真实身份. 再者，k 通过 TLS 信道传输给 AS，确保诚实参与者的公钥不被攻击者获取，以免对前者造成隐私危害. 此外，$Hash(k)$ 的值为 256 位，计算不可逆性和强抗碰撞性可以确保参与者的虚拟身份标识 $BidderID$ 唯一，且攻击者不能由 $BidderID$ 反向推出 k，所以，参与者的身份隐私保护是安全的.

对于参与者的感知数据隐私保护，如果压缩感知方法安全及噪扰数据填充有效，则感知数据隐私保护算法(算法 2)能够保证感知数据的安全性. 因为压缩感知方法能够剔除冗余数据，保留最少的特征观测值 $\boldsymbol{q}_i=\boldsymbol{\Phi}'\boldsymbol{p}_i$，减少了数据中敏感信息的暴露机会，同时又结合哈希函数 $Hash(\boldsymbol{q}_i)$ 并将函数值 λ_i 填充到观测值 $\boldsymbol{q}_i$ 中，以混淆原观测值 $\boldsymbol{q}_i$，进一步增强了感知数据隐私性保护. 压缩感知方法实际是求解欠定线性方程 $\boldsymbol{y}=\boldsymbol{\Phi x}=\boldsymbol{\Phi\Psi\vartheta}=\boldsymbol{\Theta\vartheta}$ 问题，并用 $\min\|\boldsymbol{\vartheta}\|_{l_1}$ s.t. $\boldsymbol{y}=\boldsymbol{\Phi\Psi\vartheta}$ 最优化范数来恢复原始数据. 对感知数据的压缩和观测操作可以视为等效的加密过程. 所以，基于此方法的数据隐私保护具有安全性.

(2) 通信性能分析

为便于阐述，把参与者称作感知节点. 本文衡量通信开销的标准为感知节点所发出的数据包量. 假设一个数据包的大小为 μ bit. 将 IMPP 与 ICKPDA[27]、DCSPDA[28] 进行对比，分析三种方法的通信开销.

ICKPDA 方法中，由叶子节点对传输数据进行随机分片，自身存储其中一个分片后，将其余分片在规定的时延内分发给任意选择的邻居节点. 邻居节点对接收的分片先解密，之后再与其存储的数据分片融合. 各个节点都将各自的数据分片与从邻居节点接收到的数据分片循环融合后，再依据数据融合树逐层上传，最终到达树根得到汇聚结果. 由于每个节点存储的分片数 $m>2$，所以，ICKPDA 的通信开销不小于 $O(4\mu)$，且通信开销随着 m 的增大急剧增长.

DCSPDA 方法中，每个簇节点需向接收器发送一个原始数据的 DCS 观测值数据包，通过簇头从接收器接收一个稀疏系数配置包，最后向簇头发送一个隐私保护数据包. 由此可知，DCSPDA 的通信开销为 $O(3\mu)$.

IMPP 方法中，通信开销主要产生在感知节点端. 感知节点首先发送需 AS 签名的虚拟身份标识，

在完成任务后向 AS 发送一个效用回报加密数据包,待竞价成功后再向 AS 上传压缩隐私保护的感知数据包.因此,IMPP 的通信开销为 $O(3\mu)$.

综上所述,IMPP 的通信开销比 ICKPDA 降低了 25%,和 DCSPDA 相等.IMPP 实现了数据完整性保护,但通信开销并没有比 DCSPDA 增加,可见,IMPP 的通信性能良好.

5.2 实验验证

5.2.1 实验数据及环境

实验的数据集来自实验室开发的校园卫生环境监测系统,系统在 Intel Core i5-7200U 2.5 GHz CPU、8GB RAM 的 Windows 7 平台上运行.系统于 6 个月内累计发布了 100 个不同的感知任务,任务类型属于单任务,每个任务要求监测一个校园某区域(如教学区、宿舍区、活动区等)的卫生状况,系统共积累了约 21 200 条数据记录.任务属性包括感知时间、感知区域、感知数据的数量、规格等.每一个感知任务要求参与者必须在感知时间周期内完成给定感知区域的数据采集,并且在感知区域内随机选择至少 10 个采集点,采集点的间距应保持在 50m~100 m,每个采集点的数据为 5 个.则参与者的感知数据集是由多条数值型数据记录组成的线性结构,数据记录又是由三元组组成.将参与者 i 的感知数据集形式化表示为 $Data_i=\{\langle t_1,p_1,(x_{1,1},x_{1,2},\cdots,x_{1,5})\rangle,\langle t_2,p_2,(x_{2,1},x_{2,2},\cdots,x_{2,5})\rangle,\cdots,\langle t_j,p_j,(x_{j,1},x_{j,2},\cdots,x_{j,5})\rangle\}$, $j\geq 10$.每个任务招募的参与者以 50 人为一个单位,根据任务不同,招募人数范围为 100~500 人.

为了测试 IMPP 在隐私保护水平、数据完整性和数据精确性方面的性能,且在相同实验环境下与 ICKPDA、DCSPDA 进行对比,引入隐私阈值.隐私阈值代表了参与者对个人隐私泄露的容忍度,可通过隐私保护策略设置其值为[0,1].通常,阈值越低,参与者容忍度越高;相反,阈值越高,参与者容忍度越低.

为了测试 IMPP 的时间效率,将其与 ICKPDA、DCSPDA 在参与者人数范围为 100~500 人的实验环境下进行对比分析.

为了测试 IMPP 的数据效用值评估数据质量的准确率,将其与传统异常数据检测方法——拉依达准则(Pauta 准则)[32]在相同的异常数据比例情况下进行对比分析.拉依达准则是对预先假设只含有随机误差的一组检测数据进行计算处理,得到标准偏差,并确定满足一定概率的区间,凡是超过该区间的误差均属于粗大误差,而包含粗大误差的数据应被剔除[32].它通常适用于测量次数较大的情况.

为了测试 IMPP 对原始数据的压缩重构性能,将本文优化型观测矩阵 OP-OST 与 OST 矩阵[35]、传统的 Toeplitz 矩阵在相同的压缩率下进行对比分析.

为了测试 IMPP 的激励效果,将其与 RADP[29]、RSFP[29]在参与者人数范围为 100~500 人的实验环境下进行对比分析.

5.2.2 度量指标

(1)隐私保护水平

攻击者成功获取到超出参与者容忍程度的身份信息或与身份相关的敏感信息,则造成隐私泄露.隐私泄露概率是指攻击者在对某个参与者的某次攻击中获得的综合隐私信息超出参与者对于隐私泄露容忍程度的概率.通过式(19)来计算隐私保护水平.其中,p 为隐私泄露概率,$Privacy\_Level$ 为隐私保护水平.隐私泄露概率越小,隐私保护水平越高.

$$Privacy\_Level=\left(\frac{1}{2}\right)^{p-1}-1 \tag{19}$$

(2)数据完整性

通过分析方法能否抵御重放、伪造、篡改攻击来衡量方法的数据完整性保护性能.

重放攻击就是把窃听到的原始数据再原样重新发送给接收方.伪造攻击是假冒发送方向接收方发送虚假数据.攻击是用作伪的手段对网络传输的数据进行改动、增加或删除,造成数据损坏.

(3)数据精确性

由于对数据施加了隐私保护,所以感知数据的精确性会受到一定影响,从而影响数据的可用性.数据精确性通常用误差平方和(Sum of Squared Errors,SSE)进行度量,如式(20)所示.SSE 的值越小,数据精确性越好,数据的可用性也越好.

$$SSE=\sum_{p_i\in N}\sum_{a_j^k\in p_i}(dist(a_j^k,(a_j^k)'))^2 \tag{20}$$

其中,a_j^k 和 $(a_j^k)'$ 分别是原始数据集和隐私保护数据集中的第 j 个记录的第 k 个属性.$dist()$ 是距离函数.p_i 是节点 i 的原始数据,N 是所有节点原始数据的集合.

(4)时间效率

用 t_i 表示隐私保护方法的一次运行时间,$T(n)$ 表示方法执行 n 次的平均运行时间,如式(21)所示.用 $T(n)$ 来衡量方法的时间效率,$T(n)$ 值越小,方法的效率越高.

$$T(n)=\frac{1}{n}\sum_{i=1}^{n}(t_i) \tag{21}$$

(5)评估准确率

数据质量评估准确率是衡量数据效用值有效性的重要因素之一.评估准确率越高,越能客观评判参与者感知数据的质量,平台获得的总体数据质量会越好,参与者可能获得的收益也越大.消错决策理论从规避错误角度出发,通过降低错误损失对满足需求的数据进行识别.因此,本文用异常数据识别准确率来反映评估准确率.异常数据识别准确率 YC 定义为式(22).其中,n_{sum} 表示异常数据中的真实异常数据个数,N_{sum} 表示异常数据总数.$YC\in[0,1]$,其值越趋近于1,说明数据质量评估准确率越高.

$$YC=\frac{n_{sum}}{N_{sum}} \tag{22}$$

(6)压缩率和匹配度

用压缩率 CR 衡量原始数据的压缩度[36].CR 值和压缩度呈正相关,与观测数呈负相关.将 CR 定义为式(23).其中,ℓ_n 是原始数据长度,ℓ_m 是压缩观测数据长度.

$$CR=\frac{\ell_n-\ell_m}{\ell_n}\times 100\% \tag{23}$$

经验表明,压缩感知时压缩率应合理取值,若太大则会影响重构原始数据的精确效果;若太小则会导致达不到压缩效果.因此,压缩率取值为 $60\%\leq CR\leq 85\%$ 最理想.

用匹配度 MD 衡量原始数据的重构性能[36],MD 是指重构数据与原始数据的相似程度,其计算如式(24).

$$MD=1-\frac{|\|\hat{\boldsymbol{p}}_i\|_2-\|\boldsymbol{p}_i\|_2|}{|\|\hat{\boldsymbol{p}}_i\|_2+\|\boldsymbol{p}_i\|_2|} \tag{24}$$

其中,$MD\in[0,1]$,其值越大,重构匹配度越高,表明压缩重构性能越好.

(7)激励效果

参与率和满意度能够反映激励机制的效果.定义参与率 P_rate 为竞价人数与注册人数(参与者人数)的比值,如式(25)所示.其中,n_bid 表示竞价人数,N_reg 表示注册人数.

$$P\_rate=\frac{n\_bid}{N\_reg} \tag{25}$$

参与者在完成竞价后对感知平台做出评价,评价形式为打分制,参与者所打分数记为 S.平台为参与者提供5个分数等级,1:$90\leq S\leq 100$(非常满意);2:$80\leq S<90$(较满意);3:$70\leq S<80$(满意);4:$60\leq S<70$(一般);5:$S<60$(不满意).

定义满意度 S_degree 为"非常满意"、"较满意"、"满意"人数之和与评价总人数的比值,如式(26)所示.其中,ν_1、ν_2、ν_3 分别表示"非常满意"、"较满意"、"满意"人数,S_ν 表示评价总人数.

$$S\_degree=\frac{\nu_1+\nu_2+\nu_3}{S_\nu} \tag{26}$$

5.2.3　实验结果

三种方法的隐私保护水平对比如图5所示.隐私阈值设置为[0.1,1.0].随着隐私阈值的逐渐增大,三种方法的隐私保护水平均逐渐上升.因为隐私阈值设置得越高,隐私泄露概率就越小,相应数据的隐私保护强度也越大.在相同隐私阈值下,IMPP 和 DCSPDA 的隐私保护水平明显优于 ICKPDA.因为 IMPP 和 DCSPDA 都是利用 DCS 观测矩阵对原始感知数据压缩处理,对数据的压缩和观测操作等效于加密保护,且两种方法还在原观测值中加入噪扰数据以形成混淆数据集,进一步提升了隐私保护强度.而 ICKPDA 是通过对感知数据随机分片,再由数据融合树逐层聚合的方式实现数据隐私保护,分片数直接影响着隐私保护效果.要提高隐私保护水平,需要增加分片数量,则通信能耗也会随之增加.IMPP 的隐私保护性能较 DCSPDA 略有提升,因为 IMPP 中噪扰数据为256位哈希函数值,哈希函数的单向性决定了哈希值取值范围不受观测值的影响,而 DCSPDA 中伪装数据的取值范围有限,影响了对真实数据隐藏的效果.

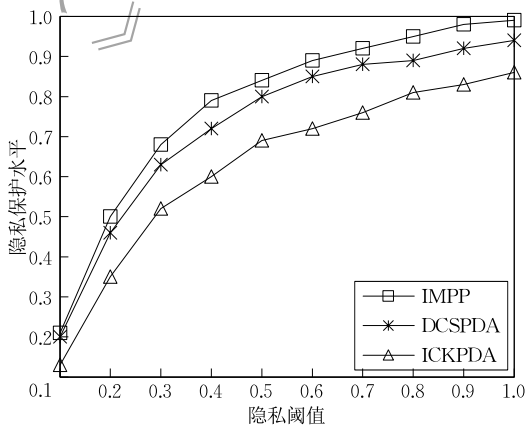

图5　三种方法隐私保护水平对比

三种方法的数据完整性对比如表1所示.对于重放数据攻击,三种方法都不能抵御.IMPP 和 DCSPDA 中,由于感知数据的压缩观测数据在重放攻击前后没有发生改变,所以无法判断是否为重放数据.ICKPDA 中节点通过逐跳传输二元数据到融合节点、汇聚节点,汇聚节点对二元数据解密恢复原始数据,恢复结果没有异常.所以,ICKPDA 也不能

识别重放攻击.

对于伪造数据攻击,只有 IMPP 能有效抵御.因为胜出者向 AS 发送的压缩感知观测值是经过胜出者数字签名的隐私数据包,并加密传输给 AS,到达 AS 端后 AS 能够用胜出者的公钥进行验证,数字签名技术的作用之一就是检测伪造攻击,所以 AS 能够检测隐私数据包是否被伪造.而 DCSPDA 设计时没有考虑数据完整性校验,所以不能抵御伪造攻击.ICKPDA 中,相邻节点之间利用共享密钥建立安全链接进行数据分片传递,但是共享密钥并不能对节点存储的数据分片验证其来源的真伪性.所以,ICKPDA 不能检测伪造数据.

对于篡改数据攻击,IMPP 和 ICKPDA 都具有较强的抵御性能.ICKPDA 中,如果发生了数据片篡改,则汇聚节点能够检测出关联的密钥发生错误,因此,篡改攻击能被发现.IMPP 利用单向安全哈希函数来验证 AS 收到的观测值是否和原观测值一致,如果观测值被篡改,会导致 $\lambda_i \neq \lambda_i'$,AS 能够检测出来,IMPP 在设计时特别考虑了这一问题,所以,IMPP 具有较强的抵御篡改攻击的性能.

表 1 抵御攻击性能分析

	重放数据攻击	伪造数据攻击	篡改数据攻击
IMPP	×	↑	↑
DCSPDA	×	×	×
ICKPDA	×	×	↑

注:×表示不能抵御攻击,↑表示抵御攻击性能较强.

三种方法的数据精确性对比如图 6 所示.同样设置隐私阈值为[0.1,1.0].随着隐私阈值的增大,ICKPDA 的数据精确性下降幅度较大,而 IMPP 和 DCSPDA 的数据精确性略微降低.因为 ICKPDA 将密钥填充在传输数据中,并将填充密钥后的数据进行分片以增强数据的隐私保护.当隐私阈值增大,隐私保护强度也随之增大,数据分片数目会大大增加,导致大量的分片数据通过密钥之间的关联性在汇聚节点重组,加上受到网络链路等因素影响,势必对数据的精确性产生较大影响.IMPP 相比 DCSPDA 的数据精确性略高一些.因为 IMPP 利用分布式压缩感知得到观测值的过程相当于轻量级数据加密,代替了使用复杂加密技术实现数据的隐私保护,并且 IMPP 不同于 DCSPDA,不需要在簇头节点聚合感知节点的隐私数据,避免了隐私数据在到达汇聚节点之前经历多跳带来的数据丢失,IMPP 是由感知节点直接发送压缩隐私数据包到 AS 端聚合,并且观测值经过最优范数重构能够得到相等的信息量,所以 IMPP 的数据精确性更高.

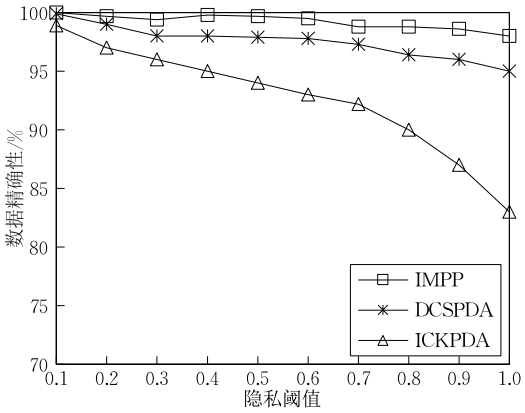

图 6 三种方法数据精确性对比

三种方法的时间效率对比如图 7 所示.三种方法中,ICKPDA 的 $T(n)$ 值受参与者数量变化影响最大,随着参与者数量的增大,$T(n)$ 值上升较快,而方法的时间效率与 $T(n)$ 值呈负相关,所以 ICKPDA 的时间效率较低.因为 ICKPDA 中随着参与者数量的增加,数据分片数量也随之增加,相应地在汇聚节点重组的分片数量也增加,大量的分片重组产生了较长的计算时间.而 IMPP 和 DCSPDA 的 $T(n)$ 值受参与者数量变化影响并不明显,当参与者数量达到 500 时,二者依然保持较高的运行效率,明显优于 ICKPDA.在相同参与者数量下,IMPP 相比 DCSPDA 的 $T(n)$ 值略大,时间效率方面稍劣于 DCSPDA.因为 DCSPDA 只考虑数据的隐私保护,并不涉及数据的完整性校验,而 IMPP 需要检测数据的完整性,但 IMPP 的时间花销并没有比 DCSPDA 多出太多,所以综合考虑,IMPP 在时间效率方面还是较有优势的.

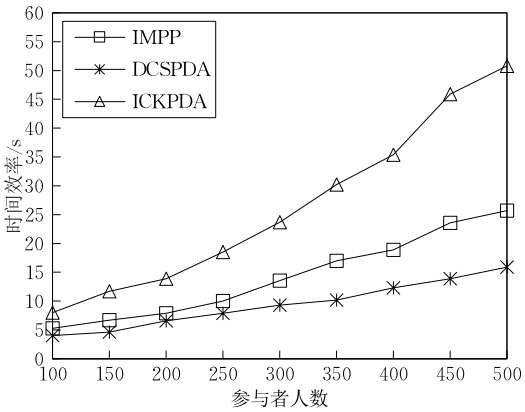

图 7 三种方法时间效率对比

对于数据质量评估准确率的测试,两种方法对比如图 8 所示.从图中看出,Pauta 准则的准确率随着异常数据比例的增加下降较快,而 IMPP 的准确率几乎保持不变.可见,IMPP 用数据效用值评估数

据质量明显优于 Pauta 准则. 因为数据效用值的出发点是对超出合理误差值范围的异常数据进行剔除，只与每一类数据质量测评参数值域有关，与异常数据所占比例无关，因此，评估准确率几乎不受异常数据比例的影响. 而 Pauta 准则是采用方差的方法对异常数据进行判断和处理，异常数据的增加会使得数据的整体方差偏差增大，造成与真实值偏差很小的异常数据被错误地认为是正确数据，从而影响了 Pauta 准则的评估性能.

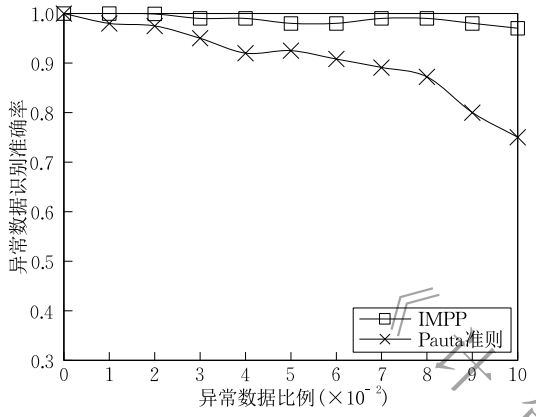

图 8　两种方法评估准确率对比

对于分布式压缩感知方法的压缩重构性能的测试，三种观测矩阵获得的匹配度对比如图 9 所示. 从图中看出，在 $60\% \leq CR \leq 85\%$ 的情况下，随着压缩率的不断增大，各矩阵对原始数据的重构匹配度都不断减小. 因为压缩率越大，需要的观测数目越少，重构的难度就越大，所以，重构数据和原始数据相似程度会有一定的差距. 其中，Toeplitz 矩阵的重构匹配度下降最快，OST 矩阵次之，而优化型 OP-OST 矩阵下降最慢. 当压缩率小于 75% 时，OP-OST 矩阵的匹配度略有减小，而其它两种矩阵匹配度减小的幅度较大. 这是因为 OP-OST 矩阵通过阈值迭代法来增大观测矩阵和稀疏变换基之间的非相关性，使得重构性能有了一定程度的提高. 可见，本文构造的 OP-OST 矩阵在原始数据重构方面的性能良好.

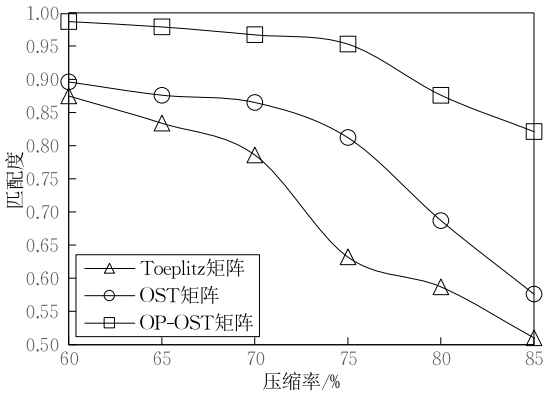

图 9　三种观测矩阵的重构匹配度对比

对于参与率的测试，三种方案对比如图 10 所示. 从图中看出，当注册人数达到 500 时，IMPP 的用户参与率仍大于 95%，而 RADP 和 RSFP 均下降到 90%以下. 因为 RADP 是通过动态价格逆向拍卖机制来保证预算的可控性和参与者数量的稳定性，RSFP 是基于固定价格随机选择胜出者，并且两者仅向胜出者奖励报酬，对竞价失败的参与者不给予补偿，而且都未研究隐私保护问题. 而 IMPP 是兼顾参与者身份和数据隐私保护，并且关注感知数据质量的激励机制，除了对胜出者奖励报酬外，还考虑对失败者的付出给予一定补偿. 因此，IMPP 拥有较高的用户参与率.

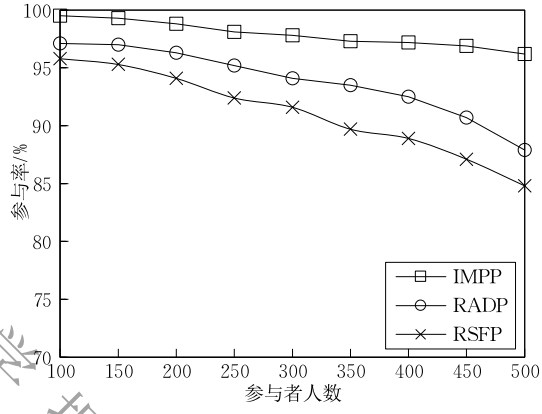

图 10　三种方案用户参与率对比

对于满意度的测试，三种方案对比如图 11 所示. IMPP、RADP 和 RSFP 的平均满意度分别为 95.22%、83.5%、79.08%. 因为 IMPP 是综合考虑隐私保护、数据质量以及失败者补偿等多方面问题的激励机制，对参与者的个人隐私保护、公平竞争以及数量的维持等方面都具有促进作用，所以 IMPP 具有较高的用户满意度.

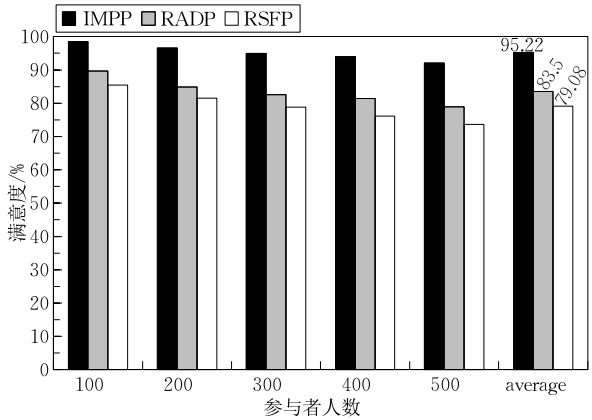

图 11　三种方案用户满意度对比

从图 5～图 11 以及表 1 实验结果分析知，IMPP 在隐私保护水平、数据完整性、数据精确性、时间效率、评估准确率、重构匹配度及激励效果等方面均表现出良好性能，证明了该方案的有效性.

6 结束语

为了同时兼顾参与者的隐私保护和激励问题，本文提出了一种 IMPP 综合机制. 该机制基于轻量级隐私保护思想，采用单向安全哈希函数生成 256 位哈希值作为参与者匿名参与感知任务的身份标识，实现参与者身份隐私保护；依据参与者的数据效用值、期望回报和感知任务预算等多种因素进行参与者最优选择，达到面向数据质量激励的目的；借助分布式压缩感知理论及方法，对胜出者的原始感知数据压缩处理，以剔除冗余数据，之后在观测值中添加哈希函数值等噪扰数据于服务器端聚合，增强感知数据的隐私性保护，且在服务器端对隐私数据集的完整性进行校验，并经过重构恢复出原始感知数据. 通过大量真实数据的实验，与 DCSPDA、ICKPDA、RADP、RSFP 以及 OST 矩阵、Toeplitz 矩阵等进行多方面对比，验证了 IMPP 在隐私保护水平、数据完整性、数据精确性、时间效率、评估准确率、重构匹配度及激励效果等方面的性能. 当然，IMPP 也存在一些不足，不能很好地适应多源异构数据的应用场景. 研究多源异构感知数据的隐私保护和激励是未来的工作方向.

参 考 文 献

[1] Guo B, Chen H, Yu Z, et al. FlierMeet: A mobile crowdsensing system for cross-space public information reposting, tagging, and sharing. IEEE Transactions on Mobile Computing, 2015, 14(10): 2020-2033

[2] Chen Y, Li B, Zhang Q. Incentivizing crowdsourcing systems with network effects//Proceedings of the 35th IEEE International Conference on Computer Communications. San Francisco, USA, 2016: 1-9

[3] Peng Z, Gui X, An J, et al. Multi-task oriented data diffusion and transmission paradigm in crowdsensing based on city public traffic. Computer Networks, 2019, 156(6): 41-51

[4] Ganti R K, Ye F, Lei H. Mobile crowdsensing: Current state and future challenges. IEEE Communications Magazine, 2011, 49(11): 32-39

[5] Li J, Zhu Y, Hua Y, et al. Crowdsourcing sensing to smartphones: A randomized auction approach. IEEE Transactions on Mobile Computing, 2017, 16(10): 2764-2777

[6] Jin H, Su L, Nahrstedt K. Centurion: Incentivizing multi-requester mobile crowd sensing//Proceedings of the 36th IEEE International Conference on Computer Communications. Atlanta, USA, 2017: 1-9

[7] Yang D, Xue G, Fang X, et al. Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones. IEEE/ACM Transactions on Networking, 2016, 24(3): 1732-1744

[8] Ren Xue-Bin, Yang Xin-Yu, Yang Shu-Sen, et al. The privacy-preserving in big data processing and analysis: A literature review. Journal of Northwest University (Natural Science Edition), 2019, 49(1): 1-11(in Chinese)
(任雪斌，杨新宇，杨树森等. 大数据处理和分析中的隐私保护研究综述. 西北大学学报(自然科学版)，2019，49(1)：1-11)

[9] Tian F, Gui X, An J, et al. A DCT-based privacy-preserving approach for efficient data mining. Security and Communication Networks, 2015, 8(18): 3641-3652

[10] Wang Tao-Chun, Liu Ying, Jin Xin, et al. Research on k-anonymity-based privacy protection in crowd sensing. Journal on Communications, 2018, 39(Z1): 170-178 (in Chinese)
(王涛春，刘盈，金鑫等. 群智感知中基于 k-匿名的位置及数据隐私保护方法研究. 通信学报，2018，39(Z1)：170-178)

[11] Wang Y, Xu D, Li F. Providing location-aware location privacy protection for mobile location-based services. Tsinghua Science and Technology, 2016, 21(3): 243-259

[12] Li Jie, Bai Zhi-Hong, Yu Rui-Yun, et al. Mobile location privacy protection algorithm based on PSO optimization. Chinese Journal of Computers, 2018, 41(5): 1037-1051(in Chinese)
(李婕，白志宏，于瑞云等. 基于 PSO 优化的移动位置隐私保护算法. 计算机学报，2018，41(5)：1037-1051)

[13] Alsheikh M A, Jiao Y, Niyato D, et al. The accuracy-privacy trade-off of mobile crowdsensing. IEEE Communications Magazine, 2017, 55(6): 132- 139

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

[15] Shen Li-Yan, Chen Xiao-Jun, Shi Jin-Qiao, et al. Survey on private preserving set intersection technology. Journal of Computer Research and Development, 2017, 54(10): 2153-2169(in Chinese)
(申立艳，陈小军，时金桥等. 隐私保护集合交集计算技术研究综述. 计算机研究与发展，2017，54(10)：2153-2169)

[16] Boubiche S, Boubiche D E, Bilami A, et al. Big data challenges and data aggregation strategies in wireless sensor networks. IEEE Access, 2018, 6: 20558-20571

[17] Xiong Jin-Bo, Ma Rong, Niu Ben, et al. Privacy protection incentive mechanism based on user-union matching in mobile crowdsensing. Journal of Computer Research and Development, 2018, 55(7): 1359-1370(in Chinese)
(熊金波，马蓉，牛犇等. 移动群智感知中基于用户联盟匹配的隐私保护激励机制. 计算机研究与发展，2018，55(7)：1359-1370)

[18] Zhang Yi-Xuan, He Jing-Sha, Zhao Bin, et al. A privacy protection model on game theory. Chinese Journal of Computers, 2016, 39(3): 615-627(in Chinese)
(张伊璇，何泾沙，赵斌等. 一个基于博弈理论的隐私保护模型. 计算机学报，2016，39(3)：615-627)

[19] Sweeney L. k-Anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570

[20] Zhou Zhi-Ping, Li Zhi-Cong. Data anonymous collection protocol without trusted third party. Journal of Electronics & Information Technology, 2019, 41(6): 1442-1449(in Chinese)
(周治平，李智聪. 无可信第三方的数据匿名化收集协议. 电子与信息学报，2019，41(6)：1442-1449)

[21] Zhang L, Xuan J, Si R, et al. An improved algorithm of individuation k-anonymity for multiple sensitive attributes. Wireless Personal Communications, 2017, 95(3): 2003-2020

[22] Ganti R K, Pham N, Tsai Y-E, et al. PoolView: Stream privacy for grassroots participatory sensing//Proceedings of the 6th International Conference on Embedded Networked Sensor Systems. Raleigh, USA, 2008: 281-294

[23] Zhang F, He L, He W, et al. Data perturbation with state-dependent noise for participatory sensing//Proceedings of the 31st IEEE International Conference on Computer Communications. Orlando, USA, 2012: 2246- 2254

[24] Cristofaro E D, Soriente C. Extended capabilities for a privacy-enhanced participatory sensing infrastructure (PEPSI). IEEE Transactions on Information Forensics and Security, 2013, 8(12): 2021-2033

[25] Xiong J, Zhang Y, Li X, et al. RSE-PoW: A role symmetric encryption PoW scheme with authorized deduplication for multimedia data. Mobile Networks and Applications, 2018, 23(3): 650-663

[26] Zhang J, Zhu J, Jia Z, et al. A secret confusion based energy-saving and privacy-preserving data aggregation algorithm. Chinese Journal of Electronics, 2017, 26(4): 740-746

[27] Zhou Qiang, Yang Geng, Li Sen, et al. An integrity-checking private data aggregation algorithm. Journal of Electronics & Information Technology, 2013, 35(6): 1277-1283(in Chinese)
(周强，杨庚，李森等. 一种可检测数据完整性的隐私数据融合算法. 电子与信息学报，2013，35(6)：1277-1283)

[28] Wu D, Yang B, Wang H, et al. Privacy-preserving multimedia big data aggregation in large-scale wireless sensor networks. ACM Transactions on Multimedia Computing, Communications, and Applications, 2016, 12(4s): 60:1-60:19

[29] Lee J S, Hoh B. Dynamic pricing incentive for participatory sensing. Pervasive and Mobile Computing, 2010, 6(6): 693-708

[30] Han K, Zhang C, Luo J, et al. Truthful scheduling mechanisms for powering mobile crowdsensing. IEEE Transactions on Computers, 2016, 65(1): 294-307

[31] Nan Wen-Qian, Guo Bin, Chen Hui-Hui, et al. A cross-space, multi- interaction-based dynamic incentive mechanism for mobile crowd sensing. Chinese Journal of Computers, 2015, 38(12): 2412-2425(in Chinese)
(南文倩，郭斌，陈荟慧等. 基于跨空间多元交互的群智感知动态激励模型. 计算机学报，2015，38(12)：2412-2425)

[32] Huang Hao-Ran, Cai Ken. A method of fuzzy multiple attribute decision making based on the error-eliminating theory. Journal of Intelligent & Fuzzy Systems, 2016, 31(4): 2119-2127

[33] Wang W, Zhu J, Zhang S, et al. Tradeoff between efficiency and delay of distributed source coding for uplink transmissions in machine type communications//Proceedings of the 9th International Conference on Wireless Communications and Signal Processing. Nanjing, China, 2017: 1-6

[34] Wang Xue-Wei, Dong Xiao-Xuan, Yuan Rui-Ming, et al. Compressive sensing measurement for electrical energy of pseudo random dynamic test signal. Journal of Electronics & Information Technology, 2017, 39(3): 640-646(in Chinese)
(王学伟，董晓璇，袁瑞铭等. 压缩感知伪随机动态功率信号的电能测量方法. 电子与信息学报，2017，39(3)：640-646)

[35] Bottcher A. Orthogonal symmetric Toeplitz matrices. Complex Analysis and Operator Theory, 2008, 2(2): 285-298

[36] Guo Jun-Feng, Dang Jiang-Ting. Data compression collecting method for vibration signals based on optimal deterministic measurement matrix. Journal of Vibration and Shock, 2019, 38(7): 195-203(in Chinese)
(郭俊锋，党姜婷. 基于最优型确定性测量矩阵的振动信号数据压缩采集方法. 振动与冲击，2019，38(7)：195-203)

[37] Li Guo-Rui, Meng Jie, Peng San-Cheng, et al. A distributed data reconstruction algorithm based on Jacobi ADMM for compressed sensing in sensor networks. Journal of Computer Research and Development, 2020, 57(6): 1284-1291(in Chinese)
(李国瑞，孟婕，彭三城等. 基于 Jacobi ADMM 的传感网分布式压缩感知数据重构算法. 计算机研究与发展，2020，57(6)：1284-1291)

LIANG Yan, M. S., lecturer. Her research interests focus on crowd sensing.

AN Jian, Ph. D., senior engineer. His research interests include crowd sensing, Internet of Things.

HU Xian-Zhi, M. S., engineer. His research interests include Internet of Things, information security.

YANG Qian, M. S. candidate. Her research interests focus on crowd sensing.

SI Hai-Feng, M. S., senior engineer. His research interests focus on crowd sensing.

Background

With the explosive popularity of mobile smart devices, mobile crowd sensing (MCS) is becoming an effective means to solve large-scale and complex social sensing tasks. Accordingly, the application of crowd sensing is more and more extensive, and has attracted the attention of many researchers. MCS needs a large number of users to participate in sensing tasks, and hopes that users can collect high-quality sensed data and provide better service quality to the public. However, when the user participates in sensing tasks, the cost is relatively high because it not only consumes the user's time, but also consumes the user's equipment resources such as power, computing, storage, and communication. In addition, sensed data collected by the user may carry sensitive information related to the user, such as identity information, living habits, health status and social relations, which results in the user to face the risk of personal privacy leakage, and the privacy leakage hinders the user' enthusiasm to participate. Therefore, we studied the comprehensive mechanism of privacy-preserving and compensation incentive to MCS, so as to effectively compensate the user for the cost of participating in sensing tasks, and reduce the user's concerns about privacy leakage, thereby attracting more users to participate and mobilizing their enthusiasm for a long time. It is of great significance to further expand the application of crowd sensing. In the existing research, the research on the comprehensive scheme or the integrated system framework that combines privacy-preserving and incentive mechanism is still a minority. Many studies usually focus on the user's unilateral privacy-preserving or compensation incentive, and these studies don't fully consider the communication consumption caused by the transmission of a large number of redundant sensed data, and the risk of the user's privacy leakage under the incomplete trusted sensing platform. These problems have a direct impact on the long-term development of MCS system, and it is also a research topic that needs close attention at present.

In order to solve these problems, we carry out our research work from three stages: participant registration, winner selection, and sensed data aggregation. Taking advantage of the advantages of lightweight privacy-preserving and distributed compressed sensing technology, we proposed an incentive mechanism IMPP with privacy-preserving in MCS. Firstly, based on the idea of lightweight privacy-preserving, it uses a one-way hash function to generate a 256-bit hash value as an anonymous identity of the participant. Secondly, according to the participant's data utility value, expected reward and the sensing task total budget, it has achieved the optimal selection of participants for data quality. Next, with the help of distributed compressed sensing theory and methods it has compressed and observed the winner's original sensed data and removed redundant data. By adding noise data such as hash values to the observed value, the privacy-preserving of the sensed data can be enhanced. After the winner's privacy sensed data is aggregated at the application server (AS), the application server performs integrity verification and reconstructs the original sensed data. Finally, using the real data sets, the effectiveness of IMPP is compared and analyzed through simulation experiments. The experimental results show that the IMPP mechanism is efficient in terms of privacy-preserving level, data integrity, data accuracy, time efficiency, evaluation accuracy, reconstruction matching degree and incentive effect.

This work was partly supported by the National Key Research and Development Project (No. 2018YFB1800304), the National Natural Science Foundation of China (No. 61502380), the Key Development Program in Shaanxi Province of China (Nos. 2020GY-033, 2019GY-005, 2017ZDXM-GY-011) and the Special Scientific Research Program of Department of Education in Shaanxi Province (No. 19JK0686).