计算机科学
COMPUTER SCIENCE

群智感知的隐私保护研究综述

李利，何欣，韩志杰

引用本文

李利，何欣，韩志杰. 群智感知的隐私保护研究综述[J]. 计算机科学, 2022, 49(5): 303-310.

LI Li, HE Xin, HAN Zhi-jie. Review of Privacy-preserving Mechanisms in Crowdsensing[J]. Computer Science, 2022, 49(5): 303-310.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210400077

群智感知的隐私保护研究综述

李　利[1]　何　欣[2,3]　韩志杰[3]
1 河南大学计算机与信息工程学院　河南 开封 475004
2 河南大学智能网络理论与关键技术国际联合实验室　河南 开封 475004
3 河南大学软件学院　河南 开封 475004
(kathleenlee@126.com)

摘　要　近年来,智能终端的快速普及极大地推动了集数据采集、分析、处理于一体的群智感知服务的发展。隐私保护作为保障服务安全运行和鼓励感知用户参与的必要手段,成为需要解决的首要科学问题。文中首先从群智感知的全生命周期出发,在描述其主要组成部分和业务流程之后,再从群智感知场景对隐私保护的特有需求出发,对隐私保护的定义和衡量指标进行讨论,并对现有文献设计的隐私保护机制所侧重的不同阶段进行分类,从隐私保护范围、保护强度、感知用户身份可追溯、感知数据损失和感知终端能耗的角度对文献使用的隐私保护机制进行讨论。在此基础上对文献使用的实验数据集进行梳理,最后结合群智感知应用的发展需求和全球对隐私保护的监管要求提出未来研究面临的挑战。

关键词:群智感知;群智计算;隐私保护;密码学;匿名化
中图法分类号　TP393

Review of Privacy-preserving Mechanisms in Crowdsensing

LI Li[1], HE Xin[2,3] and HAN Zhi-jie[3]
1 School of Computer and Information Engineering, Henan University, Kaifeng, Henan 475004, China
2 International Joint Laboratory of Intelligent Network Theory and Key Technology, Henan University, Kaifeng, Henan 475004, China
3 School of Software, Henan University, Kaifeng, Henan 475004, China

Abstract　In recent years, the rapid popularity of intelligent terminals has greatly promoted the development of crowdsensing service paradigm, which integrates data collection, analysis and processing. As a necessary base to ensure the safe operation of services and encourage the participation of sensing users, privacy-preserving has become the primary issue to be solved. This paper presents the state-of-the-art in privacy-preserving mechanisms for crowdsensing service. After describing its main components, this paper discusses the definition and metrics of privacy-preserving from the view of crowdsensing's whole life cycle. The privacy-preserving mechanisms designed in literatures are analyzed and discussed according to different stages in crowdsensing's whole-life-cycle, and the experimental datasets used in literatures are given. Finally, Future research challenges are proposed based on the development of crowdsensing and global regulatory requirements for privacy-preserving.

Keywords　Crowdsensing, Crowdcomputing, Privacy-preserving, Encryption, Anonymization

1　引言

据 Gartner 统计,2021 年全球智能手机销量达到了 15 亿部[1],全球可穿戴设备的总支出达到了 815 亿美元[2],智能手机和可穿戴设备的快速普及极大地推动了群智感知(Crowd Sensing,CS)应用[3-4]的发展,并且被广泛应用于智慧城市[5]、公共安全[6]、地图导航[7]、临床医疗[8]等领域。其显著的优点之一是无须提前部署静态传感网络,可以快速、高效、低成本地从用户移动可达的范围内收集数据,并进行分析、处理和群体智能提取,具有更大的时空覆盖范围和更好的上下文感知能力。Reddy 等[9]通过采集日常饮食的图像和进餐时间、地点等上下文信息来跟踪用户的饮食健康状况;Liu 等[10]将手机作为 PM2.5 监测仪,通过手机拍摄的图像和数据对空气污染状况进行检测,如检测城市细颗粒物 PM2.5;Kim 等[11]通过收集的河床水量、垃圾量、河水流速和水道图片等数据来监测河流状况;Bonino 等[12]通过群智感知的模式来监控垃圾

到稿日期:2021-04-08　返修日期:2021-07-20
基金项目:国家自然科学基金(61672209,61701170);河南省重大科技专项(201300210400);河南省重点研发与推广专项(212102210094)
This work was supported by the National Natural Science Foundation of China(61672209,61701170), Major Science and Technology Special Project of Henan Province(201300210400) and Key R&D and Promotion Special Project of Henan Province(212102210094).
通信作者:何欣(hxsyjkf@foxmail.com)

回收箱的物品，从而改善回收计划；Thomas 等[13]在马萨诸塞州波士顿哈佛桥的应用中，使用移动车辆中的智能手机收集的加速度数据来检测桥梁的模态频率，为地方政府低成本采集桥梁振动数据和管理决策提供了有效途径。

虽然 CS 有诸多好处，但同时也面临着许多挑战。鉴于群智感知应用的多样性、感知模式的开放性和感知用户的不同可靠性，群智感知中隐私安全的问题亟需解决。例如，谷歌地图为了生成实时的交通地图，需要收集司机的“匿名”位置信息，但仍然会暴露司机的行驶路线和轨迹。感知用户提交的感知数据不可避免地带有感知上下文的时空信息，即任务执行的时间和位置信息，攻击者可能利用这些时空信息推导出感知用户的生活习惯、行为规范等。若结合其提供的感知数据，可能推断出专业、教育程度、年龄、性别、语言、偏好甚至性格等个人敏感信息，一旦这些信息被恶意攻击者窃取，可能会导致严重的隐私泄露甚至人身攻击[14]。

专家学者们围绕群智感知场景中的任务分配[15-16]、质量评估[17-18]、激励机制[19-20]等环节已展开了深入的研究，隐私保护作为鼓励感知用户参与和保障任务执行的基础工作尤其受到关注。Vergara-Laurens 等[21]对隐私保护进行了综述，并就隐私保护机制的设计、实现和评估中需要考虑的重要问题进行了讨论；Khan 等[22]对群智感知应用中的隐私保护、任务管理、任务模型、激励机制进行了综述，并提出基于激励机制且受到高效隐私保护的任务分配和管理机制有望在未来得到快速发展。但现有文献并未从 CS 全生命周期的角度对隐私保护机制进行分析。

本文首次从群智感知全生命周期的视角出发，对近年来文献设计的隐私保护机制进行综述，后续部分结构如下：第 2 节介绍群智感知的典型场景及其面临的隐私保护挑战；第 3 节综合对近年来文献的分析，给出 CS 场景下隐私保护的定义及衡量指标，对文献所采用的的隐私保护机制进行讨论，并梳理了现有文献使用的实验数据集；第 4 节提出未来研究面临的挑战；最后总结全文。

2 问题和挑战

群智感知是以用户及其智能终端（如手机、可穿戴设备、智能汽车）为载体，以大量普通用户参与为基础，将所采集的大规模数据上传至任务处理平台，集数据采集、分析、提取群体智能为一体的服务模式[23-24]。典型的群智感知场景（见图 1）有 3 个角色：任务处理平台、任务请求者和感知用户，其中任务请求者和感知用户在不同的感知任务中身份可以互换。任务请求者将感知任务提交给任务处理平台，接受任务处理平台返回的结果，并提供激励成本。任务处理平台接受服务请求者发布的感知任务，根据任务特征选择不同的感知用户，完成对参与任务的感知用户的招募与选择，根据感知用户提交感知数据进行质量评估和激励支付，并进行群体智能的提取。感知用户执行感知任务后提交感知数据并根据质量评估的结果获取一定收益。

由于在任务分配和任务执行的过程中，感知用户不可避免地会暴露其身份、位置等敏感信息，感知数据在传输的过程中也面临泄露的风险，加上近年来隐私泄露事件的不断发生，使得隐私保护成为群智感知应用中需要解决的首要科学问题。感知用户执行任务时会消耗时间和物理资源，如电池电量、存储容量和通信带宽，因此，如果没有足以冲抵成本的报酬，感知用户将不愿意参与任何任务（在群智感知应用中，假设感知用户是理性的）。感知任务能否成功执行的关键因素包括用户的参与意愿和所提交感知数据的质量，因此，为了鼓励感知用户提交优质的数据，群智感知应用通常根据感知用户提交数据的质量进行激励支付，以提高优质感知用户参与任务的意愿，故对用户身份的隐私保护提出了可追溯的要求。同时，感知用户参与的感知任务越多，所贡献的数据越丰富，其敏感信息暴露的风险也就越大。感知用户携带的智能终端多为智能手机、可穿戴设备等，其能源和算力等资源通常受限，这又对群智感知场景下感知终端参与的隐私保护技术提出了低能耗的要求。

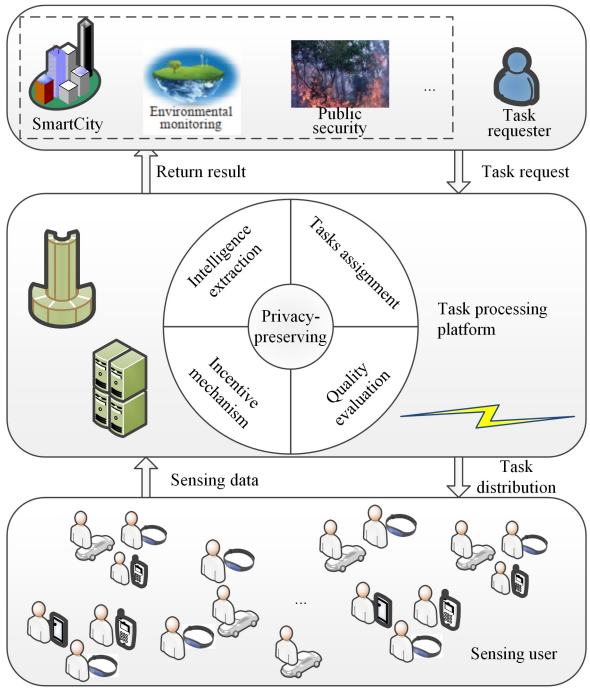

图 1 典型的群智感知场景

Fig. 1 Typical crowdsensing scenario

因此，提出既能满足低能耗和可溯源需求、又能防止非授权用户将感知数据或位置信息与感知用户身份信息相联系的隐私保护机制，对促进群智感知系统的长期稳定发展具有重要意义。

Ganti 等在提出群智感知这一服务模式时给出了针对隐私保护问题的 3 种研究方向：匿名化、密码学和数据扰动。匿名化指在将数据发送给任务处理平台前隐藏身份信息；密码学指使用密码技术转换数据；数据扰动指将数据提交至任务处理平台之前，在不影响数据可用性的前提下将噪声信息添加到数据中。

随着群智感知应用的快速发展和所面临环境的愈发复杂，Hui 等[25]提出仅有匿名化在保护隐私方面是不够的，因为攻击者可能会通过感知用户的旅行路线、社会关系等进行追踪。通过对美国 50 个州 2500 万手机用户的逾 300 亿条通话记录的研究发现，从通话记录中可以推断每个用户使用最频繁的“前 N”位置，将这些信息与公开信息（如人口普查）相关联，可以根据“前 2”位置唯一地识别出 35% 的用户，根据

"前 3"位置唯一地识别出 85%的用户。因此,在群智感知中,探索强隐私保护机制,同时防止将感知用户的身份、位置和数据相关联具有重要意义。

综合文献分析发现,隐私保护中的数据扰动技术多用于对感知用户位置信息的保护[26-27],密码学技术多用于对感知数据的加密[28-33],匿名化技术多用于感知用户身份信息的隐藏[34-39]。研究者们根据实际需要通常将多种隐私保护技术结合使用并不断地探索新的隐私保护机制。图 2 根据群智感知生命周期的不同阶段,分别呈现了本文分析文献采用的隐私保护技术。

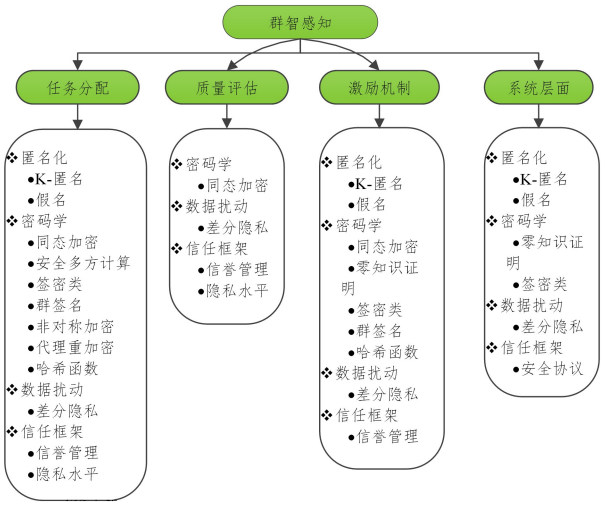

图 2　群智感知不同阶段使用的隐私保护技术

Fig. 2　Privacy-preserving technologies in Crowdsensing' different stages

3　研究现状分析

3.1　群智感知场景下隐私保护的定义

综合对近年来文献所设计的隐私保护机制的理解,群智感知场景中的隐私保护指在对感知用户的身份、位置信息和感知数据实施隐私保护之后,除了被授权方,任何非授权方都不能获取隐私保护内容,不能确认位置信息和感知数据所对应的身份信息(所有者),同时不降低位置信息和感知数据的可用性。

问题定义。1)身份信息隐私性:感知用户的姓名、邮箱、手机号码和个人偏好等信息只有被授权方可以获取。2)位置信息隐私性:感知用户的位置及其行动轨迹只有被授权方可以获取。3)感知数据隐私性:感知数据被上传至任务处理平台处理,只有被授权方可以获取数据并了解数据的所有者。

假定参加任务的感知用户数量为 n,I 表示感知用户的身份信息集合,有 $I=\{i_1,i_2,\cdots,i_n\}$;L 表示感知用户的位置信息集合,$L=\{l_1,l_2,\cdots,l_n\}$;D 表示感知数据集合,$D=\{d_1,d_2,\cdots,d_n\}$。I^{en} 表示隐私保护后的感知用户的身份信息集合,有 $I^{en}=\{i_1^{en},i_2^{en},\cdots,i_n^{en}\}$;$L^{en}$ 表示隐私保护后的感知用户的位置信息集合,有 $L^{en}=\{l_1^{en},l_2^{en},\cdots,l_n^{en}\}$;$D^{en}$ 表示隐私保护后的感知数据集合,有 $D^{en}=\{d_1^{en},d_2^{en},\cdots,d_n^{en}\}$。

f_i,f_l,f_d 分别为感知用户身份信息、位置信息和感知数据的隐私保护函数,均为单向且抗冲突的,有:

$$f_i(i_k\mid i_k\in I)=i_k^{en}(i_k^{en}\in I^{en},k\in[1,n]) \tag{1}$$

$$f_l(i_k\mid i_k\in L)=l_k^{en}(l_k^{en}\in L^{en},k\in[1,n]) \tag{2}$$

$$f_d(d_k\mid d_k\in D)=d_k^{en}(d_k^{en}\in D^{en},k\in[1,n]) \tag{3}$$

$f_i^{de},f_l^{de},f_d^{de}$ 分别表示授权手段之外可以由 i_k^{en} 获得 i_k,l_k^{en} 获得 l_k,d_k^{en} 获得 d_k 的函数;f_{li}^{de},f_{di}^{de} 分别表示授权手段之外可以由 l_k^{en} 获得 i_k,d_k^{en} 获得 i_k 的函数,则:

$$f_i^{de}(i_k^{en}\mid i_k^{en}\in I^{en})\neq i_k(i_k\in I,k\in[1,n]) \tag{4}$$

$$f_l^{de}(l_k^{en}\mid l_k^{en}\in L^{en})\neq l_k(l_k\in L,k\in[1,n]) \tag{5}$$

$$f_d^{de}(d_k^{en}\mid d_k^{en}\in D^{en})\neq d_k(d_k\in D,k\in[1,n]) \tag{6}$$

$$f_{li}^{de}(l_k^{en}\mid l_k^{en}\in L^{en})\neq i_k(i_k\in I,k\in[1,n]) \tag{7}$$

$$f_{di}^{de}(d_k^{en}\mid d_k^{en}\in D^{en})\neq i_k(i_k\in I,k\in[1,n]) \tag{8}$$

3.2　群智感知场景下隐私保护的衡量指标

研究者们就群智感知生命周期不同阶段的隐私保护展开了系统深入的研究,如任务分配[15,26-28,40-42]、质量约束[17-18,32,35,43-46]、激励机制[19-20,33-34,44],但尚未有文献就 CS 全生命周期的隐私保护机制进行衡量。基于此,本文从群智感知全生命周期出发,针对群智感知场景的特有需求,从隐私保护范围、隐私保护强度、感知用户身份可追溯、感知数据损失和感知终端能耗的角度对文献采用的隐私保护机制进行分析。从 CS 全生命周期视角出发的衡量指标打破了现有文献立足于某一阶段进行系统设计的局限,而忽略了其他阶段的关键因素可能造成的系统整体运行效率不高,甚至顾此失彼的状况。该衡量指标体系的建立将为群智感知服务系统的整体设计提供参考,为实际运营的鲁棒性扫除障碍,进而在智慧城市[5]、公共安全[6]的管理决策中发挥更大的作用。

3.2.1　隐私保护范围

群智感知场景中隐私保护内容主要涉及感知用户的身份信息、位置信息和感知数据,因一种隐私保护技术难以实现对这三者的保护,故研究者大多采用几种隐私保护技术相结合的方式进行隐私保护机制的设计。文中将隐私保护范围定义为 $Scope\in\{\text{low},\text{medium},\text{high}\}$。其中 low 表示受隐私保护的是感知用户身份、位置或感知数据三者中的任意一个;medium 表示保护三者中的任意两个;high 表示三者都受到保护。

3.2.2　隐私保护强度

在实施隐私保护之后,非授权方不能根据隐私保护后的位置信息或感知数据推知其所有者,文中将其定义为隐私保护强度 $Inten\in\{\text{low},\text{medium},\text{high}\}$。其中 low 表示文献未提及该项内容或可由位置信息或感知数据推知其所有者;medium 表示不可由位置信息推知其所有者或不可由感知数据推知其所有者,即式(7)成立或式(8)成立;high 表示既不可由位置信息推知其所有者也不可由感知数据推知其所有者,即式(7)和式(8)均成立。

3.2.3　感知用户身份可追溯

由于感知用户执行任务时存在隐私泄露的风险,同时要付出时间和资源成本,若不能获取相应的回报,用户将不愿意参与感知任务,故激励支付时被授权方要能根据感知数据追溯到其所有者。文中将身份可追溯定义为 $\text{Trace}\in\{\text{yes},\text{no}\}$,其中 yes 表示被授权方可以根据数据推断出所有者,no 表示被授权方无法根据数据推断出所有者。

3.2.4　感知数据损失

获取高质量的感知数据是群智感知应用的核心目标

之一，但在数据隐私保护的过程中，为了兼顾能耗需求，有些研究者会在数据的隐私保护和质量损失之间做出妥协。例如，采用在原始数据中添加数据扰动的方式来保护隐私，使得数据的准确性受到影响。文中将感知数据质量损失定义为 $DLoss \in \{low, medium, high\}$，其中 low 表示感知数据在感知用户端通过明文或采用密码学技术加密后传递，medium 表示感知数据在感知用户端采用数据扰动方式保护后传递，high 表示感知用户端的感知数据未全部传递出去。

3.2.5 感知终端能耗

感知终端通常在电池、内存、计算、存储和通信能力等方面受到严重限制。密码学技术要求终端有较强的计算能力和较大的内存空间，感知数据通过无线网络传输时将耗费终端的大量能源[45]。本文将综合感知终端在电池电量、存储容量、计算需求和通信传输上的要求，以能耗指标的方式体现，并将感知终端的能耗定义为 $ECons \in \{low, medium, high\}$。其中 low 表示感知终端既不需要额外耗费太多的算力，也不需要额外传输太多数据；medium 表示感知终端需要额外耗费算力，或者需要额外传输数据；high 表示感知终端既需要额外耗费算力，又需要额外传输数据。

3.3 群智感知各阶段的隐私保护机制

表 1 列出了近年来文献设计的隐私保护机制所采用的 4 类隐私保护方法中的具体技术手段，4 类隐私保护方法分别为匿名化、密码学、数据扰动和信任框架。亦有文献将区块链技术[40,44,46]作为一类单独提出，但因其采用的是区块链技术的非对称加密技术或匿名化技术，故本文将其相应地归为密码学或匿名化类别，未将区块链技术作为单独的一类列出。

表 1 文献采用的隐私保护技术

Table 1 Encription Tachniques in the Literatures

文献	匿名化	密码学	数据扰动	信任框架
Wang 等[27]			(3.1)	(4.2)
Xiao 等[28]		(2.2)		
Shen 等[26]	(1.2)		(3.1)	
Yang 等[40]	(1.2)	(2.6)		
Wu 等[48]		(2.1)(2.4)	(3.1)	
Ni 等[41]		(2.1)(2.4)(2.8)		
Wu 等[42]	(1.1)	(2.5)		
Ni 等[47]		(2.4)(2.7)		(4.1)
An 等[46]	(1.2)	(2.1)(2.6)		
Ma 等[29]		(2.1)		(4.1)
Yang 等[30]			(3.1)	
Xiong 等[31]		(2.1)		(4.2)
Miao 等[32]		(2.1)		(4.1)
Sun 等[33]		(2.1)(2.3) (2.4)(2.8)		(4.1)
Liang 等[34]	(1.2)	(2.8)		
Wu 等[35]	(1.2)	(2.5)		(4.1)
Wang 等[44]	(1.1)	(2.4)		
Lin 等[52]			(3.1)	(4.1)
Wang 等[36]	(1.1)			
Basudan 等[37]	(1.2)	(2.4)		(4.3)
Sucasas 等[38]	(1.2)	(2.3)(2.4)		
Xiong 等[39]	(1.1)		(3.1)	

(1.1)k-anonymous (1.2)pseudonym
(2.1)homomorphic encryption (2.2)Secure Multi-party Computation
(2.3)zero-knowledge proof (2.4)signcryption et al. (2.5)group signature (2.6)asymmetric encryption (2.7)proxy re-encryption
(2.8)hash function
(3.1)differential privacy
(4.1)reputation management (4.2)privacy level (4.3)secure protocol

表 2 列出了从群智感知全生命周期的角度对文献设计的隐私保护机制进行指标衡量的结果。以下分别就文献所侧重的任务分配、质量约束、激励支付和系统层面 4 个阶段对文献进行详细分析。

表 2 文献设计的隐私保护机制的衡量指标

Table 2 Metrics of literature's privacy-preserving mechanism

Literature	*Scope*	*Inten*	*Trace*	*Dloss*	*Econs*	*Technologies*
Wang 等[27]	low	low	yes	low	low	③ ④
Xiao 等[28]	low	medium	yes	low	high	②
Shen 等[26]	low	low	yes	low	low	①③
Yang 等[40]	high	medium	yes	low	high	① ②
Wu 等[48]	high	medium	yes	low	high	②③
Ni et al[41]	medium	medium	yes	low	high	②
Wu 等[42]	medium	medium	yes	low	high	① ②
Ni 等[47]	medium	medium	yes	low	high	② ④
An 等[46]	medium	medium	yes	low	high	①②
Ma 等[29]	medium	medium	yes	low	high	②④
Yang 等[30]	low	low	yes	medium	low	③
Xiong 等[31]	low	low	yes	low	high	③④
Miao 等[32]	medium	low	yes	low	high	② ④
Sun 等[33]	high	medium	yes	low	high	② ④
Liang 等[34]	medium	medium	yes	low	high	①②
Wu 等[35]	low	medium	yes	low	high	①②④
Wang 等[44]	low	medium	yes	low	high	①②
Lin 等[52]	medium	low	yes	medium	medium	③④
Wang 等[36]	medium	medium	yes	medium	low	①
Basudan 等[37]	high	medium	yes	low	high	①②④
Sucasas 等[38]	low	low	yes	low	high	① ②
Xiong 等[39]	medium	medium	yes	medium	medium	①③

①anonymization；②cryptography；③data perturbation；④trust framework

3.3.1 任务分配

群智感知应用在任务分配阶段解决的主要问题是由任务处理平台根据感知任务特点将其与感知用户的服务属性进行匹配，选出可能提供优质感知数据的用户参与感知任务的执行，在此基础上尽可能兼顾感知用户的感知成本最小化和任务请求者的激励成本最小化。该阶段感知用户的身份、位置等敏感信息需要与任务处理平台交互，因而面临更大的泄露风险。

文献[26-28,40-42,46-48]均对隐私保护下的任务分配进行了研究，其中文献[27-28,47]采用传统中央平台的架构，文献[27]提出了一种个性化隐私保护的任务分配框架，工人将模糊后的距离和个人隐私级别上传到服务器进行任务分配，该框架采用数据扰动和信任框架两种隐私保护方法。文献[28]设计了一种基于贪婪策略的秘密共享用户招募(BUR)协议，该方案不依赖于加密/解密操作和任何受信任的第三方，采用密码学中秘密共享的多方安全计算技术。文献[47]提出了一种基于移动用户地理信息和信用积分的具备精确任务分配和强隐私保护能力的方案 SPOON(Strong Privacy-preserving mObile crOwdseNsing schem)，利用代理重加密和 BBS+[49]签名技术，保护感知任务，防止隐私泄露。SPOON 采用密码学和信任框架的隐私保护方法保护感知用户的身份和位置信息，并阻止未授权方通过位置信息推知其所有者。

文献[26,41-42,48]采用边缘节点辅助的方式进行任务分配，其中，文献[26] 引入边缘节点作为匿名服务器对参与者位置数据进行聚合和模糊处理以保护位置隐私，通过基于边缘节点的遗传算法来选择初始模糊策略，并利用斯坦尔博格

的隐私博弈模型获得不受后验推理攻击的最终模糊策略，该框架采用了匿名化和数据扰动的隐私保护方法对位置信息进行保护；文献[48]通过边缘节点将大量任务高效而准确地进行分配并对感知数据进行聚合，该框架采用了密码学和数据扰动的隐私保护方法；文献[41]设计了边缘辅助的BLS-PORF方案以去除冗余的感知数据，该方案通过BLS签名的方式生成边缘计算可识别的相同感知数据密钥，采用密码学方法的多种技术，利用盲签名切断感知数据与感知用户之间的联系，利用变色龙哈希函数实现感知用户的身份可追溯；文献[42]根据任务的类型、时间窗口、激励等特征使用高斯混合模型对用户进行聚类，为用户隐私提供k匿名保护，利用群签名[45]对聚类的集群用户进行批量验证，该方案采用匿名化和密码学的隐私保护方法对感知用户的身份和感知数据进行保护。

文献[40,46]基于区块链的分布式架构实现任务分配，其中，文献[40]基于激励模式来实现任务的合理分配，并使用区块链技术的匿名特征来隐藏用户的身份信息和位置隐私，同时为每个任务建立私有链，避免非授权方推知感知数据的所有者。该机制采用匿名化和密码学的隐私保护方法。文献[46]基于区块链的地址转换和数字签名技术为链上的每个感知用户生成匿名身份和签名，用轻量级同态加密算法[50-51]对感知终端的服务属性进行加密。该机制采用匿名化和密码学的方法保护感知用户的身份和感知数据，并阻止未授权方通过位置信息推知其所有者。

综合以上分析并结合图2和表1可知，在任务分配阶段研究者所采用的隐私保护技术几乎涵盖了群智感知全生命周期所使用的的隐私保护技术。

3.3.2　质量评估

获取高质量的感知数据是群智感知应用的关键环节之一。只有获取高质量的感知数据才能为群体智能提取阶段的数据处理提供良好的样本数据，只有根据感知用户所贡献的感知数据的质量进行激励支付才能更好地鼓励优质用户参与任务。因此质量评估是群智感知任务得以高效执行的关键阶段。

文献[29-32]对隐私保护下的质量评估工作进行了研究，其中[29,31-32]均采用了密码学的同态加密技术和信任框架的方法。文献[29]提出了两种保护隐私的声誉管理方案，该方案在保护隐私的同时应对恶意参与者，且云服务器不能根据接收到的感知数据推断出参与者的原始感知数据，更无法推知数据所有者；文献[31]提出了一种基于博弈论和数据加密的个性化隐私保护框架，根据感知用户的历史时空轨迹信息提出基于公共属性和个性化属性的动态度量算法；文献[32]提出了一种隐私保护真值发现框架PPTD，该框架不仅可以保护感知数据，还可以保护用户的信誉分数；文献[30]设计了一种隐私保护的感知数据聚合平台，通过添加噪声保护感知数据，分析每个感知用户添加不同的噪声对数据隐私保护和聚合结果准确性的影响。该平台采用数据扰动的方法保护感知数据。

3.3.3　激励机制

由于感知用户在执行任务的过程中不可避免要付出时间和资源成本，并且面临隐私泄露的风险，基于理性感知用户的假设：即除非获得足够的收益，否则感知用户不愿意参与感知任务，因此激励机制便成为了吸引感知用户尤其是优质用户参与任务的关键手段，且需要同时兼顾激励成本的最小化需求。

文献[33-36,44,52]就隐私保护的激励机制进行了研究，其中文献[33,35,52]均采用了信任框架的方法选择优质的感知用户，文献[33]同时采用零知识验证、单向哈希、部分盲签名认证和同态加密等多种认证和加密技术，通过高效的隐私保护协议实现数据上传、激励支付、信任管理等环节的隐私保护；文献[35]使良性用户通过群签名的方式匿名参与任务并通过信誉模型进行激励；文献[52]构建了两个基于拍卖的隐私保护激励机制框架，同时近似地实现了差分隐私和社会成本最小化。文献[34,44]均采用匿名化和密码学方法保护感知用户身份和感知数据，其中文献[34]利用安全加密哈希函数为竞标的感知用户生成可变地址序列，作为注册编号实现用户匿名参与，参与者每次竞标时采用不同的匿名编号和公钥，以防止参与者和报价之间的链接攻击；文献[44]提出了一种节点合作验证方法来实现k匿名[53]的隐私保护，并采用基于双线性映射的签密方案[54]来完成对匿名组内节点用户的验证，从而实现用户身份的可追溯。文献[36]提出一种位置聚合方案，该方案将用户分组，通过k匿名方式来保护位置隐私的同时减少信息丢失，并在此基础上通过群体值和感知代价来设计激励机制以鼓励高效的感知用户参与感知任务。

3.3.4　系统层面

文献[37-39]从系统层面的隐私保护需求进行设计，其中文献[37,39]借助边缘节点解决群智感知场景的隐私泄露问题，文献[37]提出了一种高效的无证书聚合签密隐私保护协议，系统中的移动传感器使用由其真实身份生成的伪身份来实现匿名，保护感知用户身份、位置信息和感知数据，并阻止未授权方推知感知数据的所有者。该系统采用匿名化、密码学和信任框架的隐私保护方法。文献[39]基于随机森林分类器[55]和k匿名算法提出一种三方博弈框架。该框架采用匿名化和数据扰动的方法保护感知用户身份和感知数据，并阻止未授权方由感知数据推知其所有者。文献[38]通过使用不可链接但可问责的假名机制，允许用户在自行生成不限数量假名的同时参加多个任务。该方案采用匿名化和密码学的方法。

3.4　实验数据集

实验环节是研究者对所做科学假设的检验阶段，是理论转化为实践的重要环节，在整个研究中起着极其重要的作用。实验还以数理统计专业知识为基础，因此研究者需要根据所设计的实验使用适合的实验数据集，通过对实验数据集的测试和对结果的分析来评估实验的效果并改进实验方案，进而指导研究，因此实验数据集对科学研究具有重要意义。科学合理的实验数据集可以使实验达到事半功倍的效果，帮助研究者从纷乱的数据中找出事物的内在规律，因此本文梳理了部分文献使用的数据集，详见表3。表中第一列为使用数据集的文献，第二列为数据集名称，第三列为数据集的超链接或其来源文献。

表 3 文献采用的实验数据集

Table 3 Experimental datasets used in literatures

Literature	Dataset	Hyperlink or Source
[27]	check-in data	D. Yang, D. Zhang, V. W. Zheng, and Z. Yu, "Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 45, no. 1, pp. 129-142, 2015.
[40]	Yelp dataset challenge	https://www.yelp.com/dataset_challenge
[33]	a geo-tag dataset	D. Yang, D. Zhang, V. W. Zheng, and Z. Yu, "Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 45, no. 1, pp. 129-142, 2015.
[46]	capital shared bike	2017-capitalbikeshare-tripdata. [Online]. Available: https://s3.amazonaws.com/capitalbikeshare-data/index.html
	UCI machine learning	D. Dua and E. K. Taniskidou, UCI Machine Learning Repository, School Inf. Comput. Sci., Univ. California at Irvine, Irvine, CA, USA, 2017. [Online]. Available: http://archive.ics.uci.edu/ml
[31] [39]	GeoLife project	Y. Zheng, X. Xie, and W. Ma, "Geolife: a ollaborative social networking service among user, location and trajectory," Bulletin of the Technical Committee on Data Engineering, vol. 33, no. 2, pp. 32-39, 2010.
[52]	taxi traces	L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi, "CRAWDAD data set roma/taxi (v. 2014-07-17),"
[36]	indoor locations	A. Purohit, S. Pan, K. Chen, Z. Sun, and P. Zhang, CRAWDAD dataset cmu/supermarket (v. 2014-05-27), http://crawdad.org/cmu/supermarket/20140527, doi: 10.15783/C7MW2Z, May 2014.
[57]	Cab (SFC) Dataset	M. Piorkowski, et al., "Dataset of mobility traces of taxi cabs in San Francisco, USA," CRAWDAD dataset epfl/mobility (v. 2009-02-24), http://crawdad.org/epfl/mobility/20090224.
[43]	indoor points distance	Q. Li, Y. Li, J. Gao, L. Su, B. Zhao, M. Demirbas, W. Fan, and J. Han, "A confidence-aware approach for truth discovery on long-tail data," PVLDB, vol. 8, no. 4, pp. 425-436, 2014.
[59]	GPS mobility traces	"Cabspotting Project." [Online]. Available: http://cabspotting.org/. "CRAWDAD: a community resource for archiving wireless data at artmouth." [Online]. Available: http://crawdad.cs.dartmouth.edu/
[60]	outdoor temperature sensing traces	M. A. Alswailim, H. S. Hassanein, and M. Zulkernine, "CRAWDAD dataset queensu/crowd temperature (v. 2015-11-20)," Downloaded from http://crawdad.org/queensu/crowd temperature/20151120, 2015.
[61]	Microsoft GeoLife	Y. Zheng., X. Xie, W. Y. Ma. " GeoLife: A Collaborative Social Networking Service among User, location and trajectory," IEEE Data Engineering Bulletin, vol 33, no. 2, pp. 32-40, 2010.
[62]	StatLib	Statlib. [Online]. Available: http://lib.stat.cmu.edu
	UCI Machine Learning	M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

4 未来研究方向

综合分析近年来的文献发现，群智感知场景下的隐私保护研究多集中在任务分配环节，其次是激励机制，亦有针对质量约束和系统层面的研究，但尚未发现有文献对群体智能提取环节即群智计算环节的隐私保护进行研究。随着感知终端的大规模普及、感知数据的爆炸式增长，对所采集数据进行处理的压力必然会增大。同时，隐私泄露事件的不断发生和全球对数据隐私安全形成的保护趋势，更对传统的集中式数据处理模式提出了新的挑战，如我国的《民法典》[56]和欧盟《通用数据保护条例》。因此如何在保护数据隐私和满足监管要求的前提下进行数据处理，即针对智能提取环节的研究将成为热点。

结束语 本文回顾了近年来研究者对群智感知服务设计的隐私保护机制，从群智感知服务全生命周期的视角对隐私保护进行了定义，并对其衡量指标进行了讨论。由于单一的隐私保护手段无法满足群智感知复杂场景的需求，研究者通常使用多种隐私保护技术相结合的方式进行机制设计。文中从群智感知服务全生命周期的角度出发，对文献侧重的不同阶段所采用的隐私保护机制进行分析，最终得出隐私保护、资源消耗和数据质量损失这 3 个关键环节之间的对立统一关系：一方面是隐私保护和资源消耗之间的正相关关系，另一方面是隐私保护和数据质量的负相关关系，从而更好地理解研究者在设计任务分配、质量评估、激励机制等各阶段的隐私保护机制时所做出的妥协与平衡，并指导未来的工作。最后，汇总了现有文献使用的测试数据集，并结合群智感知应用的发展需求对未来研究面临的挑战展开了讨论。

参 考 文 献

[1] GOASDUFF L, PETTEY C. Gartner Says Worldwide Smartphone Sales to Grow 11% in 2021(WSCI)[OL]. https://www.gartner.com/en/newsroom/press-releases/2021-02-03-gartner-says-worldwide-smartphone-sales-to-grow-11-percent-in-2021.

[2] RIMOL M. Gartner Forecasts Global Spending on Wearable Devices to Total $81.5 Billion in 2021(WSCI) [OL]. https://www.gartner.com/en/newsroom/press-releases/2021-01-11-gartner-forecasts-global-spending-on-wearable-devices-to-total-

81-5-billion-in-2021.

[3] GANTI R K, FAN Y, LEI H. Mobile crowdsensing: current state and future challenges[J]. IEEE Communications Magazine, 2011, 49(11): 32-39.

[4] KHAN W Z, XIANG Y, AALSALEM M Y, et al. Mobile Phone Sensing Systems: A Survey[J]. IEEE Communications Surveys & Tutorials, 2013, 15(1): 402-427.

[5] LIU L, WEI W, ZHAO D, et al. Urban Resolution: New Metric for Measuring the Quality of Urban Sensing[J]. IEEE Transactions on Mobile Computing, 2015, 14(12): 2560-2575.

[6] WU Y, WANG Y, HU W, et al. SmartPhoto: A Resource-Aware Crowdsourcing Approach for Image Sensing with Smartphones [J]. IEEE Transactions on Mobile Computing, 2016, 15(5): 1249-1263.

[7] SAREMI F, FATEMIEH O, AHMADI H, et al. Experiences with GreenGPS—Fuel-Efficient Navigation Using Participatory Sensing[J]. IEEE Transactions on Mobile Computing, 2016, 15(3): 672-689.

[8] PRYSS R, REICHERT M, HERRMANN J, et al. Mobile Crowd Sensing in Clinical and Psychological Trials-A Case Study[C]// 2015 IEEE 28th International Symposium on Computer-Based Medical Systems. IEEE Press, 2015: 23-24.

[9] REDDY S, PARKER A, HYMAN J, et al. Image browsing, processing, and clustering for participatory sensing: lessons from a DietSense prototype[C]// Workshop on Embedded Networked Sensors. ACM, 2007.

[10] LIU L, LIU W, ZHENG Y, et al. Third-Eye: A mobilephone-enabled crowdsensing system for air quality monitoring[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(1): 1-26.

[11] KIM S, ROBSON C, ZIMMERMAN T, et al. Creek Watch: Pairing usefulness and usability for successful citizen science[C]// Proceedings of the ACM Conference on Human Factors in Computing Systems. Canada: ACM, 2011: 2125-2134.

[12] BONINO D, DELGADO M T, PASTRONE C, et al. WasteApp: Smarter Waste Recycling for Smart Citizens[C]// 2016 International Multidisciplinary Conference on Computer and Energy Science (SpliTech). IEEE, 2016: 1-6.

[13] MATARAZZO T J, SANTI P, PAKZAD S N, et al. Crowdsensing Framework for Monitoring Bridge Vibrations Using Moving Smartphones[J]. Proceedings of the IEEE, 2018, 106(4): 577-593.

[14] LI Q, CAO G. Efficient and Privacy-Aware Data Aggregation in Mobile Sensing[J]. IEEE Transactions on Dependable and Secure Computing, 2014, 11(2): 115-129.

[15] WANG E, YANG Y J, WU J, et al. An Efficient Prediction-Based User Recruitment for Mobile Crowdsensing[J]. IEEE Transactions on Mobile Computing, 2018, 17(1): 16-28.

[16] YIN B, LU J, WEI X T. Correlation-Based Task Processing Plans in Crowdsensing Platforms[J]. IEEE Transactions on Network Science and Engineering, 2021, 8(2): 1542-1556.

[17] GONG X, SHROFF N B. Truthful Mobile Crowdsensing for Strategic Users With Private Data Quality[J]. IEEE/ACM Transactions on Networking, 2019, 27(5): 1959-1972.

[18] AN J, LIANG D, GUI X, et al. Crowdsensing Quality Control and Grading Evaluation Based on a Two-Consensus Blockchain [J]. IEEE Internet of Things Journal, 2019, 6(3): 4711-4718.

[19] YANG D, XUE G, FANG X, et al. Incentive Mechanisms for Crowdsensing: Crowdsourcing With Smartphones[J]. IEEE/ACM Transactions on Networking, 2016, 24(3): 1732-1744.

[20] XIONG J B, CHEN X H, YANG Q, et al. A Task-Oriented User Selection Incentive Mechanism in Edge-Aided Mobile Crowdsensing[J]. IEEE Transactions on Network Science and Engineering, 2019, 7(4): 2347-2360.

[21] VERGARA-LAURENS I, JAIMES L, LABRADOR M. Privacy-Preserving Mechanisms for Crowdsensing: Survey and Research Challenges[J]. IEEE Internet of Things Journal, 2017, 4(4): 855-869.

[22] KHAN F, REHMAN A U, ZHENG J, et al. Mobile crowdsensing: A survey on privacy-preservation, task management, assignment models, and incentives mechanisms[J]. Future Generation Computer Systems, 2019, 100(11): 456-472.

[23] GUO B, YU Z, ZHANG D, et al. From Participatory Sensing to Mobile Crowd Sensing[C]// 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (Percom Workshops). 2014: 593-598.

[24] DING S, HE X, WANG J. Multiobjective Optimization Model for Service Node Selection Based on a Tradeoff Between Quality of Service and Resource Consumption in Mobile Crowd Sensing [J]. IEEE Internet of Things Journal, 2017, 4(1): 258-268.

[25] HUI Z, BOLOT J. Anonymization of location data does not work: A large-scale measurement study[C]// Proceedings of the 17th Annual International Conference on Mobile Computing and Networking. 2011.

[26] SHEN H, BAI G, HU Y, et al. P2TA: Privacy-Preserving Task Allocation for Edge Computing Enhanced Mobile Crowdsensing [J]. Journal of Systems Architecture, 2019, 97: 130-141.

[27] WANG Z, HU J, LV R, et al. Personalized privacy-preserving task allocation for mobile crowdsensing[J]. IEEE Transactions on Mobile Computing, 2018, 18(6): 1330-1341.

[28] XIAO M, JIE W, SHENG Z, et al. Secret-sharing-based secure user recruitment protocol for mobile crowdsensing[C]// IEEE INFOCOM 2017－IEEE Conference on Computer Communications. 2017: 1-9.

[29] MA L, LIU X, PEI Q, et al. Privacy-preserving reputation management for edge computing enhanced mobile crowdsensing[J]. IEEE Transactions on Services Computing, 2018, 12(5): 786-799.

[30] YANG L, ZHANG M, HE S, et al. Crowd-empowered privacy-preserving data aggregation for mobile crowdsensing[C]// Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing. 2018: 151-160.

[31] XIONG J, MA R, CHEN L, et al. A personalized privacy protection framework for mobile crowdsensing in IIoT[J]. IEEE Transactions on Industrial Informatics, 2019, 16(6): 4231-4241.

[32] MIAO C, JIANG W, SU L, et al. Privacy-preserving truth discovery in crowd sensing systems[J]. ACM Transactions on Sensor Networks (TOSN), 2019, 15(1): 1-32.

[33] SUN G, SUN S, YU H, et al. Toward Incentivizing Fog-Based Privacy-Preserving Mobile Crowdsensing in the Internet of Vehicles[J]. IEEE Internet of Things Journal, 2019, 7(5): 4128-4142.

[34] LIANG Y, AN J, HU X Z, et al. Dynamic incentive mechanism supported privacy-preserving in mobile CrowdSensing [J]. Com-

puter Application Research, 2019, 36(11): 3404-3409.
[35] WU H, WANG L, XUE G, et al. Enabling data trustworthiness and user privacy in mobile crowdsensing[J]. IEEE/ACM Transactions on Networking, 2019, 27(6): 2294-2307.
[36] WANG X, LIU Z, TIAN X, et al. Incentivizing crowdsensing withlocation-privacy preserving[J]. IEEE Transactions on Wireless Communications, 2017, 16(10): 6940-6952.
[37] BASUDAN S, LIN X, SANKARANARAYANAN K. A privacy-preserving vehicular crowdsensing-based road surface condition monitoring system using fog computing[J]. IEEE Internet of Things Journal, 2017, 4(3): 772-782.
[38] SUCASAS V, MANTAS G, BASTOS J, et al. A signature scheme with unlinkable-yet-accountable pseudonymity for privacy-preserving crowdsensing[J]. IEEE Transactions on Mobile Computing, 2019, 19(4): 752-768.
[39] XIONG J, ZHAO M, BHUIYAN M, et al. An AI-enabled Three-party Game Framework for Guaranteed Data Privacy in Mobile Edge Crowdsensing of IoT[J]. IEEE Transactions on Industrial Informatics, 2021, 17(2): 922-933.
[40] YANG M, ZHU T, LIANG K, et al. A blockchain-based location privacy-preserving crowdsensing system[J]. Future Generation Computer Systems, 2019, 94: 408-418.
[41] NI J, ZHANG K, YU Y, et al. Providing task allocation and secure deduplication for mobile crowdsensing via fog computing [J]. IEEE Transactions on Dependable and Secure Computing, 2018, 17(3): 581-594.
[42] WU D, YANG Z, YANG B, et al. From Centralized Management to Edge Collaboration: A Privacy-Preserving Task Assignment Framework for Mobile Crowd Sensing[J]. IEEE Internet of Things Journal, 2021, 8(6): 4579-4589.
[43] ZHENG Y, DUAN H, WANG C. Learning the Truth Privately and Confidently: Encrypted Confidence-Aware Truth Discovery in Mobile Crowdsensing[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(10): 2475-2489.
[44] WANG J, LI M, HE Y, et al. A blockchain based privacy-preserving incentive mechanism in crowdsensing applications[J]. IEEE Access, 2018, 6: 17545-17556.
[45] VERGARA-LAURENS I, MENDEZ D, LABRADOR M. Privacy, quality of information, and energy consumption in Participatory Sensing systems[C]// 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom). IEEE, 2014: 199-207.
[46] AN J, YANG H, GUI X, et al. TCNS: node selection with privacy protection in crowdsensing based on twice consensuses of blockchain[J]. IEEE Transactions on Network and Service Management, 2019, 16(3): 1255-1267.
[47] NI J, ZHANG K, XIA Q, et al. Enabling strong privacy preservation and accurate task allocation for mobile crowdsensing[J]. IEEE Transactions on Mobile Computing, 2019, 19(6): 1317-1331.
[48] WU H, WANG L, XUE G. Privacy-aware task allocation and data aggregation in fog-assisted spatial crowdsourcing[J]. IEEE Transactions on Network Science and Engineering, 2019, 7(1): 589-602.
[49] MAN H A, SUSILO W, YI M. Constant-Size Dynamic k-TAA [C]// 5th International Conference on Security and Cryptography for Networks, Lecture Notes in Computer Science. Maiori, 2006: 111-125.
[50] YANG P, GUI X, JIAN A, et al. An Efficient Secret Key Homomorphic Encryption Used in Image Processing Service[J]. Security & Communication Networks, 2017, 2017: 1-11.
[51] JIANG J, ZHENG Y, SHI Z, et al. A Practical System for Privacy-Aware Targeted Mobile Advertising Services [J]. IEEE Transactions on Services Computing, 2020, 13(3): 410-424.
[52] LIN J, YANG D, LI M, et al. Frameworks for Privacy-Preserving Mobile Crowdsensing Incentive Mechanisms [J]. IEEE Transactions on Mobile Computing, 2018, 17(8): 1851-1864.
[53] SWEENEY L. k-anonymity: A model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.
[54] DENT A W, ZHENG Y, YUNG M. Practical Signcryption [M]. Berlin: Springer, 2010.
[55] ZMAB C, JMA C, YMAB C, et al. Privacy-preserving and high-accurate outsourced disease predictor on random forest[J]. Information Sciences, 2019, 496: 225-241.
[56] civil code. civil code(WSCI) [OL]. http://www.npc.gov.cn/npc/c35174/mfdgfbca.shtml.
[57] LI T, JUNG T, QIU Z, et al. Scalable Privacy-Preserving Participant Selection for Mobile Crowdsensing Systems: Participant Grouping and Secure Group Bidding[J]. IEEE Transactions on Network Science & Engineering, 2020, 7(2): 855-868.
[58] ZHENG Y, DUAN H, WANG C. Learning the truth privately and confidently: Encrypted confidence—aware truth discovery in mobile crowdsensing[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(10): 2475-2489.
[59] QIU F, WU F, CHEN G. Privacy and Quality Preserving Multimedia Data Aggregation for Participatory Sensing Systems[J]. IEEE Transactions on Mobile Computing, 2015, 14(6): 1287-1300.
[60] ZHAO C, YANG S, MCCANN J A. On the Data Quality in Privacy-Preserving Mobile Crowdsensing Systems with Untruthful Reporting[J]. IEEE Transactions on Mobile Computing, 2021, 20(2): 647-661.
[61] SONG Z, LI Z, CHEN X. Local Differential Privacy Preserving Mechanism for Multi-attribute Data in Mobile Crowdsensing with Edge Computing[C]// 2019 IEEE International Conference on Smart Internet of Things (SmartIoT). IEEE, 2019: 283-290.
[62] CHEN JW, MA H D, ZHAO D, et al. Correlated Differential Privacy Protection for Mobile Crowdsensing[J]. IEEE Transactions on Big Data, 2021, 7(4): 784-795.

LI Li, born in 1977, Ph.D candidate, is a student member of China Computer Federation. Her main research interests include crowdcomputing, crowdsensing, privacy-preserving and machine learning.

HE Xin, born in 1974, professor, Ph.D supervisor, is a senior member of China Computer Federation. His main research interests include crowdsensing, mobilecomputing, cloudcomputing and big data processing.

(责任编辑:柯颖)