参与式感知隐私保护技术

曾菊儒 陈 红 彭 辉 吴 垚 李翠平 王 珊

(中国人民大学数据工程与知识工程教育部重点实验室 北京 100872)

(中国人民大学信息学院 北京 100872)

摘 要 移动智能终端正在成为人们日常生活的核心通信设备,参与式感知的隐私保护技术已成为研究热点.研究和解决数据以及位置隐私保护问题对参与式感知的大规模安全使用具有重要意义,然而参与式感知的特征使得隐私保护技术面临诸多挑战.该文对参与式感知隐私保护现有的研究成果进行了综述,首先介绍了参与式感知的基本应用和攻击模型,然后按照基于分组统计、第三方验证、K-匿名、数字加密 4 种策略对现有成果进行了分类,阐述了代表性的隐私保护技术,接着分析和比较了各技术的性能并总结了各技术的主要优缺点,最后提出了未来的研究方向.

关键词 移动参与式感知;隐私保护;分组统计;第三方验证;K-匿名;数字加密

中图法分类号 TP311 **DOI 号** 10.11897/SP.J.1016.2016.00595

Privacy Preservation in Mobile Participatory Sensing

ZENG Ju-Ru CHEN Hong PENG Hui WU Yao LI Cui-Ping WANG Shan

(*Key Laboratory of Data Engineering and Knowledge* (*Renmin University of China*) *of Ministry of Education*, *Beijing* 100872)

(*School of Information*, *Renmin University of China*, *Beijing* 100872)

Abstract Participatory sensing (PS) is an emerging area of interest for researchers as mobile intelligent terminals are becoming the core communication device in people's everyday lives. Privacy preservation techniques in PS have attracted more and more attentions. Researching and solving the problems of data privacy preservation along with location privacy preservation is essential to widespread employment of PS. However, inherent characteristics of PS make privacy preservation face series of challenging problems. This paper surveys the state-of-the-art privacy preservation techniques in PS. First, this paper reviews the basic applications and attack models. Second, existing works are classified into four categories, including packet statistics, third-party verification, K-anonymous and digital encryption. Then this paper describes the key techniques of privacy preservation in PS, analyzes and compares the performance of these techniques, and summarizes the main advantages and disadvantages of these techniques. Finally, suggestions for future research works are put forward. Packet statistics are used when participants upload data. First, the data are cut into a number of packets or chosen from n packets data. Then the data packets are distributed to neighboring nodes. After that the sink or the qureier completes information reorganization and integrity verification. Packet statistics are always used along with

收稿日期:2015-03-12;在线出版日期:2015-09-22.本课题得到国家"九七三"重点基础研究发展规划项目基金(2012CB316205)、国家"八六三"高技术研究发展计划项目基金(2014AA015204)和国家自然科学基金(61070056,61033010,61272137,61202114)资助.**曾菊儒**,男,1987 年生,博士研究生,主要研究方向为无线传感器网络、隐私保护、参与式感知.E-mail:zeng.juru@ruc.edu.cn.**陈 红**(通信作者),女,1965 年生,博士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为数据库、数据仓库、无线传感器网络.E-mail:chong@ruc.edu.cn.**彭 辉**,男,1986 年生,博士研究生,主要研究方向为无线传感器网络、隐私保护.**吴 垚**,男,1990 年生,博士研究生,主要研究方向为无线传感器网络、隐私保护、群智感知.**李翠平**,女,1971 年生,博士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为数据仓库和数据挖掘、信息网络分析、流数据管理.**王 珊**,女,1944 年生,硕士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为高性能数据库新技术、数据仓库与商务智能.

random walker, delayed transmitting, hop-by-hop encryption and data perturbation. In order to protect the participant privacy in the case of an untrusted server, one of the present methods is to add Trusted Third-Party (TTP) between participants and server. Participants upload data or the server assigns tasks all needs to go through TTP verification. Because the server cannot contact with participants' sensitive information directly, so as to effectively prevent the correlation attack and other attacks with background knowledge. Special network coding techniques are often needed when a TTP is used for verification. *K*-anonymity strategies allow participants to hide themselves in *k* users. The server can divide anonymous region according to participants' location information or privacy strength requirements. There are many useful techniques such as camouflaged data, data perturbation, false source and perturbation histogram. Meanwhile, *L*-diversity refers to the region in the same anonymous at least exists L attribute values which cannot be distinguished from each other. *L*-diversity can effectively enhance anonymity. Digital encryption techniques are used to prevent attackers access to the participant's private data directly, followed by preventing untrusted server to get too much sensitive information. The main difficulty in traditional encryption technology is how to calculate the encrypted data directly, verify the data integrity and reduce computational overhead. This review does the basic introduction for complex encryption process and other studies. Overall, packet statistics likely to cause a large amount of communication and network latency, therefore, we need to do trade-off between privacy strength and quality of service. Although based on third-party verification removes the direct contact between the server and participants, it requires more processing time and results in a decline in the quality of service. *K*-anonymity's challenge is how to efficiently build dynamic anonymous region; Digital encryption has to take more consideration on calculation amount and implementation complexity.

Keywords mobile participatory sensing; privacy preservation; packet statistics; third-party verification; *K*-anonymous; digital encryption

1 引 言

在开始正文之前,我们首先对参与式感知(Participatory Sensing,PS)进行明确定义:所谓参与式感知,是指通过日常移动设备形成移动互联网络,其数据由公众和专业用户感知、收集、分析或者筛选,然后上传的参与式传感器网络.它十分注重"人"的参与,其中,智能手机是参与式感知应用的主要终端.今天的高端手机配备了一些专门的传感器,包括环境光传感器、加速计、数字罗盘、陀螺仪和全球定位系统 GPS 等,还有一些通用传感器如麦克风和照相机.

参与式感知在健康监测、环境监测、交通监测、行为监测、社交和商业等领域具有广阔的发展空间,然而,这种新的数据采集方式也带来新的隐私问题.在健康监测领域,Zhang 等人[1]安装可以检测参与者所处环境信息(如体温、位置)以及是否患传染病等身体参数的移动传感器,间断性获取参与者的信息,从而推断出参与者患传染病的概率,以便及时控制流行疾病的传播范围.但是,这些敏感信息可能被攻击者截获,后者利用所获信息做一些谋取利益的垃圾广告.在环境监测领域,PEIR[2]提出了个性化环境影响报告书,通过运用手机采样来计算环境对参与者日常生活的影响,为人们的出行提供参考路线,然而,攻击者一旦获得这些数据便可推测出参与者的出行轨迹.在商业领域,MobiShop[3]和 PetrolWatch[4]分别利用参与式感知实时收集超市商品和石油价格,从而帮助人们买到更加实惠的商品.由于其收集的感知数据往往与商家的利益直接相关,攻击者很可能获取并篡改这些数据,进而误导参与者大量购买.当隐私受到威胁时,参与者会失去参与的积极主动性,影响参与式感知的应用.因此,如何保护每个参与者的隐私数据且不影响服务[5-6]的质量,是一个极其重要和具有挑战性的问题.

参与式感知隐私保护不同于移动社交网络隐私

保护[7-10]和传统的传感器网络隐私保护[11-16]，其主要面临如下挑战：(1) 数据稳定性差. 由于感知的主体是人，而人具有高度的移动性和灵活性，隐私保护协议必须应对这些动态的变化；(2) 数据维度大. 传统的感知数据类型往往是事先定义好的. 而现在，手机嵌入多少种传感器，参与者就可以获得多少种数据，进而能够上传多少种数据. 同时人们还可以对获取的数据进行相应的预处理，贴上不同的语义标签. 根据网络特征和形态的不同，往往存在不同的隐私保护需求和攻击方式，需要设计出有针对性的隐私保护协议；(3) 难以监督. 参与式感知中参与者经常会出现在无人监督的环境中，安全的监督管理难以进行，攻击者既可以根据先验信息和窃听所得数据进行个体攻击，又可以较便捷地与其他参与者进行共谋攻击，从而窃取参与者的隐私信息.

本文综述了参与式感知隐私保护技术，列举了典型应用和攻击方式，并对该领域的研究成果进行系统地梳理和总结，按照所采用的技术将现有研究成果分为分组统计、第三方验证、K-匿名、数字加密 4 类，阐述了代表性协议的核心技术，分析和对比了各个协议的性能以及主要优缺点，给出了各个协议所能防御的攻击类型，最后，展望了未来的研究方向.

本文第 2 节介绍研究模型，包括攻击模型、网络模型和信任模型；第 3 节分别对基于分组统计、第三方验证、K-匿名、数字加密的隐私保护技术进行阐述；第 4 节从实现难易程度、网络通信质量、保护对象、隐私保护强度和能耗等角度全面分析和对比代表性协议的性能；第 5 节提出未来的研究方向.

2　研究模型

2.1　攻击模型

根据参与式感知中攻击者的目标不同，主要分为针对参与者隐私的诚实但好奇模型[17]和针对数据完整性的恶意攻击模型. 诚实但好奇模型：攻击者通过被俘获的节点信息推测隐私数据，从而窃隐私取信息，但其遵守协议而不篡改信息，仅破坏数据的隐私性. 诚实但好奇模型是目前参与式感知数据隐私保护中主要关注的对象. 恶意攻击模型：攻击者不但窃取用户的隐私数据，而且篡改信息，破坏数据的隐私性和完整性.

针对参与者隐私的攻击模型，主要攻击手段有：(1) 特定对象攻击. 攻击者为了锁定攻击目标，尝试向服务器请求一个只有少数参与者才会接受的任务. 例如，攻击者可以请求一个针对使用 iPhone 的参与者并且携带心率传感器的任务，他所获得的返回结果中会有很高的概率包含攻击目标的具体位置；(2) 选择性攻击. 与特定任务攻击的不同之处是，选择性攻击控制多个相关任务分配而不是请求单个任务，攻击者利用多个请求之间的相关性，将任务分配给一个或有限几个参与者，使得该任务在几个参与者之间彼此相连. 例如，如果已知只有一个参与者下载了某一个任务，那么与该任务相同的请求容易链接到同一个参与者；(3) 时间关联分析. 攻击者通过分析来自同一参与者在一段时间内的一个或多个提交信息，获取参与者的活动轨迹. 最坏的情况下，一个提交信息就能泄露参与者的隐私，例如，已知一个 IP 地址归属于某个人的房屋；(4) 共谋攻击. 主要有本地攻击者和不可信服务器共谋，本地攻击者提供时间和位置信息，不可信服务器提供参与者的通信内容. 为达到攻击目的，攻击者会提前获取参与者在众多通信会话中所处位置；(5) 背景知识攻击. 攻击者利用已获取的关于被攻击者的信息来窃取隐私数据，比如攻击者知道在某个时间段被攻击者出现在某个匿名区域，即使被攻击者以匿名的方式发布消息，但攻击者仍能通过已获取的时间信息匹配出相应的结果，从而窃取被攻击者的隐私；(6) 同质性攻击(homogeneity attack). 常出现在参与者已经实现 K-匿名后，由于 k 个匿名者具有相同的特征，比如都是某个癌症的患者，从而隐私遭到泄露；(7) 逐跳回溯攻击. 攻击者通过反向逐跳追踪信号来源来确定参与者的位置，逐跳回溯攻击最常用到的方法是三角定位法. 具体过程为：攻击者确定信号发送点位置后，立刻转移到该点，并继续对信号进行监听，重复上述过程，攻击者就可以找到参与者的位置；(8) 主成分分析攻击(Principal Component Analysis，PCA). 尽管参与者的隐私数据已经经过扰乱技术处理，但攻击者能通过求平均或剔除噪音函数等方式对真实数据进行拟合，从而获取接近真实数据的值.

针对数据隐私性和完整性的攻击模型，主要攻击手段有：(1) 数据篡改. 攻击者通过对捕获的数据进行恶意修改或者直接删除，使感知服务器得到不完整的查询或聚焦结果；(2) 数据重发. 攻击者重复转发捕获的节点数据，使网络通信量大大增加，影响网络的正常运转，同时打乱了节点转发的时序，加大感知服务器对数据完整性验证的难度；(3) 数据伪造. 攻击者通过向捕获的数据中注入虚假信息，既

可以伪装自己的真实身份，又可以降低感知数据的质量.

以上攻击手段可以通过俘获节点、伪装成正常用户或监听网络信号等方式实现(如表 1)，其中，特定对象攻击、选择性攻击和共谋攻击可采用伪装方式；时间关联攻击、背景知识攻击、同质性攻击和逐跳回溯攻击可采用监听网络信号方式；主成分分析、数据篡改、数据重发和数据伪造可采用俘获节点方式. 采用伪装或监听方式的攻击模型是参与式感知隐私保护研究的重点. 而数据篡改、数据重发和数据伪造尽管是比较严重的攻击方式，但在参与式感知隐私保护中关于这部分内容的讨论比较少. 因此，如没有特殊说明，本文所述的攻击模型都采用诚实但好奇模型. 而且，在参与式感知隐私保护中，攻击选择对象和攻击选择时刻是有所不同的. 请求方/参与者在数据上传/数据下达时的位置或数据隐私保护都可以作为独立的研究点. 为避免篇幅冗长，本文着重介绍各个协议主要解决的隐私保护问题，比如参与者上传数据时的位置隐私保护问题. 本文默认攻击时刻是数据上传.

表 1 攻击手段对比

攻击手段	所属攻击模型及主要破坏类型	攻击选择对象及攻击选择时刻	主要特点	采用攻击方式
特定任务攻击	诚实但好奇模型/数据隐私性	参与者/数据上传	攻击速度快，操作简单	伪装
选择性攻击		参与者/数据上传	攻击速度较快，操作简单	伪装
时间关联分析		请求者、参与者/数据上传	出现频率较高，需要物理硬件支持	监听
共谋攻击		请求者、参与者/数据上传、下达	出现频率高，防御需考虑多方因素	伪装
背景知识攻击		请求者、参与者/数据上传	出现频率高，需要物理硬件支持	监听
同质性攻击		请求者、参与者/数据上传	出现频率较高，需要物理硬件支持	监听
逐跳回溯攻击		请求者、参与者/数据上传、下达	出现频率较低，需要物理硬件支持	监听
主成分分析攻击		参与者/数据上传	实现复杂，需要物理硬件支持	俘获节点
数据篡改	恶意攻击模型/数据隐私性和完整性	请求者、参与者/数据上传	出现频率较低，需要物理硬件支持	俘获节点
数据重发		请求者、参与者/数据上传	出现频率较低，对网络通信影响大	俘获节点
数据伪造		请求者、参与者/数据上传	出现频率较低，需要物理硬件支持	俘获节点

2.2 网络模型

典型的参与式感知网络体系结构不同于单层的传感器网络[18]和两层传感器网络[19]，它通常由 5 部分组成(如图 1)，包括参与者(Workers 或 Mobile Nodes，MNs)、请求者(Requesters 或 App)、注册验证者(Registration Authority，RA)，服务提供者(Service Provider，SP)和外围组件(如网络接入点、数据处理器). 上述前 4 部分通过外围组件进行连接和通信.

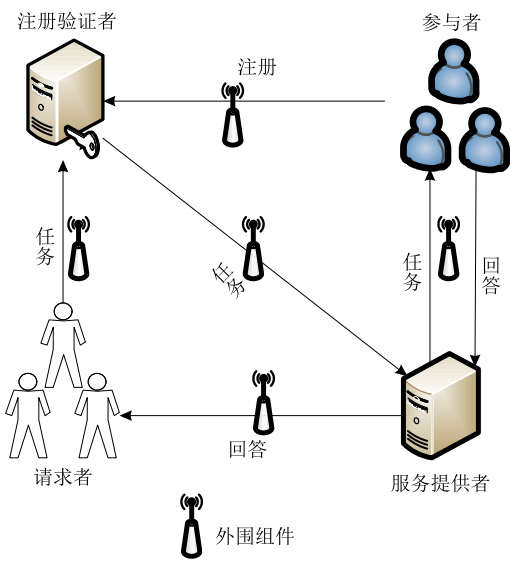


图 1 典型的参与式感知网络体系结构

参与者往往拥有能够感知、计算、存储以及在不同平台有无线通信能力的设备，诸如智能电话、个人数字助手 PDA、笔记本电脑. 他们既可以访问个人传感器(如 iPhone 上的加速器)上的数据，同时又具备通过一些开放式网络接口接入互联网的功能. 当某个移动网络用户想加入参与式感知系统时，他首先需要向 RA 注册自己的移动设备. 待验证成功后，RA 为参与者设备装入相关认证信息. 任务请求及下达阶段，请求者首先将任务提交给 RA，验证通过后，RA 将任务转发送给 SP. 任务获取及上传阶段，参与者根据自身条件和任务要求从 SP 处选择性下载任务. 一旦接受任务，参与者根据任务描述收集数据，既可以把感知数据返回给 SP，也可以把结果直接返回给请求者，而第一种返回方式是现阶段参与式感知网络中隐私保护研究的重点.

参与式感知经常和机会式感知(Opportunity Sensing)一起被讨论和研究. 由于两者都包含参与者的主观意愿(上报自带的或附加的传感器设备的参数)，只是在监督力度上有所不同，前者结合参与者的全程监督，后者可以在参与者的间隔性甚至无

监督的情况下完成. 而在隐私保护层面上，二者基本相同，因此本文不做区别处理.

文献[20]从参与者的角度按照任务分派和提交方式分为主动和被动两种参与模式. 主动参与模式是指服务提供者将参与式感知任务上传到网络，参与者根据任务要求和自身条件选取任务，任务范围不受服务提供者影响；被动参与模式要求参与者事先将自己当前的敏感信息(比如所处位置)告知服务提供者，后者将任务推送给邻近节点. 被动参与模式由于提前将信息上传，数据整合和可控性都比较好，但提前上传的隐私数据给攻击者留下大量的背景知识和攻击时间，因此如何保护此时的隐私是现阶段研究的热点.

2.3　信任模型

参与式感知网络中参与者和请求者都信任注册验证者能够正确执行它的认证过程. 请求者相信来自认证后的参与者的感知数据. 但外围组件有时可信性不高，它们很可能成为攻击者俘获的对象.

3　隐私保护技术分类

在参与式感知隐私保护中，K-匿名策略能够使得用户的位置或时空数据被识别出的概率不大于$1/k$，从而成为较早被使用的一类重要的保护策略. K-匿名策略往往和可信第三方同步使用. 可信第三方将用户的原始位置信息转换到另一个空间或模糊处理用户的确切位置点，将其置于隐匿区. 然而，可信第三方拥有太多用户的敏感信息而容易成为首要的攻击目标. 因此，为避免可信第三方可能出现的问题，研究者引入了分组统计和数字加密策略. 分组统计和数字加密可以在请求方进行分片组合或解密操作，从而避免可信第三方拥有过多的敏感数据. 但是后述两者增加了网络能耗和系统开销，需要均衡处理.

本节中，按照分组统计、第三方验证、K-匿名和数字加密 4 种策略，对现有的参与式感知网络隐私保护技术进行了分类. 如上所述，在具体实现时上述策略并不是完全对立的，它们往往被组合使用. 例如，分组统计结合数字加密一起使用，第三方验证与K-匿名以及数字加密组合运行. 因此，分类按照隐私保护协议侧重点来进行. 本节对各个代表协议的基本原理和突出特点进行了详略描述，并简要分析了各个协议的性能.

3.1　基于分组统计技术

分组统计技术是指在数据上传时，首先把感知数据切分成若干份或者从 n 份数据中选取 k 份，然后将分组后的数据分发给邻居节点，最后在汇聚节点或者最终用户处进行分片信息重组和完整性验证. 分组统计技术经常与随机转发、延迟转发、逐跳加密、数据扰乱等技术组合使用.

HP^3(Hot-Potato-Privacy-Protection Algorithm)[21]在服务器不可信的情况下保护参与者的数据隐私，主要防止服务器将参与者和感知数据相结合以及恶意参与者攻击. 主要思想为：用户将数据分成多片，然后分别转发给邻居节点，当跳数达到阈值时，各转发节点才将分片信息直接传送给服务器. 由于参与式感知分片数据的稳定性差，最后，服务器需做数据完整性验证. 算法相对简单，执行过程如下：

首先，假设初始发送节点有 k 个邻居节点，现从 k 中随机选取 i 个节点作为转发节点，同时选取一个随机数 $\rho(\rho\in(0,1)$，ρ 按照 $1/k$ 递减). 如果 ρ 的值小于等于跳数阈值 τ(the hopping threshold)，则将数据直接传送给服务器；否则减小 ρ. 继续从当前节点的邻居节点中随机选取若干节点作为转发节点，若中间节点只有一个邻居节点，表示其数据转入和将要转出的节点是同一个，则算法重新选择 ρ，重复执行上述步骤，直到 ρ 满足条件为止.

为了保护转发过程中的隐私信息，各个节点与服务器利用公开密钥对数据进行加密，同时邻居节点两两之间利用协商后的共享密钥对数据再次进行加密. 由图 2 可知服务器不参与转发，节点到服务器

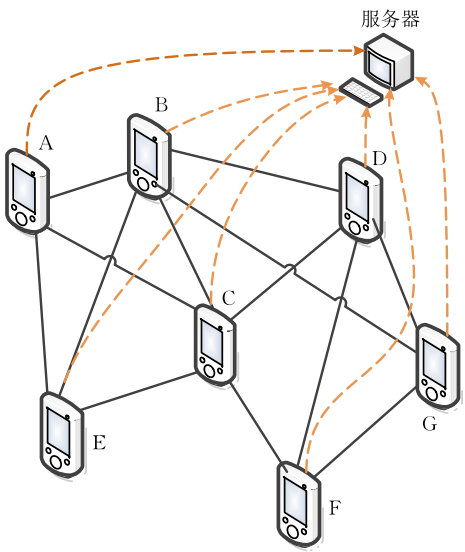


图 2　HP^3 网络模型示意图

具有多个路径且包括一条直达服务器的路径(图 2 中虚线部分).

恶意用户会对数据进行伪造、篡改或者丢弃.对于数据伪造和篡改问题,原有的公钥和私钥能起到很好的效果,而对于数据丢弃问题,则需要对 HP^3 进行扩展. HP^{3+}[21]在 HP^3 的基础上做了备份处理,具体方法如下:将数据分成多份,比如分成 n 份,其中 k 份是另外 $n-k$ 份的备份数据,然后分组转发.这样将降低因恶意用户丢掉转发包而导致数据不完整的概率.理论上,备份越多,保证数据完整性的概率越大,但系统开销也会增大,因此需要在能耗和隐私保护强度上做一个均衡处理.服务器无法获得感知数据的具体来源,从而保护参与者的数据隐私.

PESP(Privacy Enhanced State-dependent Perturbation)[22-23]对 HP^3 进行了改进,加入了数据扰乱技术,提高了隐私保护强度,进一步防御主成分分析攻击.首先,主成分分析攻击需要提前获得噪音参数以便准确估计出真实数据,但在 PESP 中,每个参与者的噪音分布参数是不同的,攻击者截获一个参与者的噪音参数并不影响其他参与者;其次,PESP 中加入的噪音和参与者的各时间段的状态值有关,但在时间上没有关联性;最后,PESP 选取的噪音参数能每隔一段时间更新一次,让攻击者难以区分,但数据分片和数据扰乱增加了系统的通信和计算代价.PESP 主要分为生成噪音和数据重构 2 个步骤.具体方法如下:

(1)对于参与者 i,为其设置一个常量 C^i 和一个服从期望为 μ_i,方差为 σ_i^2 的独立随机噪音变量 η^i,即 $\eta^i \sim N(\mu_i,\sigma_i^2)$;

(2)生成噪声序列 $n^i=S^i\eta^i+C^i$,其中 S^i 与状态值 x^i 紧密相关,它既可以等于 x^i,也可以是某一段时期内 x^i 的平均值,比如: $S^i=\sum_{j=1}^{t}x_j^i/t$. 为了简便,取 $S^i=x^i$;

(3)向原始的真实数据序列 x_k^i 加入第 2 步生成的噪音变量 n_k^i,得到扰乱后的数据 $y^i=x^i+S^i\eta^i+C^i=x^i(1+\eta^i)+C^i$;

(4)每隔一个时间段 T,更新 μ_i 和 σ_i 的值,并把新的参数值上传给服务器.从而防御主成分分析攻击.

M-PERM(Mutual Privacy Preserving Regression Modeling)[24]在服务器不信的情况下,不仅保护感知节点的数据隐私,而且提供相对准确的聚集结果. M-PERM 的核心思想是通过建立回归模型,利用泰勒公式和矩阵转换导出最佳的数据分解维度和参与分组的参与者数,并用统计学方法找到最优模型.最终用户对数据进行信息重组和完整性验证,这种方式对共谋攻击十分有效,但动态建立回归模型和矩阵转换增加了系统的计算复杂度. M-PERM 分为 3 个步骤:

(1)感知节点上的数据聚集.每个感知节点 $V^{(ij)}$ 首先聚集 w 个下列元组 $\{(x_{k,1}^{(ij)},x_{k,2}^{(ij)},\cdots,x_{k,w}^{(ij)},b_k^{(ij)})\mid k=1,2,\cdots,w\}$,然后计算其隐私数据聚集值 $\boldsymbol{\Theta}^{(ij)}$.

(2)簇头节点上的数据聚集.首先,簇头节点 $P^{(i)}$ 生成 c 个不同的正数 $\rho_1,\rho_2,\cdots,\rho_c$ 并分发给节点 $V^{(ij)}$,与此同时,节点 $V^{(ij)}$ 随机生成向量 $\boldsymbol{r}^{(ij)}=(\rho_1,\rho_2,\cdots,\rho_{c-w^2+3w+3})^{\mathrm{T}}$. 计算

$$s_{(k)}^{(ij)}=[1\ \rho_k\cdots\rho_k^c]\begin{bmatrix}\boldsymbol{\Theta}^{(ij)}\\ \boldsymbol{r}^{(ij)}\end{bmatrix}.$$

节点 $V^{(ij)}$ 保留 $s_{(j)}^{(ij)}$ 的值,同时将 $s_{(k)}^{(ij)}$ 发送给其他节点 $V^{(ik)}(k\neq j)$. 节点 $V^{(ij)}$ 计算来自同一簇中其他节点的数据,簇头节点数据聚集过程如图 3 所示.

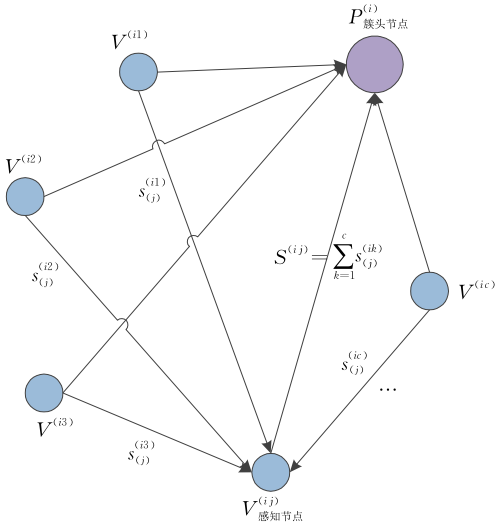


图 3 M-PERM 中簇头节点上的数据聚集过程演示

簇 i 中的聚集结果用 $\boldsymbol{\Phi}^{(i)}$ 表示,则

$$\boldsymbol{\Phi}^{(i)}=\begin{bmatrix}\sum_{j=1}^{c}\boldsymbol{\Theta}^{(ij)}\\ \sum_{j=1}^{c}\boldsymbol{r}^{(ij)}\end{bmatrix}.$$

(3)汇聚节点(最终用户)上的数据聚集:由于在 $\boldsymbol{\Phi}^{(i)}$ 中,上半部分的数据是真实的数据,下半部分是随机扰乱数据,因此在汇聚节点上计算最终的聚

焦结果 $\boldsymbol{\Theta}$.

$$\boldsymbol{\Theta}=\sum_{i=1}^{m}\sum_{j=1}^{c}\boldsymbol{\Theta}^{(ij)},$$

其中，m 为簇头节点个数，数据聚集过程如图 4 所示.

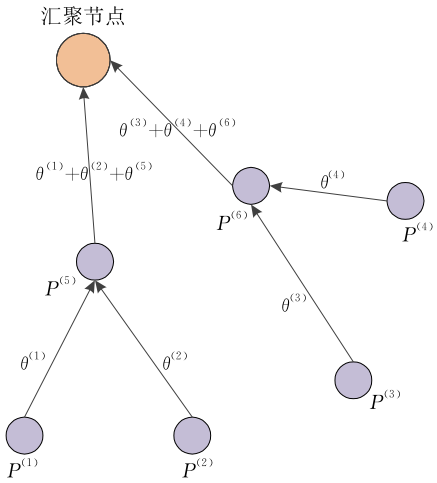


图 4　M-PERM 中汇聚节点上的数据聚集过程演示

聚集过程中，节点 $V^{(ij)}$ 可能会收到来自节点 $V^{(ik)}$ 的数据，但是因为 $f^{(ik)}(j)$ 是 c 个不同正数的累积和，节点 $V^{(ij)}$ 很难从 $f^{(ik)}(j)$ 中反向推导出 $\boldsymbol{\Phi}^{(ik)}$ 的值，从而保证感知节点之间的隐私数据是相互保密的.

PMG (Privacy-Preserving High-Quality Map Generation)[25] 能够在参与者参与构建地图并上传数据给服务器时，有效防止服务器或监听者获取参与者的轨迹信息. PMG 与 M-PERM 不同的是，它不需要添加扰乱数据，而且分组数据是从 n 份数据中选取 k 份即参与者可以从其位置轨迹中自主选择若干个位置信息，以乱序组合的方式上传给服务器. PMG 主要从两个方面保护用户隐私：一是去除上传地点和上传时间的关联性；二是在一段时间和区域内限制上传者的数量. 分组数据的随机选取支持用户不同的隐私保护要求，但是轨迹数据的动态传输增加了网络通信量. 其具体方法如下：当参与者需要报告自己的位置信息之前，先判断所在位置是否满足 γ-抽样，即判断 $\gamma\geq\gamma'$ 是否成立，其中 γ' 是参与者自定义的隐私保护强度. 曲线 F(图 5 中的方框)属于点集 S(图 5 最中间的椭圆)的 γ-抽样，需满足：对曲线 F 上任意一点 p，在点集 S 中存在一个离 p 最近的点 s，符合 $D(p,s)/LFS(p)\leq\gamma$，其中，$D(p,s)$ 表示点 p 和 s 的距离，$LFS(p)$(Local Feature Size)是曲线 F 上的点 p 到中轴线(图 5 中椭圆虚线)的最短欧式距离. 如果下一个位置是安全的，参与者才会选择上传感知数据，从而保护用户的轨迹安全.

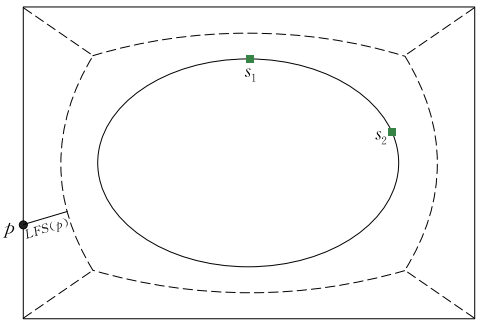


图 5　PMG 中的 $LFS(p)$示意图

PriSense[26] 首次对参与者的隐私保护强度和能耗进行均衡处理，防止在难以监督的环境下，服务器和被俘获参与者直接进行共谋攻击，其中，切分重组和数字加密可以有效应对共谋攻击. PriSense 主要包括切分重组和二分查找两个部分. 其中，切分重组与 M-PERM 有很大的不同. 首先，它将用户分为满足查询条件和不满足查询条件两种类型，满足查询条件的用户只会将分片转发给不满足查询条件的感知节点；其次，PriSense 中不需要提前分簇，节点和节点之间是动态变化的. 节点的选择有 3 种方式：随机选择、一跳选择和多跳选择. 一旦选择完成，节点之间会生成具有时效性的端到端的密钥，从而防止其他用户的共谋攻击. 同时，它能够支持各种非增量型数据聚集功能，如总和、平均值、方差、最大值/最小值等，并能得到准确度较高的聚焦结果.

PriSense 采用二分查找法，感知节点和感知节点之间由服从 802.11 协议的网络连接，即如果节点地址由 λ 位二进制组成，那么节点有 2^{λ} 个地址. 例如：求最大值时，首先将查询条件设为 $Q_1=2^w$(整数 w 满足 $0\leq w<\lambda$)，如果存在大于 Q_1 的值，那么接下来将查询条件设为 $Q_2=2^w+2^{w+1}$，否则，$Q_2=2^{w-1}$，然后继续查询比较. 二分查找法能够快速获取聚集值，但实时动态的节点链接增加了系统通信代价.

Wang 等人[27-28] 用于在攻击者已获取参与者的部分背景知识时保护参与者的轨迹隐私. 它采用马尔可夫模型来捕捉上下文感知数据之间的转换，并用两个状态的马尔可夫链模拟感知数据中人的行为和活动. 参与者和攻击者之间的竞争被当作零和随机游戏(zero-sum stochastic game)，参与者通过事先安装的隐私保护中间件控制上下文感知数据的粒度以保护隐私，数据粒度越小，隐私保护越弱，反之则相反；攻击者则试图通过选择感知数据源进行攻击. 参与者和攻击者之间的互动游戏被认为持续若干个上下文感知阶段. 历史数据和攻击结果被当作马尔可夫链中的系统状态，在系统状态上双方进行

信息统计并调整自己的策略,从而得到隐私保护状态.如果攻击者成功推断出参与者的位置是在其推断区域,则参与者的隐私遭到破坏.参与者的最佳防守策略在零和游戏中的纳什均衡(Nash equilibrium)点处获得,因此,可将隐私保护问题转化为既能够获得纳什均衡点又高效收敛的算法.感知节点因需要整合历史数据并预测位置坐标而带来新的计算代价和能耗,但此协议能有效地防止背景知识攻击.

3.2 基于第三方验证技术

参与式感知是一个面向服务的网络,其中,服务提供商是重要的组成部分.在传统的两层传感器网络中,存储节点被攻击的概率较大;而在参与式感知系统中,服务器被攻击的概率大大增加.因此,如何在服务器不可信的情况下对参与者的隐私进行保护,也是参与式感知中隐私保护的热点问题.现阶段常用到的方法是在参与者和服务器之间添加权威第三方,任务上传和下达都需要先经过可信第三方验证,称之为定向通信.由于服务器端不能直接接触到参与者的隐私信息,从而有效防止数据上传和下达时的时间关联攻击、背景知识攻击等攻击方式,为此,可信第三方进行验证时经常会用到特殊的网络编码技术.

PEPSI(Privacy-Enhanced Participatory Sensing Infrastructure)[29-30]主要用于保护参与者的数据隐私,参与者的查询请求和个人信息不被服务器同时获取.PEPSI 的主要思想是利用可信第三方为参与者提供匿名服务,参与者的数据传输由可信第三方负责.它主要包含移动节点(参与者)、注册认证(可信第三方)、服务提供商、网络运行商以及查询请求者(如图 6).

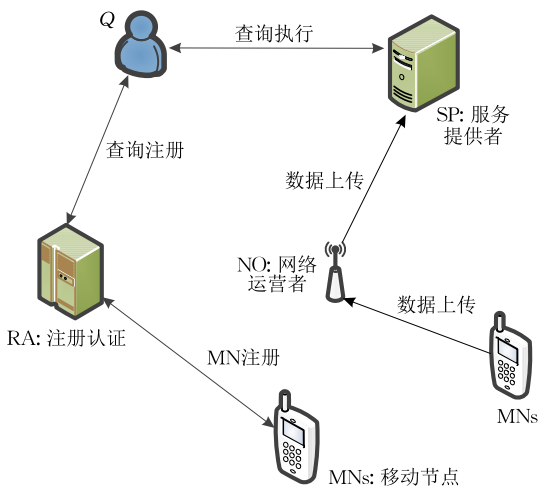


图 6 PEPSI 系统示意图

PEPSI 与 2.2 节网络模型中介绍的方法类似:参与者在可信第三方处注册自己的传感器装置并安装参与感知的软件,此时,参与者可以从请求数据类型列表中选择其数据请求类型.这个请求数据类型列表是公开的,既可以从服务提供商处获得又可以从可信第三方处获得.以查询请求者为例,它先向参与式感知应用中发出的查询请求需要先经过可信第三方验证.验证通过后,可信第三方将请求者的个人信息映射到新的用户空间,然后以新用户的身份向服务器请求数据.移动节点使用网络运行商提供的网络接口提交请求,理想情况下,这种操作数据不能服务提供商或者未经授权的查询请求者获得.隐私数据包括数据类型、参与者的位置信息或定量信息,例如每立方米 35 毫克碳氧化物.查询执行应在服务提供商不知情的情况下进行,服务提供商只是负责匹配查询条件和查询结果,从而防止服务提供商同时获取参与者的个人信息和感知数据.请求者/参与者可用通过少数几次链接而获得/上传相应的感知数据,这对资源有限的参与式感知节点非常重要,但 PEPSI 应对特定任务攻击和选择性攻击都相对较弱.

PEPPeR(Privacy Enhancing Protocol for PaRticipatory sensing)[31]用于保护请求者发出请求时的数据隐私,它和 PEPSI 的目的和解决思路相似.首先,请求者和参与者由可信第三方服务器分配令牌;然后,请求者的数据请求和参与者的数据上传都需要通过第三方验证.数据上传和下达都需要经过参与者、请求者和可信第三方服务器之间的 Tor(The onion router)网络:第二代洋葱路由(onion routing)的一种实现,用户通过 Tor 可以获得一个匿名,并在因特网上使用匿名交流.由于请求者的请求数据首先需要到 RA 处进行验证,然后 RA 将验证后的数据转发给 SP,从而形成了一个定向通信;同时,请求者的数据又经过了 Tor 网络,使得 SP 更加难以对请求者进行背景知识或时间关联攻击,但是定向通信和匿名网络增加了通信时延.

ARTSense(Anonymous Reputation and Trust in Participatory Sensing)[32]用于保护参与者的数据隐私.当参与者提供感知数据时,服务器在不知道参与者的个人信息情形下对感知数据做相似度计算.其主要思想与基于第三方验证策略类似,即防止服务器获得用户和感知数据之间的关联信息.具体方法如下:首先,参与者向服务器提供数据之前,先申

请一个任务 ID. 然后，参与者利用随机数匿名绑定任务 ID 和自己的特征数据，并结合感知数据组成上传内容传送给服务器. 接着，服务器对上传内容中的感知数据做相似度计算并给出反馈的信誉值. 计算完成后，服务器将加密后的信誉值替换感知数据并返回给用户. 最后，用户向服务器上传真实的特征信息以方便在信誉数据库中查找并更改相应的信誉值. 整个过程中，服务器不能同时获取用户的特征信息和感知信息，从而保护了用户的隐私. ARTSense 主要由声誉管理、感知数据信任评估和匿名管理组成，实现比较复杂. 它首次将隐私保护匿名机制和移动节点的信任度、声誉进行结合，并且充分考虑“人”的因素，进一步保护移动节点主动或被动泄露等隐私问题.

LotS(Lord of the Sense)[33-34] 与 ARTSense 一样将参与者的信任度融合到参与者数据隐私保护技术当中，它主要防止服务器同时获取参与者的个人信息和感知数据，并由 RA 负责转发数据. LotS 架构中主要包括 RA、MN、C(Community)和 ADM(Application Distributed Market). 其中，C 相当于服务器 SP，ADM 相当于物理地址转换器. MN 在 RA 处注册之前先从 ADM 处获得一个由加密数字串、时间戳和别名组成的证书. RA 解密证书得到数字串并通过哈希函数验证数字串是否有效，若有效，则给 MN 分配组签名. MN 向 C 请求上传数据时，C 首先验证 MN 的组签名，然后再判断 MN 的证书是否在有效期内，只有两者同时满足，MN 才能上传数据，这样能有效防止恶意用户攻击. MN 上传的数据会被打分，打分结果与 MN 的隐私保护强度直接关联. 打分越高，组越大，获得时间戳越长，隐私保护强度越高. LotS 不仅保护参与者的数据隐私，而且间接激励用户提供高质量的感知数据.

TrPF(Trajectory Privacy-Preserving Framework)[35-36] 提出参与式感知中针对参与者声明中需要保护的隐私数据进行操作的隐私保护框架. TrPF 与 PEPSI 一样，主要依靠可信第三方为用户提供匿名区域. 可信第三方在综合考虑参与者隐私强度要求和通信代价后，将应用信息发送给参与者. 应用信息包括证书和匿名信息. 证书用于验证参与者的有效性，从而排除恶意攻击者；匿名可以消除参与者的时空数据和其特征信息之间的直接联系，因为时空信息泄露会威胁到参与者的隐私. 为了减少系统的通信代价，TrPF 没有选择虚假位置和数据扰乱，而是建立轨迹混合区并在混合区模型的建立时添加时延，以保护参与者的运动轨迹. 同时，TrPF 只针对参与者声明需要保护的隐私数据进行操作，这不仅能够节省计算代价，而且能够为参与者提供不同的隐私保护强度.

AnonySense[37-38] 构建系统框架如图 7 所示，用于保护参与者上传或从服务器方获取感知数据时的数据隐私. 它是较早被提出的利用可信第三方验证的隐私保护协议. 其基本思想是利用可信第三方去除用户与服务器的之间关联. AnonySense 在系统中建立多个同构网络，它允许不同类型的感知节点参与到应用当中，有效解决参与式感知中数据维度大的问题.

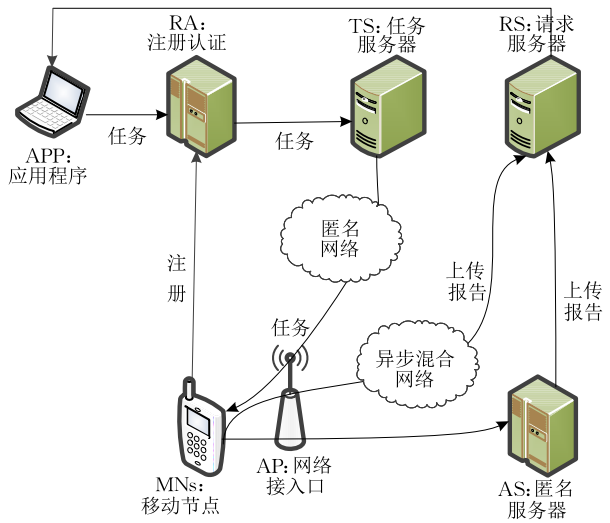


图 7　AnonySense 框架示意图

为了给应用提供高可信度的感知数据，AnonySense 增加了很多细节处理. (1) RA 负责验证提交或请求的信息；(2) 建立参与者之间的各种私钥，分组密钥、全局公钥；(3) 利用随机延迟和固定延迟成批发送数据，防止时间关联攻击；(4) 对 MAC 地址进行转换，地址转换不仅需要物理层硬件的支持，还需要对网络进行编码，从而防止 IP 地址攻击等. 图 7 中，匿名网络主要指的是 Tor 类型的网络，而异步混合网络是指同步和异步网络混合在一起. 数据传输路径的变化将有效地防止服务器同时获取用户的个人信息和感知数据.

3.3　基于 K 匿名技术

不管是通过参与式感知服务器给参与者分配 $k-1$ 个邻居节点，还是参与者自己与邻居节点协商，从而组成一个拥有 k 个用户的匿名区域，都被称之为 K-匿名策略. K-匿名策略让参与者隐藏在 k 个

用户中,使攻击者不能区分出参与者与其他 $k-1$ 个用户.服务器可以根据参与者的地理位置信息或者用户对隐私强度的要求来划分匿名空间,常用到的技术有伪装数据、数据扰乱、虚假源和扰动直方图等.同时,L-多样性能够加强匿名效果,L-多样性是指在同一匿名区域内,至少存在 L 种相互之间不能被区分的属性值.

To 等人[39]提出一种整合差分隐私(Differential Privacy,DP)[40]和地域群播的保护机制,主要保护参与者上传感知数据时的位置隐私.差分隐私通过结合拉普拉斯分布函数,在原始数据中添加随机噪音,既可以隐藏真实数据又能够相对准确地计算出文献[39]中所需 *count* 的聚集结果;地域群播通过有效控制和选取节点数来保证高效的服务,基本步骤如下:

(1)初始化阶段.参与者根据自身对隐私保护强度的要求设定差分隐私预算 ε 值.

(2)网格分割阶段.系统根据 ε 值和敏感度 Δf($\Delta f=1$,因为是计算 *count* 值),向查询结果中添加随机的拉普拉斯噪音,其标准差 $\lambda=\Delta f/\varepsilon$.Cormode 等人[41]首次提出了私人空间分解的概念(Private Spatial Decompositions,PSD),文献[42]表明两层自适应网格(adaptive grid)(如图 8)往往比执行递归分割的树形结构更有优势.

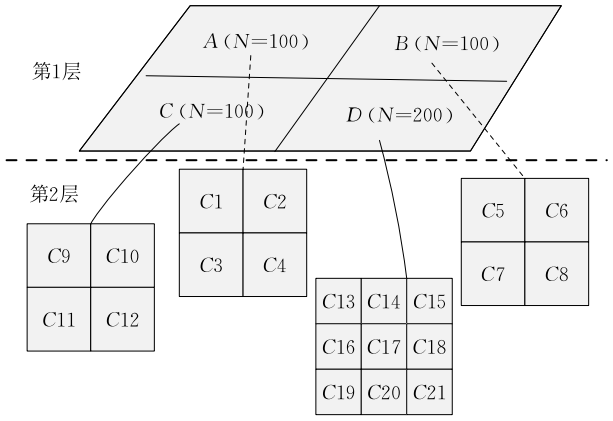


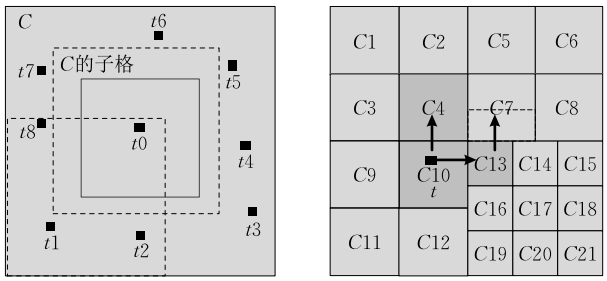
图 8 文献[39]中两层自适应网格示意图

网格的分割通常根据参与者对隐私保护强度的要求即 K-匿名策略中 k 的大小来改变.为了在两层可自适应网格实现 ε-差分隐私保护,需要在第 1 层加入的噪音参数 $\lambda_1=2\times\Delta f/\varepsilon_1$,第 2 层加入噪音参数 $\lambda_2=2\times\Delta f/\varepsilon_2$,其中 $\varepsilon=\varepsilon_1+\varepsilon_2$,文献[43]证明了按照上述方法添加噪音后能够达到预期的匿名效果.

(3)任务下达阶段.根据参与者对某一任务的接受率 p^a(the Probability of Acceptance Rate)和请求者对该任务能被顺利完成的最大期望值 EU 计算出地域群播的范围.PA 受参与者与任务之间的距离影响,距离越近,接受率越高;距离越远,接受率越低;当距离超过最大行进距离(Maximum Travel Distance,MTD)时,接受率为 0.接受率也可以是服从斜率为 s 的 Zipf 分布,s 值越大,p^a 下降越快.通常 EU 的值会比单个参与者的 p^a 值大,因此,地域群播需要把任务分配给多个参与者.定义 X 为参与者是否接受任务的随机函数:$P(X\text{ 为接受})=p^a$,$P(X\text{ 为拒绝})=1-p^a$.假设现在有 w 个互相独立的参与者,则 X 服从(w,p^a)二项式分布.$U=1-(1-p^a)^w$,U 表示至少有一个参与者接受任务的概率,推广到一个任务需要 k 个人合作的情形,此时

$$U=1-\sum_{i=1}^{k}\binom{w}{i}(p^a)^i(1-p^a)^{w-i}.$$

设 U_{c_i} 表示第 c_i 个网格对任务的接受率,U_{c_i} 以递减的方式存入一个堆栈中,算法按照堆栈中 U_{c_i} 的值逐个添加网格,每次添加一个网格,新的 U 值 $U'=1-(1-U)(1-U_{c_i})$,直到网格边界超出 MTD 或者 $U'\geq EU$ 为止.最后,为了进一步减少系统开销,网格被切分成更小的块(如图 9).



(a)单个网格切分 (b)多个网格切分

图 9 文献[39]中网格切分示意图

在实际应用中,增加一个网格中少数几个节点就可以达到 EU 值的下界.对于单个网格,从中心点开始切分,而对于多个网格的组合方式,为了确保切分之后网格仍具有连续性和紧凑性,切分从紧邻已选网格的边进行.例如:c_1,c_2,c_3,c_4 比 c_1,c_2,c_4,c_7 更加紧凑,从而在保护参与者位置隐私的同时减少下发任务时的跳数,降低系统开销.

Tessellation[44]为保护参与者的位置信息,将参与者的位置坐标映射到一个二维空间(tile).在参与式感知应用中,当参与者上传感知数据时,参与者使

用 tile 的 ID 代替自己的真实位置，从而保护自己的位置隐私. 其核心思想是在 tile 中圈住 k 个不同的参与者，达到 K-匿名的效果. tile 的空间通过横向和纵向扩展，直到满足参与者数大于等于 k 为止，最后构建出一个 tile 地图. 横向和纵向的扩展对离群点比较敏感：当参与者的位置分布比较稀疏时，为了获得 K-匿名，区域需要扩展的空间也会成比例增加. 因此，当数据分布比较稀疏时，应当有一种更好的获取匿名区域的方法. 当数据分布比较密集时，Tessellation 能够快速地保护参与者的位置信息. 它比文献[39]实现简单，不需要添加噪音等操作，但其应对背景知识攻击能力弱.

Microaggregation[45]原本是数据库隐私保护的一个概念，而在参与式感知隐私保护中，主要在参与者位置分布比较稀疏时使用. 其基本思路是：先递归生成一个具有 k 个参与者的等价类(Equivalence Class，EC)，然后对未加入等价类的边界值进行处理，称之为等价类扩展(EC extension)，通过计算离群点到等价类的平均欧式距离或离群点到等价类的中心距离等方法进行扩展. 前者是为了确保 K-匿名而强制实行的，它使得高地理相似度的参与者簇拥在一块；后者是允许未被划入等价类的参与者和邻近的、已达到 K-匿名的等价类进行合并，以适应参与者的真实分布. 相比 Tessellation，这不仅节省了系统为一个参与者而大范围寻找 $k-1$ 个参与者的开销，而且加快了等价类的生成速度，但 Microaggregation 同样不能有效防止背景知识攻击.

HV-MDAV(Hybrid Variable Size Maximum Distance to Average Vector)[46]结合 Tessellation 方法和 Microaggregation 的概念(如表 2)，其主要目的是保护参与者位置隐私，并且能够在参与式感知动态的环境下，能够同时处理参与者分布稀疏与密集的情况，进一步提供高质量的服务.

表 2　MT 和 Microaggregaion 结合

User ID	Tile ID	Location	Tile Mean	Class Number/Mean
1	1	(1.5,6.0)	(3.5,3.67)	1/(4.33,5.17)
2	1	(4.5,4.0)	(3.5,3.67)	1/(4.33,5.17)
3	1	(4.5,1.0)	(3.5,3.67)	2/(6.33,1.33)
4	2	(6.5,2.0)	(7.17,2.83)	2/(6.33,1.33)
5	2	(7.0,5.5)	(7.17,2.83)	1/(4.33,5.17)
6	2	(8.0,1.0)	(7.17,2.83)	2/(6.33,1.33)

HV-MDAV 首先对 Tessellation 方法进行改进(Modified version of Tessellation，MT)，参与者改用 tile 的中心点代替真实位置，例如表 2 中，假设 k 等于 3，则 User 1、2、3 的位置都用 tile 1 的中心点(3.5,3.67)表示；User 4、5、6 的位置都用 tile 2 的中心点(7.17,2.83)表示. 由上面的介绍可知，Microaggregation 可以使得地理相似度高的参与者簇拥在一起，按照此方法，可以延伸得到 V-MDAV(Variable Size Maximum Distance to Average Vector)方法. 它根据欧几里得距离得到等价类，此时 k 仍被设为 3，从而 User 1、2、5 被分为一类其位置都用等价类 1 的中心点(4.33,5.17)表示；User 3、4、6 分为另一类其位置都用等价类 2 的中心点(6.33,1.33)表示. HV-MDAV 分两种不同的情况使用了 MT 和 V-MDAV 方法：(1)当参与者在不同区域的分布比较稀疏且分布均匀时，V-MDAV 能够做出更好的决策；(2)当参与者分布比较密集时，MT 执行效果更好. 因此，判断分布是否密集的方法直接影响系统性能.

最后，HV-MDAV 在第三方不可信的情况下，采用随机高斯分布噪音进行数据扰乱，将有效地保护参与者位置分布稀疏或密集时的位置信息.

LD-VMDAV(L-diverse version of Hybrid Variable Size Maximum Distance to Average Vector)[47]是 HV-MDAV 的扩展，其使用环境和目的和 HV-MDAV. 它在 HV-MDAV 的基础上添加了 L-多样性. 例如，在表 2 中，如果先不考虑属性列 Tile ID，将 User 1、2、5 组成匿名区间，则获得 2-多样性，即匿名区间内存在(4.33,5.17)和(6.33,1.33)两个不同的位置信息. LD-VMDAV 在实现 L-多样性时具体采用的是时间维度的多样性. LD-VMDAV 先根据参与者位置的分布情况进行 K-匿名，然后在 K-匿名策略的上增加时间多样性，使得同一个匿名区域拥有不同特征的地理位置信息，实现多层匿名. 时间多样性去除了时间和参与者的关联性，进一步增强了匿名效果，有效地防止了背景知识攻击和同质性攻击.

PiRi(Partial-inclusivity and Range independence)[48]用于保护参与者上传感知数据时的位置隐私，其主要思想是扩大匿名半径以及查询覆盖区域. PiRi 中的区域划分方法与 LD-VMDAV 有所不同，它不是通过横向或纵向地移动，而是根据参与者的位置构建维诺图(Voronoi Diagram)(如图 10). 维诺图又被称为泰森多边形，指的是将平面上有区别但同性质的 N 个主节点划分成 N 个区域，每个主节点对应一个区域，区域内的若干次节点到该主

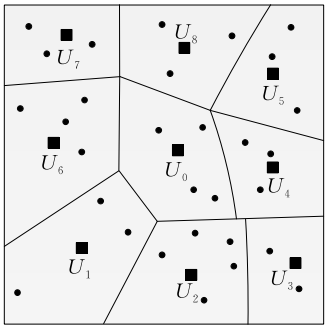


图 10 维诺图示例

节点的距离最近. PiRi 主要分为以下 3 个步骤：

(1) 数据预处理. 根据参与者的位置构建维诺图.

(2) 范围依赖性(range dependency)处理. 范围依赖性是指查询范围是依赖于参与者的维诺图单元的大小. 由于参与者 U_i 被要求发送自己的匿名区域，其匿名区域半径 r_i 至少包含其所在单元格，而每个参与者的匿名区域和所处单元格大小是一一对应的，因此存在不同大小的 r 值，这使得隐私数据存在暴露的风险. 比如：在图 11 中，当 k 为 2 时，U_1 需要结合 U_2 来达到匿名效果，它向位置服务器发送 U_1 和 U_2 以及自己的匿名半径 r_1. 后者会查询与 r_1 半径相匹配的查询位置即同参与者所处维诺图单元格相同的位置信息，这样就得出数据来自 U_1 而不是 U_2，匿名失败. 为防止因范围依赖性导致隐私数据泄露，参与者 U_i 不仅要和其他 $k-1$ 个参与者一起形成 K-匿名，而且上传匿名半径 r 选择 k 中最大的 $r_{\max}$.

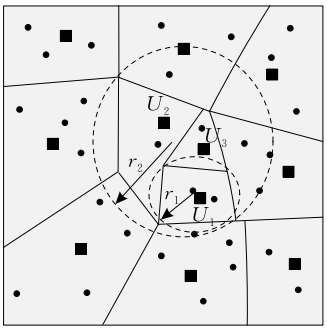


图 11 PiRi 中背景知识攻击示意图

(3) 查询相关性处理. 由于服务器接收所有用户的查询信息(all-inclusivity)，从而可以利用所收集的信息对参与者形成选择性任务攻击. 图 12 给出了具体事例，简单解释，现在假设只有 U_1、U_2 和 U_3 3 个参与者，U_1 通过结合 U_2 隐藏自己的位置信息. 同理，U_2 通过结合 U_1 隐藏自己，结果使得 U_1 和 U_2 具有相同的查询半径和查询区域. 从图 12 中我们可以看到，U_3 也是通过结合 U_1 来隐藏自己. 当服务器选择 3 个不同的任务让 3 个参与者分别提交结果，服务器很容易辨别出 U_3，因为 U_3 是唯一一个只出现一次的参与者. 观察可知 3 个参与者之间的查询区域有很大一部分是重叠的. 因此，当某一查询区域完全包含另一个查询区域时，后者的数据应当结合前者的查询区域进行提交. 问题转化为：如何用最少的查询区域覆盖所有的顶点. 顶点覆盖(V-cover)问题可以规约到最小元素覆盖问题，NP-hard 问题. 在集中式架构中，可获得环境的全局性信息，因此可以使用一种大家比较熟悉的启发式算法-贪婪算法来求解该问题. 在每一次迭代中，选取覆盖顶点数最多的查询区域. 在分布式体系结构中，需要对启发式贪婪算法进行扩展，由此设计一种表决机制使得参与者同意在本地和邻近网格中选择一个代表，选择依据是：① 查询区域包含的顶点数目；② 查询区域被其他查询区域覆盖的次数. 前者用来度量候选查询区域的大小，后者用来度量候选查询区域的使用频度. 选取完成后，参与者结合所在的代表查询区域的位置信息提交感知数据，这使得服务器无法获知参与者的具体位置，但通信代价会相应地增加.

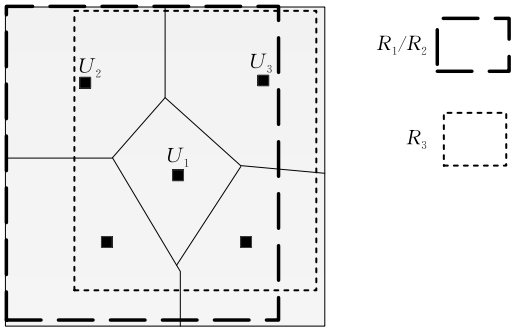


图 12 PiRi 中查询相关示意图

Gao 等人[49]通过参与者与周围其他参与者建立可信连接，形成 K-匿名，从而保护参与者的轨迹隐私. 参与者有数据需要提交时，先把数据转发给邻近的节点，然后分别上传数据，从而防止不可信第三方泄露隐私数据. 文献[49]中的协议与分组统计有所不同，它并不把隐私数据先进行切分，然后在服务器端或者最终用户端进行数据重组，而是在需要匿名时与周围的参与者达成某种约定，然后将数据加密整体转发给 $k-1$ 个用户，其他用户再按照自己所在位置向服务器进行数据上传操作，数据一经上传

就会被销毁.该协议的最大特点就是及时性和动态性好.尽管能量消耗偏大,但它能够有效防止服务器同时获取参与者的个人信息和感知数据.参与式感知更加侧重于服务质量和隐私保护强度.以能耗换服务质量和隐私保护强度也是参与式感知中一种常用的折中方式.

文献[50]提出在服务器不可靠的情况下,既保护参与者上传数据时的位置信息又保护请求者向服务器查询数据时的位置隐私,它结合了 K-匿名和可信第三方策略.以请求者为例,具体方法如下:首先,请求者将查询信息发送给可信的匿名服务器,匿名服务器将请求者的位置隐匿在相应的匿名区间.然后,匿名服务器代表请求者将查询请求发送给参与式感知服务器,参与式感知服务器根据请求信息整合出备选的感知数据并将数据返回给匿名服务器.最后,匿名服务器去除匿名信息,将最终结果发送给请求者.上述匿名过程中,文献[50]利用局部敏感哈希(Locality-Sensitive Hashing,LSH)将原始数据空间中的两个相邻数据点通过相同的映射或投影变换后,这两个数据点在新的数据空间中仍然相邻的概率很大,而不相邻的数据点被映射到同一个桶的概率很小.这样不仅可以将位置数据散列到至少包含 k 个用户的高维分区空间,而且使得位置接近的点映射在一个桶中.尽管局部敏感哈希的选取会增加系统的计算代价,但该协议既保护了用户的隐私,又可以提高查询效率,特别是 KNN 查询.

文献[51]提出一种用户自适应的位置隐私保护技术.主要防止那些试图通过网络信道截获感知数据的位置信息或利用历史位置信息等背景知识推测用户的位置信息的攻击.具体方法如下:首先,为了保证用户最低的匿名效果,系统为用户定义了一个用户位置隐私下界.当用户上传感知数据时,他可以自定义一个不小于位置隐私下界的值,从而使得用户的隐私保护级别随着具体的应用而有所不同.其次,系统根据用户自定义隐私保护级别生成多个网格单元并加入扰乱噪音,用户可以从中任意挑选一个当作自己上传感知数据的位置.仅当挑选出的网格单元大小和用户自定义的隐私保护级别对等时,用户的位置才被认为是安全的.因此,文献[51]既满足了用户自定义位置隐私级别的需求,又防止了背景知识攻击和窃听攻击,但网格单元的动态生成增加了系统计算代价.

文献[52]提出一种结合匿名隐私保护机制、数据扰乱和具体的数字加密技术来保护参与者位置隐私的同时提高感知数据的质量和减少网络能量消耗,主要防止攻击者获取参与者的位置信息.与文献[51]类似,文献[52]中服务器是可信的.首先,服务器通过参与者的位置信息划分出不同的网格单元;然后,匿名的隐私保护机制则根据参与者位置所在网格单元的大小动态地变化.当网格单元较小的时,利用具体的数字加密技术来保护用户的隐私.由于感知数据中的位置信息更接近于参与者的真实位置,因而增加了感知数据的质量;当网格单元较大时,则利用数据扰乱技术,因为单元格比较大,本身可构成一个匿名区域,从而有效降低攻击者攻击成功的概率.该协议在使用过程中可以有效提高数据质量和减少网络能耗,但维护和确定网格单元的大小会带来比较大的计算开销.

3.4　基于数字加密技术

数字加密技术主要是防止攻击者直接获取用户的隐私数据,其次是防止不可信服务器获取过多用户的敏感信息.传统的加密技术中的难点是如何在加密数据上直接进行计算和对数据进行完整性验证以及减少计算开销.但在参与式感知应用中,研究的侧重点是数字加密能带来的端到端的保密服务,对于加密过程中的复杂计算等偏密码学方面的研究相对较少,因此本综述只是对加密思想、加密过程、加密方法和加密效果做了基本介绍.

IBE(Identity Based Encryption)[53-54]基于身份加密是一种有效的保护参与者数据隐私的方法.它不依赖于传统公钥基础设施(Public Key Infrastructure,PKI)以及通过认证中心来向参与者绑定公钥,用户的身份信息(如电话号码、身份证号码和邮件地址等)可以直接作为用户的公钥[55].IBE 中相应的私有密钥则由受信任的私钥生成器(Private Key Generator,PKG)管理和颁发.IBE 能为移动节点和请求者提供有效安全的通信,它分为以下 4 个步骤:

(1) Setup(1^λ)阶段.PKG 设置初始化系统参数和主密钥.

(2) BlindExtract 阶段.用户根据自己和 PKG 之间的交互式协议获取公钥,协议保证 PKG 无法截取用户的敏感信息.

(3) IBEn 阶段.用户用任意一个身份公钥为指定的接受方加密一条消息.

(4) IBDec 阶段. 接收方从 PKG 处获取相应的私钥,从而得到明文信息.

IBE 中用户的身份与用户有着直接的联系,因此无需通过数字证书进行绑定,从而避免了传统公钥密码体制中因管理大量用户证书带来的网络能耗和存储资源,比较适合资源受限的参与式感知应用,但是 IBE 中主密钥和密钥的托管存在比较大的安全隐患.

AIBE(Additively homomorphic Identity Based Encryption)[56] 使用一种基于身份特性的求和同态加密函数,它是 IBE 的扩展. 与 IBE 相比,AIBE 更容易对密文进行扩展,它实现密文逐元素同态相加,但当参与者数量较小时,单一的同态相加方法不足以抵御共谋攻击. AIBE 主要分为 4 个阶段:

(1) 初始化阶段. 生成两个双线性组,$(G=\langle g\rangle, q, e: G\times G\rightarrow G_T)$,其中,$G_T=\langle \bar{g}\rangle$,$\bar{g}=e(g,g)$,$q$ 为质数. 选择参数 $x\in Z_q^*$,令参数 $y=g^x$,取哈希函数 $H:\{0,1\}^*\rightarrow G^*$ 消息空间设为 $M=Z_N=\{0,1,\cdots,N-1\}$,其中 $N=p(n)<q$,同时,设密文空间为 $C=G^*\times G_T$,最后输出私钥参数 $mpk=(q,g=\langle g\rangle, G_T=\langle \bar{g}\rangle, e, y, H)$ 和 $msk=x$.

(2) 信息提取阶段. 计算并输出密钥 $sk_{id}=H(id)^x$.

(3) 加密阶段. 选择随机数 $r\in Z_q^*$,输出密文 $c=(g^r, \bar{g}^m\cdot e(H(id),y)^r)$.

(4) 解密阶段. 解析密文 c,使之成为 (c_1,c_2),$\log_{\bar{g}}\bar{m}=\log_{\bar{g}}\left(\dfrac{c_2}{e(sk_{id},c_1)}\right)=m$,通过计算可知 $\bar{m}=c_2/e(sk_{id},c_1)$,$m=\log_{\bar{g}}\bar{m}$,加密和解密阶段数据中加入随机变量,服务器只充当一个信息传递者,从而较为有效地保护了参与者的数据隐私.

PC(Paillier Cryptosystem)[57-59] 与 AIBE 类似,也属于加同态加密技术,在 AIBE 的基础上,PC 将用户的位置信息转换成可操作的数字符号. 它可以从操作数密文的代数运算得到代数运算的密码文本. 这样既可以保护参与者的数据隐私,又可以保护参与者的位置隐私,但信息的转换和选取提高了实现难度. 其主要内容分为以下 4 个步骤:(1) 初始化阶段. 服务器从用户位置数据库中抽取用户位置和用户位置采集地点,并将其发送给参与者;(2) 密码生成准备阶段. 参与者通过 PC 生成一对密钥,并将公钥和公钥加密后的位置信息发送给服务器;(3) 加密阶段. 服务器在密文空间上计算参与者位置和位置数据库中每一个位置之间的欧式距离,并将结果返回给参与者;(4) 解密阶段. 参与者首先将接收到的欧式距离解密,找出 k 个最小的距离,然后根据 k 个最小的距离和初始化阶段获得的信息估算自己当前的位置. 同理,将上述的用户位置和欧式距离分别换成用户数据和相似度以保护参与者的数据隐私. 由于用户的真实位置和数据都是在用户端计算所得,因此不可信服务器不能获取用户的真实信息,从而有效抵御共谋攻击.

CP-AABE (Ciphertext Policy Anonymous Attribute-Based Encryption)[60] 基于属性的匿名加密技术是 CP-ABE(Ciphertext Policy Atttribute-Based Encryption)基于属性加密技术的扩展,它主要保护参与者上传数据时的位置隐私. CP-ABE 与上述介绍的加密技术类似,主要分为初始化阶段、密码生成阶段、加密阶段和解密阶段. 其中,在加密阶段,它允许加密方将访问权限嵌入密文中且每个解密都是基于一些属性集合. 因此,在解密阶段,有且仅有那些属性集合满足要求的解密方才能解开密文. 与 CP-ABE 不同的是,CP-AABE 不仅保证了数据的机密性,而且提供匿名机制. 尽管一个合法的解密方可以解开密文,但他只能获取加密方的匿名位置信息. CP-AABE 通过加密技术和访问权限控制成功防止不可信第三方和非法用户窃取参与者的隐私信息,但匿名和属性值嵌入增加了系统的通信代价.

4 对比分析与评价

参与者的参与方式、参与者对隐私强度的要求、参与式感知应用场景、网络环境等都与隐私保护技术的性能密切相关. 因此对隐私保护技术的性能评价应当从多个角度进行综合评价. 表 3 按照各协议所属策略、实现难易程度、网络通信质量、隐私保护强度和能耗等方面对现有研究成果的优缺点进行了总结评估.

表 4 则从协议主要隐私保护对象、所用学科方法和所防御的攻击手段等多个方面对隐私保护技术进行了汇总对比.

表 3　隐私保护协议的优缺点

代表协议	所属策略	采用技术	主要优点	主要缺点
HP³[21]	分组统计	逐跳加密	实现相对简单	通信代价大
PESP[22]		随机转发	隐私保护度强	通信和计算代价较大
M-PERM[24]		数据扰乱 逐跳加密	精确聚集	计算复杂
PMG[25]			支持不同隐私保护强度	通信代价大
PriSense[26]			支持多种类型聚集	通信代价较大
Wang&Zhang[27]		延迟转发	有效防御信号追踪攻击	对节点的处理能力要求高
PEPSI[29]	第三方验证	定向通信	通信代价低	隐私保护强度弱
PEPPeR[31]			通信代价较低	通信时延较长
ARTSense[32]			高可靠性	实现复杂
LotS[34]			支持不同隐私保护强度	需要物理层硬件支持
TrPF[35]			计算代价小	通信时延较长
AnonySense[37]		网络编码	支持多种用户类型	需要物理层硬件支持
To&Ghinita[39]	*K*-匿名	数据扰乱	支持不同隐私保护强度	实现复杂，计算代价大
Tessellation[44]		伪装数据	计算代价小	隐私保护度弱
Microaggregation[45]			通信代价小	隐私保护度弱
HV-MDAV[46]			隐私保护度强	性能不稳定
LD-VMDAV[47]		数据扰乱	隐私保护度强，服务精度高	通信时延较长
PiRi[48]		扰动直方图	隐私保护度强	通信和计算代价大
Gao&Ma[49]		虚假源	隐私保护度较强	通信代价大
文献[50]		数据扰乱	隐私保护度较强	计算代价大
文献[51]		数据扰乱	用户可自主选择	计算代价大
文献[52]		数据扰乱	通信和计算代价小	隐私保护弱
IBE[53]	数字加密	同态加密函数 IBE scheme	支持非线性聚集	安全等级较低
AIBE[56]			密文容易扩展	应对共谋攻击能力较弱
PC[58]		同态加密函数 paillier scheme	应对共谋攻击能力强	实现相对复杂
CP-AABE[60]		双线性映射	应对共谋攻击能力强	通信代价大

表 4　隐私保护协议的保护对象和防御的攻击手段

代表协议	保护对象	学科方法	防御的攻击手段
HP³[21]	参与者数据隐私	无向图	数据篡改
PESP[22]		正态分布	背景知识攻击
M-PERM[24]		回归模型	背景知识攻击
PriSense[26]		概率统计	共谋攻击
PEPSI[29]		集合	背景知识攻击
ARTSense[32]		模运算	时间关联分析
LotS[33]		哈希函数	共谋
AnonySense[37]		集合	数据篡改
PriSense[26]		概率统计	共谋攻击
IBE[53]		模运算	背景知识攻击
AIBE[56]		模运算	背景知识攻击
PC[58]		模运算	共谋攻击
To&Ghinita[39]	参与者位置隐私	拉普拉斯分布	背景知识攻击
Tessellation[44]		欧式距离	背景知识攻击
Microaggregation[45]		异种属性相似度	背景知识攻击
HV-MDAV[46]		欧式距离	背景知识攻击
LD-VMDAV[47]		欧式距离	同质性攻击
PiRi[48]		维诺图	共谋攻击
PC[58]		模运算	共谋攻击
Gao&Ma[49]		关系映射	逐跳回溯追踪
文献[50]		哈希函数	共谋攻击
文献[51]		信息熵	背景知识攻击
文献[52]		图论	共谋攻击
文献[60]		模运算	共谋攻击
TrPF[35]	参与者轨迹隐私	图论，积分，信息熵	逐跳回溯追踪
Gao&Ma[49]		关系映射	逐跳回溯追踪
PEPPeR[31]	请求者数据隐私	模运算	特定对象攻击、选择性攻击

(1)从总体上看,基于分组统计的技术容易造成通信量大和网络延迟增加等问题,因此,需要对隐私强度和服务质量做适当的权衡;基于第三方验证的技术尽管去除了服务器与参与者的直接联系,但第三方验证需要更多的系统处理时间,导致了服务质量的下降;基于 K-匿名的技术的难点是如何高效动态地构建匿名区域;基于加密的技术需要更多地考虑计算量和实现复杂度.现有的参与式感知隐私保护方法,例如,文献[26,39,48]很大程度上依赖于可信的基础设施(通常是汇聚节点或者基站)和加密方法.实际上,参与式感知中的数据通常在参与者端进行收集和分析[61-62],而不是在一些可信的基础设施上,这可能泄露参与者的隐私数据.此外,现有的隐私保护数据聚合方法往往只能提供有关求和、平均、最大值、最小值等有限的信息,这是远远不能满足参与者的需要,并在很大程度上限制了参与者在参与式感知应用中进行复杂的数据分析能力.例如,目前的参与式感知隐私保护中的回归建模方法通常是预先假定已知模型表达式后对模型系数进行估计,因而缺乏找到最佳拟合数据的能力[63],或者在模型拟合期间不得已将模型系数泄露给参与者[64].因此,在参与式隐私保护技术中需综合考虑实现难易程度、隐私保护强度、网络通信质量、性能等各方面因素.

(2)分组统计策略,以保护参与者数据隐私为主,对位置隐私保护的研究较少.在实现难易程度和通信方面,PESP 和文献[27]因为采用随机转发和延迟转发,实现简单,不需要节点增加新的数据包,因此节点的存储代价和计算代价相对较小,通信时延和数据丢失率都比较低;在防御的攻击手段和隐私保护方面,HP^3 和 PMG 引入了逐跳加密机制应对背景知识攻击逐跳回溯追踪攻击.数据扰乱和逐跳加密方法比随机转发和延迟转发方法的隐私保护效果好;在能耗方面,M-PERM、PMG 和 PriSense 由于既采用了数据扰乱技术又采用了逐跳加密机制,通信能耗比 PESP 协议大;在网络适用方面,PMG 支持不同隐私保护强度,而 PriSense 则支持多种不同数据类型聚集.

(3)第三方验证策略,尽管保护参与者数据隐私的研究占绝大多数,然而,其中也有专门为了保护请求者数据隐私的协议,如 PEPPeR.在实现难易程度方面 AnonySense 由于需要物理层硬件支持而变得比较难实现;在通信方面,所有的协议都在参与者和服务器之间增加了一个中间验证,形成一个定向通信,这增加了网络时延.同时,AnonySense 因为用到网络编码技术,虽然增加了计算能耗,但其通信能耗较低;在防御的攻击手段和隐私保护方面,PEPPeR 可以防御背景知识攻击,AnonySense 能防御主动攻击者对数据的篡改,ARTSense 和 TrPF 提供高可靠性隐私保护;在能耗方面,由 ARTSense 和 LotS 的数据传输量大而导致能耗较高.

(4)K-匿名策略,主要是为了解决参与者的时空位置隐私问题.在实现难易程度方面,PiRi 和文献[42]因为需要提前构建维诺图或对参与者的位置信息进行关系映射,导致数据预处理时任务比较繁重.文献[29]结合了差分隐私机制,可以为用户提供不同强度的隐私保护,而拉普拉斯分布噪音的加入让网络的通信量变大.文献[37-38]采用最原始的 K-匿名机制,实现简单,但隐私保护强度低.HV-MDAV 和 LD-VMDAV 都能防御背景知识攻击,而后者因为引入了 L-多样性,能有效防御时间关联攻击.文献[50-52]的主要思想是通过数据扰乱来达到 K-匿名,其中文献[50]需要可信第三方支撑,而文献[51]在很大程度上给用户提供自主性,文献[52]则更多地从能耗出发,既保证参与者的隐私又降低网络能耗.

(5)数字加密策略,在实现难易程度、网络通信质量、能耗方面,IBE 和 AIBE 由于采用的是同态加密函数,相对比较简单,计算代价小,通信和能耗也相对较小;在防御的攻击手段和隐私保护方面,PC 使用加同态加密函数 Paillier,它与 CP-AAPE 一样,应对共谋攻击能力强,能为参与者提供较强的隐私保护.同时,PC 引入了可并行的代数运算,大大提升了加密和解密的速度.

5 未来工作展望

由于参与式感知网络隐私保护技术涉及传感器网络、数理统计、运筹学等多个学科且属于新兴研究领域,有很多热点难点问题值得去关注.参与式感知给用户带来便利的同时也为隐私保护技术提出了新的要求,很多具有挑战性的问题有待进一步研究.

(1)基于隐私数据分析和隐私保护模型建立的技术

参与式感知相对于传统的传感器网络而言具有随时随地可感知、输入简单、参与者负担小、数据维度大等特征.与此同时,参与感知的节点很可能出现在无人监督、环境复杂的地方导致数据难以被监督

管理，数据稳定性差等弊端. 因此，为参与式感知应用提供一种快速有效的跨数据类型的隐私保护模型建立和移动多维数据分析的隐私保护技术是十分紧迫的.

(2) 新型的隐私保护体系结构

为了能够处理传统的单个服务器且不可信问题，可考虑新型的隐私保护体系结构. 比如：多个服务器分工合作、权限分割，一部分服务器负责处理参与者的公开数据，一部分服务器负责处理参与者的敏感数据. 不同服务器之间通过保密协议控制彼此，不可以共谋. 多个服务器的出现既能避免成为首要的攻击目标，又可以减少请求者和参与者的请求次数和能耗. 同时，利用对称加密和非对称加密方法，对数据分段进行加密，从而防止不同服务器同时获取参与者的公开数据和敏感数据. 因此，如何设计多个服务器隐私解决方案中的保密协议以及如何对数据分段加密是进一步需要研究的方向.

(3) 隐私设置与参与者的相关因素相结合的隐私保护技术

在参与式感知应用中，参与式服务端、参与者和最终用户属于 3 个利益相关者，未来的一个关键挑战是如何更好地在隐私设置时考虑参与者的相关因素. ① 量身定做的隐私接口：隐私是高度个人，它取决于参与者的观点和看法. 因此，简化参与式感知应用中的复杂配置，同时减少参与者周边存在的隐私威胁是十分重要的. 然而，现有的参与式感知隐私保护协议很少能既易于参与者去理解又能提高隐私保护强度；② 结合参与者的反馈信息：参与者的感知度和满意度应被列入隐私保护的评价指标，这可以促进参与者和应用程序之间的紧密合作，根据参与者的直接反馈信息构建隐私保护解决方案，能提高参与者的满足度.

(4) 组合隐私解决方案

参与式感知保护对象包括请求者和参与者，保护时刻包括数据上传和数据下达. 上述 4 个方面可以组合成多种需要隐私保护的方向. 然而，现有隐私保护技术多数侧重于某一个或有限几个研究方向. 因此，未来的研究问题应包括如何在动态的、不可预测的参与式感知环境下组合使用现有隐私保护技术与隐私分级. 其中，隐私分级是指同一应用场景具备不同的隐私保护等级. 隐私分级模型通过实时的可信度计算和安全强度离散处理来为海量数据提供动态隐私判定方案.

(5) 权衡隐私保护强度、性能和数据完整性的隐私保护技术

强的隐私保护机制可能影响数据的感知时延或数据完整性，而保护数据完整性的同时通常会降低隐私保护强度. 因此，隐私保护强度和数据完整性之间的折衷是必须的：① 隐私保护强度和数据完整性：在参与式感知应用中，为了鼓励参与者积极参与，需要对参与者进行隐私保护，但隐私保护强度和数据完整性之间存在一个内在的冲突. 如果为参与者提供完全的匿名，那么很难保证感知数据的可信度和完整性. 因此，设计既满足隐私保护需求又满足数据完整性的隐私保护系统是一个重要的研究方向；② 隐私保护强度、性能和数据完整性：参与者提供的感知数据的完整性将直接影响参与者的隐私保护强度. 感知数据越完整和精确，隐私保护强度越强. 数据完整性可通过将参与者提供的感知数据同其他参与者的感知数据进行相似度计算所得，得到的反馈值作用于参与者的可信度或者信誉等级. 同时，新的隐私保护技术需要考虑参与者与服务器之间的多次通信对系统性能产生的影响.

(6) 隐私可度量和评估基准

现有研究中采用不同的度量标准，使得不同协议之间难于直接进行比较. 如何形成一个统一的度量方案来量化隐私，是一个长期的研究目标. 为获得这样的度量标准，目前使用的隐私度量需要进行检验，以确定哪些输入参数. 例如，同一区域参与者的数量，实际的和扰动位置的轨迹等被认为是计算隐私度的必要步骤. 此外，其他应用程序领域的隐私度量应该被借鉴. 比如数据库中的用户等级划分方法，并分析其在参与式感知中的适用性，通用的隐私指标框架可进一步研究.

致　谢　本文审稿专家和编辑老师提出了很多宝贵的意见和建议，在此表示感谢！

参　考　文　献

[1] Zhang Zhaoyang, Wang Honggang, Lin Xiaodong, et al. Effective epidemic control and source tracing through mobile social sensing over WBANs//Proceedings of the INFOCOM. Turin, Italy, 2013: 300-304

[2] Mun M Y, Reddy S, Shilton K, et al. PEIR the personal environmental impact report as a platform for participatory sensing systems research//Proceedings of the MobiSys. Krakow, Poland, 2009: 55-68

[3] Sehgal S, Kanhere S S, Chou C T. MobiShop: Using mobile phones for sharing consumer pricing information//Proceedings

of the Conference on Distributed Computing in Sensor Systems (DCOSS). Santorin Island, Greece, 2008

[4] Dong Yifei, Blazeski L, Sullivan D, et al. PetrolWatch: Using mobile phones for sharing petrol prices//Proceedings of the 7th Annual International Conference on Mobile Systems, Applications and Services (MobiSys). Krakow, Poland, 2009

[5] Krontiris I, Freiling F C, Dimitriou T. Locaition privacy in urban sensing networks: Research challenges and directions. IEEE Wireless Communications, 2010, 17(5): 30-35

[6] Kapadia A, Kotz D, Triandopoulos N. Opportunistic sensing: Security challenges for the new paradigm//Proceedings of the Communication Systems and Networks and Workshop (COMSNETS). Bangalore, India, 2009: 1-10

[7] Liang Xiaohui, Zhang Kuan, Shen Xuemin, Lin Xiaodong. Security and privacy in mobile social networks: Challenges and solutions. IEEE Wireless Communications, 2014, 21(1): 33-41

[8] Huo Zheng, Zhang Xiao-Jian, Meng Xiao-Feng. Privacy-preserving in social, mobile and location-based services. Sciencepaper Online, 2013: 1-10(in Chinese)
(霍峥, 张啸剑, 孟小峰. 移动社交网络及位置隐私保护研究综述. 中国科技论文在线, 2013: 1-10)

[9] Xu Jian-Liang, Hu Hai-Bo, Chen Qian. Location privacy-preserving in social, mobile network. Communications of the China Computer Federation, 2014, 10(6): 58-62(in Chinese)
(徐建良, 胡海波, 陈乾. 移动社交网络中的位置隐私保护. 中国计算机学会通讯, 2014, 10(6): 58-62)

[10] Wu Zhen-Gang, Sun Hui-Ping, Guan Zhi, Chen Zhong. Survey on location privacy preservation of continuous spatial queries. Application Research of Computers, 2015, 32(2): 321-342(in Chinese)
(吴振刚, 孙惠平, 关志, 陈钟. 连续空间查询的位置隐私保护综述. 计算机应用研究, 2015, 32(2): 321-342)

[11] Liu Chenxu, Liu Yun, Zhang Zhenjiang, Cheng Ziyao. High energy-efficient and privacy-preserving secure data aggregation for wireless sensor networks. International Journal of Communication Systems, 2013, 26(3): 380-394

[12] Zeng Weini, Lin Yaping, Yu Jianping, et al. Privacy-preserving data aggregation scheme based on the P-function set in wireless sensor networks. Ad Hoc & Sensor Wireless Networks, 2014, 21(12): 21-58

[13] Hu Ronghua, Dong Xiaomei, Wang Daling. Protecting data source location privacy in wireless sensor networks against a global eavesdropper. International Journal of Distributed Sensor Networks, 2014: 1-17

[14] Kiran Mehta, Donggang Liu, Matthew Wrigh: Protecting location privacy in sensor networks against a global eavesdropper. IEEE Transactions on Mobile Computing, 2012, 11(2): 320-336

[15] Peng Hui, Chen Hong, Zhang Xiaoying, et al. Location privacy preservation in wireless sensor networks. Journal of Software, 2015, 26(3): 617-639(in Chinese)
(彭辉, 陈红, 张晓莹等. 无线传感器网络位置隐私保护技术. 软件学报, 2015, 26(3): 617-639)

[16] Yang Dejun, Fang Xi, Xue Guoliang. Truthful incentive mechanisms for *K*-anonymity location privacy//Proceedings of the INFOCOM. Turin, Italy, 2013: 2994-3002

[17] Groat M M, He Wenbo, Forrest S. KIPDA: *K*-indistinguishable privacy-preserving data aggregation in wireless sensor networks//Proceedings of the INFOCOM. Shanghai, China, 2011: 2024-2032

[18] Sun Li-Min, Li Jian-Zhong, Chen Yu, Zhu Hong-Song. Wireless Sensor Networks. Beijing: Tsinghua University Press, 2005(in Chinese)
(孙利民, 李建中, 陈渝, 朱红松. 无线传感器网络. 北京: 清华大学出版社, 2005)

[19] Fan Yong-Jian, Chen Hong, Zhang Xiao-Ying. Data privacy preservation in wireless sensor networks. Chinese Journal of Computers, 2012, 35(6): 1131-1146(in Chinese)
(范永健, 陈红, 张晓莹. 无线传感器数据隐私保护技术. 计算机学报, 2012, 35(6): 1131-1146)

[20] Kazemi L, Shahabi C. GeoCrowd: Enabling query answering with spatial crowdsourcing//Proceedings of the SIGSPATIAL/GIS. Redondo Beach, USA, 2012: 189-198

[21] Hu Ling, Shahabi C. Privacy assurance in mobile sensing networks: Go beyond trusted servers//Proceedings of the PerCom Workshops. Mannheim, Germany, 2010: 613-619

[22] Zhang Fan, He Li, He Wenbo, Liu Xue. Data perturbation with state-dependent noise for participatory sensing//Proceedings of the 32nd IEEE International Conference on Computer Communication (INFOCOM). Orlando, USA, 2012: 2246-2254

[23] Ganti R K, Pham N, Tsai Yu-En, Abdelzaher T F. PoolView: Stream privacy for grassroots participatory sensing//Proceedings of the SenSys. Raleigh, USA, 2008: 281-294

[24] Xing Kai, Wan Zhiguo, Hu Pengfei, Zhu Haojin, et al. Mutual privacy-preserving regression modeling in participatory sensing//Proceedings of the INFOCOM. Turin, Italy, 2013: 3039-3047

[25] Chen Xi, Wu Xiaopei, Li Xiangyang, et al. Privacy-preserving high-quality map generation with participatory sensing//Proceedings of the INFOCOM. Toronto, Canada, 2014: 2310-2318

[26] Shi Jing, Zhang Rui, Liu Yunzhong, Zhang Yanchao. PriSense: Privacy-preserving data aggregation in people-centric urban sensing systems//Proceedings of the INFOCOM. San Diego, USA, 2010: 758-766

[27] Wang Wei, Zhang Qian. A stochastic game for privacy preserving context sensing on mobile phone//Proceedings of the INFOCOM. Toronto, Canada, 2014: 2328-2336

[28] Gotz M, Nath S, Gehrke J. MaskIt: Privately releasing user context streams for personalized mobile applications//Proceedings of the SIGMOD Conference. Scottsdale, USA, 2012: 289-300

[29] De Cristofaro E, Soriente C. Short paper: PEPSI: Privacy-enhanced participatory sensing infrastructure//Proceedings of the WISEC. Hamburg, Germany, 2011: 23-28

[30] De Cristofaro E, Soriente C. Participatoty privacy: Enabling privacy in participatory sensing. IEEE Network, 2013, 27(1): 32-36

[31] Dimitriou T, Krontiris I, Sabouri A. PEPPeR: A querier's privacy enhancing protocol for PaRticipatory sensing// Proceedings of the MobiSec. Frankfurt am Main, Germany, 2012: 93-106

[32] Wang X O, Cheng Wei, Mohapatra P, Abdelzaher T F. ARTSense: Anonymous reputation and trust in participatory sensing//Proceedings of the INFOCOM. Turin, Italy, 2013: 2517-2525

[33] Michalas A, Komninos N. The lord of the sense: A privacy preserving reputation system for participatory sensing applications//Proceedings of the ISCC. Funchal, Portugal, 2014: 1-6

[34] Christin D, Robkopf C, Hollick M, et al. IncogniSense: An anonymity-preserving reputation framework for participatory sensing applications. Pervasive and Mobile Computing, 2013, 9(3): 353-371

[35] Gao Sheng, Ma Jianfeng, Shi Weisong, et al. TrPF: A trajectory privacy-preserving framework for participatory sensing. IEEE Transactions on Information Forensics and Security, 2013, 8(6): 874-887

[36] Ma Mumtaj Muthu Gadhiza, Sd Sd Akthar Basha, Babu P. Privacy preserving for participatory sensing using trajectory mix-zone model. International Journal of Research Studies in Computer Science and Engineering, 2014, 1(3): 8-14

[37] Cornelius C, Kapadia A, Kotz D, et al. AnonySense: Privacy-aware people-centric sensing//Proceedings of the MobiSys. Breckenridge, USA, 2008: 211-224

[38] Shin M, Cornelius C, Peebles D, et al. AnonySense: A system for anonymous opportunistic sensing. Pervasive and Mobile Computing, 2011, 7(1): 16-30

[39] To H, Ghinita G, Shahabi C. A framework for protecting worker location privacy in spatial crowdsourcing. Proceedings of the VLDB Endowment, 2014, 7(10): 919-930

[40] Dwork C. Differential privacy. ACM Crossroads, 2006, 20(1): 1-12

[41] Cormode G, Procopiuc C M, Srivastava D, et al. Differentially private spatial decompositions//Proceedings of the ICDE. Washington, USA, 2012: 20-31

[42] Qardaji W H, Yang Weining, Li Ninghui. Differentially private grids for geospatial data//Proceedings of the ICDE. Brisbane, Australia, 2013: 757-768

[43] McSherry F, Mironov I. Differentially private recommender systems: Building privacy into the netfix prize contenders// Proceedings of the KDD. Paris, France, 2009: 627-636

[44] Kapadia A, Triandopoulos N, Comelius C, et al. AnonySense: Opportunistic and privacy-preserving context collection// Proceedings of the Pervasive. Sydney, Australia, 2008: 280-297

[45] Domingo-Ferrer J, Mateo-Sanz J M. Practical data-oriented Microaggregation for statistical disclosure control. IEEE Transactions on Knowledge and Data Engineering, 2002, 14(1): 189-201

[46] Huang Kuanlun, Kanhere S S, Hu Wen. Towards privacy-sensitive participatory sensing//Proceedings of the PerCom Workshop. Galveston, Texas, USA, 2009: 1-6

[47] Huang Kuanlun, Kanhere S S, Hu Wen. Preserving privacy in participatory sensing systems. Computer Communications, 2010, 33(11): 1266-1280

[48] Kazemi L, Shahabi C. Towards preserving privacy in participatory sensing//Proceedings of the PerCom Workshops. Seattle, USA, 2011: 328-331

[49] Gao Sheng, Ma Jianfeng, Shi Weisong, Zhan Guoxing. Towards location and trajectory privacy protection in participatory sensing//Proceedings of the MobiCASE. Los Angeles, USA, 2011: 381-389

[50] Vu Khuong, Zheng Rong, Gao Jie. Efficient algorithms for *K*-anonymous location privacy in participatory sensing// Proceedings of the INFOCOM. Orlando, FL, USA, 2012: 2399-2407

[51] Agir B, Papaioannou T G, Narendula R, et al. User-side adaptive of location privacy in participatory sensing. GeoInformatica, 2014, 18(1): 165-191

[52] Vergara-Laurens I J, Mendez D, Labrador M A. Privacy, quality of information, and energy consumption in participatory sensing systems//Proceedings of the PerCom. Budapest, Hungary, 2014: 199-207

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

[54] Meiklejohn S, Mowery K, Checkoway S, Shacham H. The phantom tollbooth: Privacy-preserving electronic toll collection in the presence of driver collusion//Proceedings of the USENIX Security Symposium. San Francisco, USA, 2011: 1-20

[55] Zeng Bing, Yang Yu, Cao Yunfei. Study on identity-based encryption technology. Academic Research, 2011: 64-67(in Chinese)
(曾兵，杨宇，曹云飞. 基于身份加密(IBE)技术研究. 学术研究，2011：64-67)

[56] Gunther F, Manulis M, Peter A. Privacy-enhanced participatory sensing with collusion resistance and data aggregation //Proceedings of the CANS. Heraklion, Crete, Greece, 2014: 321-336

[57] Paillier P. Pubilc-Key cryptosystems based on composite degree residuosity classes//Proceedings of the EUROCRYPT. Prague, Czech Republic, 1999: 223-238

[58] Shu Tao, Chen Yingying, Yang Jie, Williams A. Multi-lateral privacy-preserving localization in pervasive environment// Proceedings of the INFOCOM. Toronto, Canada, 2014: 2319-2327

[59] Sun Li-Min, Li Hong, Wang Xiao-Han, He Yun-Hua. Survey on the location privacy preservation in the Internet of Things. Journal of Software, 2014, 25(1): 1-10(in Chinese) (孙利民, 李红, 王笑寒, 何云华. 物联网位置隐私保护综述. 软件学报, 2014, 25(1): 1-10)

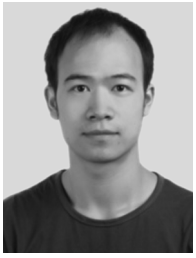
[60] Shao Jun, Lu Rongxing, Lin Xiaodong. FINE: A fine-grained privacy-preserving location-based service framework for mobile devices//Proceedings of the INFOCOM. Toronto, Canada, 2014: 244-252

[61] Marusic S, Gubbi J, Sullivan H T, et al. Participatory sensing, privacy, and trust management for interactive local government. IEEE Technology and Society Magazine, 2014, 33(3): 62-70

[62] Sheikh S B, Deshmukh P. Review on enabling enhanced privacy in participatory sensing. IOSR-JCE, 2014: 13-17

[63] Ahmadi H, Pham N, Ganti R K, et al. Privacy-aware regression modeling of participatory sensing data//Proceedings of the SenSys. Zurich, Switzerland, 2010: 99-112

[64] Amintoosi H, Kanhere S S. A trust-based recruitment framework for multi-hop social participatory sensing//Proceedings of the DCOSS. Cambridge, USA, 2013: 266-273



ZENG Ju-Ru, born in 1987, Ph. D. candidate. His research interests include wireless sensor network, privacy preservation & participatory sensing.

CHEN Hong, born in 1965, Ph. D., professor, Ph. D. supervisor. Her research interests include database, data warehouse and wireless sensor network.

PENG Hui, born in 1986, Ph. D. candidate. His research interests include wireless sensor network, privacy preservation.

WU Yao, born in 1990, Ph. D. candidate. His research interests include wireless sensor network, privacy preservation and crowd sensing.

LI Cui-Ping, born in 1971, Ph. D., professor, Ph. D. supervisor. Her research interests include data warehouse and data mining, information network analysis and flow data management.

WANG Shan, born in 1944, M. S., professor, Ph. D. supervisor. Her research interests include the new technology of high-performance database, data warehouse and business intelligence.

Background

Participatory sensing (PS) has very wide application prospects including real-time weather forecast, health monitoring, traffic surveillance, etc. As the personal information associated to the users, the strong focus on the collection of sensor data may compromise user privacy in different ways. For example, by tracking a user's current location, an adversary can further utilize the leakage of user's location and judge the user is at home or not. Hence, privacy preservation is needed in PS and has a critical effect on the widespread deployment of this network.

There are three inherent characteristics of PS, which bring big challenge for data privacy preservation and location privacy preservation in PS. Firstly, the PS data are not stability due to users with a high degree of mobility and flexibility. Privacy policy must respond to these dynamic changes. Secondly, users can gain many kinds of data and affix different semantic tags on them, making the PS data dimensions become larger. Thirdly, users often occur in an unsupervised environment which leads it's difficult to monitor user's activities.

Over the recent years, the development of PS has attracted much more attentions. Extensive researches focused on enhancing privacy preservation and improving the quality of service and resource efficiency. This paper mainly surveys the previous works and gives some indexes for researchers to this interesting and important issue.

This research is supported by the Natural Basic Research Program (973 Program) of China (No. 2012CB316205), the National High Technology Research and Development Program (863 Program) of China (No. 2014AA015204) and the National Natural Science Foundation of China (Nos. 61070056, 61033010, 61272137, 61202114).