计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2020.20190579
57(11): 2337-2347, 2020

移动群智感知中融合数据的隐私保护方法

王涛春　金　鑫　吕成梅　陈付龙　赵传信
(安徽师范大学计算机与信息学院　安徽芜湖　241002)
(网络与信息安全安徽省重点实验室(安徽师范大学)　安徽芜湖　241002)
(wangtc@nuaa.edu.cn)

Privacy Preservation Method of Data Aggregation in Mobile Crowd Sensing

Wang Taochun, Jin Xin, Lü Chengmei, Chen Fulong, and Zhao Chuanxin
(*School of Computer and Information, Anhui Normal University, Wuhu, Anhui* 241002)
(*Anhui Provincial Key Laboratory of Network and Information Security (Anhui Normal University), Wuhu, Anhui* 241002)

Abstract Serious privacy leakage problems are on the rise with the wide application of mobile crowd sensing owing to the popularity of mobile smart devices. In general, the existing privacy protection schemes assume that the third-party service platform is credible, which therefore sets a high requirement on the application context. Based on this, the paper proposes a new privacy preservation data aggregation algorithm based on elliptic curve cryptography (ECPPDA) in mobile crowd sensing. The server randomly divides the participants into g clusters and forms respective cluster public key for each cluster. The nodes in the cluster encrypt the data through their own cluster public keys and merge the data aggregation results. The server obtains the aggregation result by cooperating with the members in the cluster. Since what the server receives is the ciphertext of aggregation and the ciphertext decryption requires all the nodes in the cluster to cooperate together, the server cannot obtain the data of a single participant. In addition, the updating of the cluster public key by the server can facilitate the participants to dynamically join or leave. The experimental result shows that ECPPDA has the characteristics of high security, low consumption, low communication and high precision.

Key words mobile crowd sensing; aggregation data; privacy preservation; collusion attack; cluster

摘　要　随着移动智能设备的普及,群智感知得到广泛应用,也面临严重的隐私泄露问题.现有隐私保护方案一般假设第三方服务平台是可信的,而这种假设对应用场景要求较高.基于此,提出了群智感知中一种新的数据融合隐私保护算法 ECPPDA(privacy preservation data aggregation algorithm based on elliptic curve cryptography).服务器将参与者随机划分成 g 个簇,并形成簇公钥.簇内节点通过簇公钥加密数据并融合得到簇融合结果数据.服务器通过与簇内成员协同合作得到融合结果原文,由于服务器接收到的是融合密文且密文解密需要簇内所有节点共同协作,因此服务器不能得到单个参与者的数据.

收稿日期:2019-08-23;**修回日期**:2019-11-25
基金项目:国家自然科学基金项目(61402014,61972439,61972438,61871412);赛尔网络下一代互联网创新项目(NGII20170312);安徽省教育厅高校自然科学研究重点项目(KJ2019A1164);安徽师范大学博士启动项目(2018XJJ66);安徽师范大学创新项目(2018XJJ114)
This work was supported by the National Natural Science Foundation of China (61402014, 61972439, 61972438, 61871412), the CERNET Next Generation Internet Creative Project of China (NGII20170312), the Key Program of Universities Natural Science Research of the Anhui Provincial Department of Education (KJ2019A1164), the Anhui Normal University PhD Startup Fund (2018XJJ66), and the Anhui Normal University Innovation Fund (2018XJJ114).

此外,通过服务器对簇公钥的更新,能够方便参与者动态加入或失效.实验结果显示 ECPPDA 具有高安全性、低消耗、低通信、高精度的特点.

关键词 移动群智感知;融合数据;隐私保护;共谋攻击;簇

中图法分类号 TP391

随着集成多种传感器(温度传感器、GPS、加速度传感器、相机等)的移动智能设备(例如智能手机、智能手环等)的大量普及,一种新的传感范式群智感知逐渐兴起.群智感知[1]是一种利用便携式移动智能设备携带的传感器收集数据,然后将数据上传给群智感知服务商,获得一定的奖励或应用.随着移动智能设备集成的传感器越来越多,群智感知收集的数据类型也越来越广,使得应用领域更加广泛.服务提供商利用参与者收集大量的信息向需求者提供服务,需求者利用这些信息进行分析解决问题[2-6],例如交通导航、环境监测[7]、社区服务[8]等.

然而群智感知的应用也引出一系列隐私安全问题.例如攻击者通过获取参与者发送的医疗等敏感信息来推导参与者的健康状况,从而进行一系列恶意攻击.同时,参与者传输给应用服务器的数据带有时空信息(参与者收集数据时的位置),攻击者可能利用这些时空信息推导出参与者的生活习惯、行为规律等敏感信息来进行恶意攻击.因此,确保群智感知中参与者敏感信息的隐私是推动群智感知应用的关键因素.数据融合是群智感知数据收集的主要操作方式之一,目前群智感知中融合数据的隐私保护前提是服务提供商平台是可信的,文献[9]提出了一种基于同态加密的数据融合方案,该方案可以对密文进行融合操作,能够有效进行各种数据融合(如均值、方差、偏差等),然而第三方服务提供商能够获得每个节点的信息,所以该方案的应用前提是第三方是可信的.此外,现有的群智感知中隐私保护数据融合方法没有考虑到节点(移动智能设备)移动性的特点,文献[10]提出了一种基于 Diffie-Hellman 的加密方案,当节点加入或是失效时,需要更新所有节点的加密密钥,节点的动态加入或失效处理复杂困难.

基于此,提出了群智感知中基于椭圆曲线的数据融合隐私保护算法(privacy preservation data aggregation algorithm based on elliptic curve cryptography, ECPPDA),服务器对所有节点进行随机分簇,簇内节点通过自身的公钥构建簇公钥,并利用簇公钥对簇内节点采集到的数据进行加密,簇内节点对密文数据进行融合,最后将融合结果传输给应用服务器.应用服务器通过与簇内节点协作完成数据解密得到数据融合结果.由于应用服务器不能直接解密密文,从而保证节点数据的隐私性,解决了第三方必须是可信的应用场景.同时,ECPPDA 算法中,节点的加入或失效只需要更新簇内节点的少量数据信息,因此本算法对节点动态加入或失效处理简单,较适用于群智感知这类节点具有移动性特点的分布式网络.理论分析和实验结果表明 ECPPDA 不仅能够保护用户的数据隐私抵御共谋攻击且兼顾节点移动性,而且能有效地降低节点的通信开销.

1 相关工作

根据服务器收集的数据类型,群智感知数据收集分为 2 类:1)原始数据收集,移动智能设备直接上传采集到的数据(如 GPS 坐标、加速度读数等).2)融合数据,服务器需要得到区域内节点数据的融合结果(求和、平均值、方差等).例如,服务器通过获得参与者平均运动量(步数、运动传感器)能够推导出参与者普遍的公共健康状况.本文将这 2 类分别称为基于原始数据的收集和基于融合数据的收集.当前国内外学者对群智感知中隐私保护问题进行了广泛研究.现有方案主要利用 4 类方法[11]:分组统计、可信第三方验证、k-匿名和数字加密.

基于分组统计的技术思想是参与者在上传感知数据阶段,将数据进行切片分成多个数据片,然后将数据片转发给邻节点,之后邻节点上传数据,应用服务器收集数据片并进行融合从而得到原始数据.文献[12]提出了 HP^3(hot-potato-privacy-protection algorithm)方案,在服务器不可信的情况下实现保护隐私的分组转发.HP^3 方案中参与者不是直接将数据上传到服务器,而是将数据传输给参与者的一个朋友(可信),朋友选择另一个朋友并传输数据,以此类推,直到到达参与者定义的阈值,然后最后一个用户将数据上传到服务器.这种方案虽然解决了在服务服不可信情况下保证参与者的位置隐私,但前提是参与者有可信朋友.文献[13]提出了一种基于编码 k-匿名方案(coding-based privacy preserving

scheme)SLICER,SLICER 集成了数据编码技术和消息传输策略,以实现对参与者隐私的保护,同时保持高数据质量.SLICER 方案没有考虑共谋攻击.文献[14]提出了 PEPPeR(privacy enhancing protocol for participatory sensing)方案,PEPPeR 能够保护查询者的数据隐私,查询者和参与者由可信第三方服务器分配令牌,查询者和参与者传输数据时需要经过它们和第三方服务器之间的 Tor(the onion router)网络.这些方案的隐私保护全都需要可信的第三方,这种第三方更容易被攻击.k-匿名是指区域内 k 个参与者和它们发出的信息不能被攻击者区分出来.文献[15]提出了 AnonySense 方案,该方案通过 k-匿名技术来保护参与者的位置隐私.文献[16]的作者利用可信的邻居构建多个 k-匿名区域发布虚假查询,来防止攻击者构建真实轨迹.数字加密是利用密码学的知识对数据进行加密防止攻击者窃听参与者在传输时的数据内容.这 4 种技术常用来组合使用保护参与者的数据隐私.

针对群智感知中数据融合的隐私保护问题,近年来国内外学者已经提出了许多解决方案.文献[17]提出的方案先利用多假名机制来克服由于参与者密度低而导致的漏洞,接着提出 2 种基于 Paillier 密码系统的方法抵御 sybil 攻击.文献[10]提出了一种基于 Diffie-Hellman 的加密方案,但这些方案无法有效地支持动态连接和失效.以文献[10]中的工作为例,当节点加入或失效时,需要更新所有节点的加密密钥,这意味着群智感知应用程序具有高额的通信开销.文献[8]提出了一种基于同态加密的数据融合方案,该方案对密文能够有效地进行各种数据融合(如均值、方差、偏差等),然而第三方服务提供商能够获得每个节点的信息,所以该方案应用前提是具有可信的第三方.文献[18]中提出了不需要可信第三方来保护融合数据隐私的协议,该协议通过节点间参数共享来抵御共谋攻击.文献[19]提出了 2 种具有隐私保护能力的加法融合数据方案:第 1 种方案是基于簇的隐私数据融合方案(cluster-based private data aggregation, CPDA),利用多项式的聚类协议和代数属性完成隐私保护的数据融合;第 2 种方案是基于切片-混合的数据融合方案(slice-mix-aggregate, SMART),基于切片技术和加法关联属性.SMART 方案的优点是计算开销相对较少,但 SMART 对每个数据都分成多段,因此节点的通信开销较大.文献[20]提出了一种基于信任的数据融合方案 ERTDA(energy-efficient reliable trust-based data aggregation),ERTDA 通过节点行为来计算和评估节点的信任值,并能够及时检测和排除受损节点.ERTDA 能够有效地提高融合数据的精确度,降低节点能耗,提高数据传输的可靠性,但节点信任值的评估以及受损节点的检测需要大量的计算开销,且 ERTDA 基于可信节点进行操作.

2 模型介绍

2.1 群智感知网络框架

群智感知网络主要由 4 个部分组成:移动节点、数据请求者、应用服务器和接入网络,如图 1 所示:

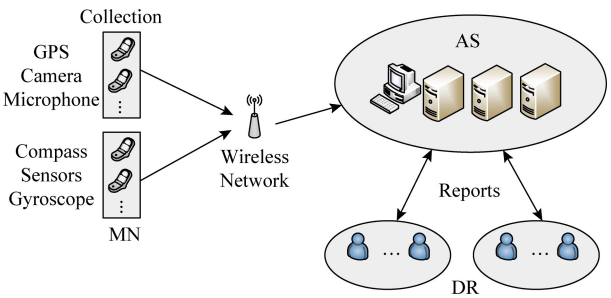

Fig. 1 Basic framework of the crowd sensing application

图 1 群智感知应用基本框架

1) 移动节点(mobile node, MN).MN 是具有感知、计算、存储、通信等基本功能的移动智能设备,如智能手机、运动手环等.它们是群智感知网络的基本单元,主要用于各类数据的采集,如温度、湿度、位置、声音等,并通过无线网络将采集来的数据上传到应用服务器.这些移动智能设备常由用户随身携带,所以能够随时随地进行数据采集,在群智感知中也常被称作参与者.

2) 数据请求者(data requester, DR).DR 通过应用服务器发布感知任务,然后应用服务器将任务下达给参与者,服务器收集到请求者所需的数据时,将数据传输给请求者.在群智感知系统中参与者也可以通过他们的移动智能设备向服务器发布任务,所以参与者也可以是数据请求者.

3) 应用服务器(application server, AS).AS 是对节点上传的数据进行处理,如解密、分类、融合、存储等,并与数据请求者实现数据的共享.根据移动节点提供的数据来回答数据请求者的各种问题,如请求者查找距离居住宾馆最近的餐厅等.本文应用场景是最常见的半诚实模型,即参与者和服务器严格按照协议要求执行操作,另一方面,他们通过获取信息推导出其他参与者的隐私信息.

4)接入网络.指数据传输的通道,通常是通过无线网络进行数据传输的例如移动蜂窝网络、蓝牙、WIFI等无线网络,目前感知网络主要利用移动蜂窝网络进行数据传输.

2.2 攻击模型

根据攻击来源可将攻击划分为2类:外部攻击和内部攻击.1)外部攻击指的是攻击者通过窃听群智感知网络获取感知数据,破坏数据机密性.2)内部攻击指的是群智感知网络的参与者或服务提供商进行的攻击,例如参与者获取其他参与者的隐私信息,或服务提供商获取参与者的隐私数据,并通过这些隐私信息进行恶意攻击.本文主要采用的是半诚实模型,即群智感知中所有成员(服务提供商和参与者)严格执行设计的协议,但成员试图通过协议执行过程获得的数据来推导出其他成员的敏感信息.通过对传输数据加密能够有效抵抗外部攻击,而内部攻击以及共谋(多个参与者或服务器与参与者)攻击是ECPPDA研究的重点.

3 ECPPDA算法

本节主要介绍ECPPDA给出问题定义、算法思想,以及算法的具体执行步骤,最后对参与者的动态加入和失效处理过程进行描述.

3.1 问题定义及算法思想

ECPPDA假设有 m 个参与者和1个应用服务器,参与者采集感知数据、加密并融合,应用服务器得到融合后的密文数据,参与者 n_i 采集的感知数据为 $d_i(i=1,2,\cdots,m)$.

ECPPDA中,服务器将网络中所有节点随机分成 $g=\lceil m/k\rceil$ 个簇,每簇包含 k 个节点,并将分簇信息和簇公钥发送给每个节点.节点利用簇公钥对感知数据进行加密,簇内节点融合簇内密文后再传输给服务器,服务器需要与每个簇内所有节点协作才能解密获得簇内融合数据的明文,从而得到区域内所有节点的感知数据总和.

3.2 ECPPDA算法步骤

ECPPDA融合参与者采集到的感知数据,其处理主要包括3个阶段:

1)初始化.服务器随机将所有节点划分成 g 个簇,再对每个簇内所有节点的公钥进行融合形成簇公钥 $P_i^{CK}(i=1,2,\cdots,g)$,最后服务器将簇公钥和簇信息发分别发送给对应簇内的所有节点.

2)数据传输.簇头节点随机选择簇内某个节点作为簇头节点,簇头节点随机选择一个始节点,始节点通过簇公钥加密自身的感知数据并将密文传输给中间节点,中间节点接收到前一个节点发送的密文并与自己的密文融合,再将融合结果发送给下一个中间节点,以此继续,直至终节点得到簇内所有节点密文的融合结果并发送给服务器.

3)解密.服务器接收每个簇的密文融合结果后,需分别与每个簇内节点协同合作计算得到簇融合结果的明文,最后服务器对所有的簇明文进行融合从而得到所需数据.

3.2.1 初始化

现有的相关算法没有对节点进行分簇处理如文献[9],这类算法的缺点是缺少动态处理机制,当节点动态加入或失效时会导致大量的数据更新.由于在群智感知中节点具有移动性,节点可能动态地加入或离开感知任务,因此ECPPDA对所有参与节点进行分簇,当节点动态加入或失效时只需要更新少量的数据.服务器随机将所有参与节点划分成 $g=\lceil m/k\rceil$ 个簇,当节点动态地加入或失效时只需要更新失效节点所在簇的簇内节点的簇公钥.为防止簇内只有一个诚实节点的极端情况或者簇内节点数量过多导致节点动态加入/离开需要进行大量数据的更新,ECPPDA初始每个簇的节点数为 k,其计算式为 $k=\gamma\times m+2$,其中 γ 是恶意节点的概率,$\gamma\times m$ 为恶意节点数.所以能够保证每个簇内至少有2个诚实节点,从而避免出现簇内除目标节点之外其他节点都为恶意节点的情况.

簇划分具体过程为:在任务区域内接受任务节点 n_i 发送公钥 Y_i 给服务器,服务器随机选择 k 个节点形成一个簇,从而将任务区域划分为 g 个簇,并给每个簇随机选择一个节点作为簇头节点.对于每个簇,服务器对簇内 k 个节点的公钥进行融合形成簇公钥,然后将簇公钥和簇信息(包括簇头节点和簇成员信息)发送给簇内的所有节点.其操作包括3个步骤:

步骤1. 节点 $n_i(i=1,2,\cdots,n)$ 选择一个随机数 $x_i\in\mathbb{Z}_q^*$ 作为自身节点私钥,然后计算公钥 $Y_i=x_i\times G$ 并发送给服务器.

步骤2. 服务器将参与节点随机划分成 $g=\lceil m/k\rceil$ 个簇,服务器接收到所有节点发送过来的公钥按簇融合得到每个簇 C_i 的簇公钥 $P_i^{CK}=\sum_{j=1}^{k}Y_{s_j}(n_{s_j}\in C_i)$.

步骤 3. 服务器分别将簇公钥 P_i^{CK} 和分簇信息包括簇头信息发送给所有节点.

具体过程见算法 1.

算法 1. 节点分簇.

① for($i=1;i\leqslant n;i++$) /* 节点 n_i 操作 */

② $x_i\in\mathbb{Z}_q^*$; /* 生产随机数 x_i */

③ $Y_i=x_i\times G$;

④ $Y_i\rightarrow N_{AS}$; /* 将公钥 Y_i 发送给应用服务器 N_{AS} */

⑤ end for

⑥ for($i=1;i\leqslant g;i++$) /* 服务器操作 */

⑦ 簇 $C_i=\{n_1,n_2,\cdots,n_k\}$; /* 随机选择 k 个节点构成簇 C_i */

⑧ 簇 $C_i:P_i^{CK}=\sum_{j=1}^{k}Y_{s_j}(n_{s_j}\in C_i)$; /* 服务器计算簇 C_i 的簇公钥 P_i^{CK} */

⑨ 随机选取簇头节点 n_h;

⑩ 簇 $C_i:D_{C_i}\rightarrow$ 节点 $n_s,n_s\in C_i$; /* 包括簇公钥和簇头等分簇信息 D_{C_i} 发送给簇内所有成员节点 */

⑪ end for

3.2.2 数据传输

每个簇的操作相同,现以一个簇操作为例介绍数据传输进程.节点 n_i 通过簇公钥 P^{CK} 加密它的数据 d_i,簇头节点随机选择一个节点作为始节点(密文发送的初始节点),始节点将密文发送给随机选择的下一个节点(中间节点),中间节点接收始节点的密文并与自身的密文进行融合,再发送给下一个中间节点,如此继续,直至簇内所有节点完成密文融合,最后一个节点(终结点)把最终的融合结果发送给服务器.

节点 n_i 的感知数据为 d_i,椭圆曲线参数为 $E(F_p)$,阶数为 q,G 是基点.节点 n_i 选择一个随机数 x_i 作为节点的私钥,然后计算节点 n_i 的公钥 $Y_i=x_i\times G$.$E(F_p)$,q,G 是群智感知中公共参数,并广播给簇内所有成员.

步骤 1. 每个节点 n_i 选择一个随机数 $r_i\in\mathbb{Z}_q^*$,计算密文 (C_a^i,C_b^i),其中 $C_a^i=r_i\times G$,$C_b^i=d_i\times G+r_i\times P^{CK}$.

步骤 2. 簇头随机选择一个成员节点 $n_i(i=1,2,\cdots,k)$ 为始节点,并通知节点 n_i.以此,可以将节点分为始节点、中间节点和终节点.

① 始节点.节点 n_i 随机选取下一个节点 n_j(中间节点),并给其发送密文 (C_a^i,C_b^i).

② 中间节点.节点 n_j 接收上一个节点 n_{j-1} 发送的密文信息 $(C_a^{agg,j-1},C_b^{agg,j-1})$(前 $j-1$ 个节点密文之和),并与自身密文数据 (C_a^j,C_b^j) 进行融合得到 $(C_a^{agg,j},C_b^{agg,j})$,再随机传输给下一个节点 n_{j+1},如此继续,直到终节点.

步骤 3. 终节点 n_{end} 接收到上一个节点融合的密文并与自身密文数据 (C_a^{end},C_b^{end}) 进行融合得到簇内最终的融合密文 (C_a,C_b) 并发送给服务器,其中,$C_a=\sum_{i=1}^{k}C_a^i$,$C_b=\sum_{i=1}^{k}C_b^i$.

3.2.3 解密

服务器接收到簇 C_i 发送过来的密文数据后,首先与簇成员协同合作计算 D_i,进而计算得到簇明文的融合结果 Sum_j,最后对所有簇数据进行融合得到任务区域内所有节点的数据融合结果 Sum.每个簇 C_i 的密文解密步骤为:

步骤 1. 服务器把簇融合密文中的 C_a 广播给簇内所有成员节点 $n_i(i=1,2,\cdots,k)$.

步骤 2. 节点 n_i 计算 $D_i=x_i\times C_a$ 并以密文传输的方式将 D_i 传输给服务器.

步骤 3. 服务器通过接收到的 $\sum_{i=1}^{k}D_i$ 并利用式(1)进行解密,假设所有感知数据符合 $d_i\in[0,L]$,所以每个簇融合数据 $Sum_j(j=1,2,\cdots,g)$ 的范围是 $[0,kL]$,因为 $kL\ll q$,所以 Sum 可以采用复杂度为 $O(\sqrt{kL})$ 的 Pollard's lambda 算法[21]求解.

$$Sum_j=\log_G(C_b-\sum_{i=1}^{k}D_i). \tag{1}$$

步骤 4. 服务器将所有簇的数据 Sum_j 进行融合得到任务区域内所有节点数据的融合结果 Sum,$Sum=\sum_{j=1}^{g}Sum_j$.

式(1)的正确性证明:

$$Sum_j=\log_G(C_b-\sum_{i=1}^{k}D_i)=$$

$$\log_G(C_b-\sum_{i=1}^{k}x_i\times C_a)=$$

$$\log_G(\sum_{i=1}^{k}C_b^i-\sum_{i=1}^{k}x_i\times C_a)=$$

$$\log_G(\sum_{i=1}^{k}(d_i\times G+r_i\times P^{CK})-$$

$$\sum_{i=1}^{k}x_i\times\sum_{i=1}^{k}r_i\times G)=$$

$$\log_G(Sum_j\times G+r\times P^{CK}-\sum_{i=1}^{k}x_i\times r\times G)=$$

$$\log_G(Sum_j \times G + r \times P^{CK} - \sum_{i=1}^{k} x_i \times r \times G) =$$

$$\log_G(Sum_j \times G + r \times \sum_{i=1}^{k} Y_i - \sum_{i=1}^{k} x_i \times r \times G) =$$

$$\log_G(Sum_j \times G + r \times \sum_{i=1}^{k} x_i \times G -$$

$$r \times \sum_{i=1}^{k} x_i \times G) = \log_G(Sum_j \times G). \quad \text{证毕.}$$

由公式 $Sum_j = \log_G\left(C_b - \sum_{i=1}^{k} D_i\right)$ 推导至公式 $Sum_j = \log_G(Sum_j \times G)$过程可知式(1)是正确的.

群智感知中所有数据通过无线传输,且参与节点具有动态性特征,数据传输过程中出现数据丢失是普遍情况.ECPPDA 采用简化的超时重传机制来处理传输过程中数据丢失情况.重传机制的主要思想为:数据传输过程中,当数据接收者收到数据时要立刻返回一个确认信息给数据发送者,如果数据发送者的重传超过时间(retransmission time out, RTO)T,即在时间 T 内没有收到数据接收者的确认信息则默认为数据丢失,再次发送一份相同的数据.其中 $T=2T_R$,T_R 为数据在 2 个端点平均往返时间.简化的超时重传机制能够弱化网络传输中的各种复杂问题.

3.3 节点加入和节点失效

群智感知是一种特殊的无线传感器网络,它的节点由移动智能设备构成,而设备持有人由于自身原因离开任务区域或关闭移动智能设备,所以与传统的无线传感器网络相比,群智感知节点具有更显著的移动性特征,而现有的方法很少考虑群智感知节点移动性特征.基于此,ECPPDA 考虑节点移动性特征,更加方便快捷处理节点的加入和节点失效情况.

3.3.1 节点加入

节点加入有 2 种情形:1)加入节点的数量 $b<\gamma\times n+2$(簇内节点数下限),节点 n_i 向服务器发送申请加入感知任务信息,服务器接收到信息后查询成员节点数量最少的簇 C_i,判断加入节点后簇内节点数量是否超出上限.如果没有服务器将节点 n_i 加入该簇,更新该簇的簇公钥,即 $P_i^{CK}=P_i^{CK}+\sum_{j=1}^{b}Y_{s_j}$,其中 P^{CK}为原簇公钥和 Y_{s_j} 为加入节点公钥,并将新的簇公钥传输给簇内所有成员节点(包括新加入节点 n_i).如果节点加入后簇内节点数量超出上限,服务器将从簇 C_i 中随机选取 $k-b$ 个节点与 b 个新加入的节点建立新簇 C_j 并将簇公钥与簇信息传输给簇 C_j 内所有节点.其次更新簇 C_i 内剩余节点的簇公钥 $P_i^{CK}=P_i^{CK}-\sum_{j=1}^{k-b}Y_{s_j}$ 和簇信息.2)加入节点的数量 $b>\gamma\times n+2$,直接生成 $g=\lfloor b/(\gamma\times n+2)\rfloor$ 个新簇.剩余不满足下限数量的节点按照情形 1 的方式加入,具体见算法 2.

算法 2. 节点加入.

① 输入 b; /* 其中 b 是加入任务的节点数 */
② for ($i=1$;$i<b$;$i++$)
③ $x_i\in\mathbb{Z}_q^*$; /* 产生随机数 x_i */
④ $Y_i=x_i\times G$; /* 计算节点 n_i 公钥 */
⑤ $Y_i\rightarrow N_{AS}$;
/* 将公钥 Y_i 发送给服务器 N_{AS} */
⑥ end for
⑦ if ($b<\gamma\times n+2$) /* 判断加入节点数是否超过下限 */
⑧ 查询并选择节点数量最少的簇 C_i;
⑨ if ($b+|C_i|<2\times\gamma\times n+4$) /* $|C_i|$为簇 C_i 内节点数量 */
⑩ 簇C_i:$P_i^{CK}=P_i^{CK}+\sum_{j=1}^{b}Y_{s_j}$; /* 计算簇 C_i 的新公钥 */
⑪ $P_i^{CK}\rightarrow$节点 n_s,$n_s\in C_i$; /* P_i^{CK} 发送给簇内成员节点 */
⑫ else
⑬ 从簇 C_i 随机选取 $k-b$ 个节点与 b 个新节点建立新簇 C_j;
⑭ $P_j^{CK}=\sum_{n_s\in C_j}Y_s$; /* 计算簇 C_j 公钥 */
⑮ 随机选取簇头 n_j;
⑯ 簇 C_j:$D_{C_j}\rightarrow$节点 n_s,$n_s\in C_j$; /* 分簇信息 D_{C_j} 发送给簇内所有成员节点 */
⑰ 簇C_i:$P_i^{CK}=P_i^{CK}-\sum_{j=1}^{k-b}Y_{s_j}$; /* 修改簇 C_i 公钥 */
⑱ 随机选取簇头 n_i;
⑲ 簇C_i:$D_{C_i}\rightarrow$节点 n_s,$n_s\in C_i$; /* 修改后的簇 C_i 的分簇信息 D_{C_i} 发送给簇内所有成员节点 */
⑳ end if
㉑ end if
㉒ if ($b\geqslant\gamma\times n+2$) /* 假设簇节点数目 $k=\gamma\times n+2$ */
㉓ $g_1=\lfloor b/(\gamma\times n+2)\rfloor$;

㉔ for $i=1;i\leqslant g_1;i++$)

㉕ 簇$C_i=\{n_{s_1},n_{s_2},\cdots,n_{s_k}\}$; /* 随机选择 k 个节点划入簇 C_i 内 */

㉖ 簇$C_i:P_i^{CK}=\sum_{j=1}^{k}Y_{s_j}$ $(n_{s_j}\in C_i)$;

㉗ 随机选取簇头节点 $n_s,n_s\in C_i$;

㉘ 簇$C_i:D_{C_i}\rightarrow$节点 $n_s,n_s\in C_i$; /* 簇 C_i 的分簇信息 D_{C_i} 发送给簇内所有成员节点 */

㉙ end for

㉚ 查询并选择节点数量最少的簇 C_i;

㉛ 簇$C_i:P_i^{CK}=P_i^{CK}+\sum_{j=1}^{b-g_1\times k}Y_{s_j}$; /* 将剩余节点加入到簇 C_i 中 */

㉜ 簇$C_i:D_{C_i}\rightarrow$节点 $n_s,n_s\in C_i$;/* 分簇信息 D_{C_i} 发送给簇内所有成员节点 */

㉝ end if

3.3.2 节点失效

簇 C_i 内节点 n_i 失效有 2 种情形:

1) 主动失效即节点 n_i 失效前向服务器发送离开信息.节点 n_i 发送离开信息给服务器申请离开簇,服务器接收信息后查看簇 C_i 内剩余成员节点的数量,当剩余成员节点数量大于 $\gamma\times n+2$ 时,服务器计算新的簇公钥 $P^{CK}=P^{CK}-Y_i$,并随机选取簇头,更新簇 C_i 内剩余成员节点的簇信息.当剩余成员节点数量小于 $\gamma\times n+2$ 时,服务器解散簇 C_i,然后将该簇剩余节点按照流程加入到其他簇.

2) 被动失效节点 n_i 没有发送任何信息就离开簇.簇 C_i 的节点 n_i 被动失效有 2 种假设:①假设节点 n_i 在接收数据前失效,节点 n_j 传输数据给节点 n_i,节点 n_i 已经失效无法返回确认信息给节点 n_j.由于超时重传机制,节点 n_j 将不停地传输数据直到重传次数达到设置的阈值(超时重传机制中最大重传次数)时节点 n_i 没有发送返回信息则认为节点 n_i 失效,然后节点 n_j 向服务器发送信息表示节点 n_i 失效,服务器接收信息后按照节点失效情形 1)的方式更新簇信息.②假设节点 n_i 在接收信息后失效,节点 n_i 失效无法继续传输数据,因此服务器不能接收簇 C_i 的数据,服务器对簇 C_i 内节点进行失效查询(服务器向簇 C_i 内所有节点发送信息,没有响应的节点被确认为失效节点),确定失效节点 n_i 后服务器按照节点失效情形 1 的方式进行处理.

4 性能分析

本节 ECPPDA 算法与现有算法在应用环境、算法的安全性和复杂度等方面进行分析.

4.1 应用环境

应用环境主要从第三方是否确定可信、能否抗共谋攻击、节点动态性、数据丢失处理以及数据实用性等方面进行对比.从表 1 可以看出,与 PDAIF[22](private data aggregation with integrity assurance and fault tolerance)相比,ECPPDA 可以抵御共谋攻击且不需要可信的第三方平台;与 CTPPSP[18](collusion-tolerable privacy-preserving sum and product calculation)相比,ECPPDA 对数据丢失进行了处理,提高了方案的健壮性,且 ECPPDA 没有降低数据实用性且保留了节点移动性的特征.

Table 1 Comparison in Advantages and Disadvantages of Existing Schemes

表 1 现有方案优缺点对比

Properties	PDAIF[22]	CTPPSP[18]	ECPPDA
Trusted Third Party	Yes	No	No
Anti-collusion Attack	No	Yes	Yes
Node Dynamics	Yes	No	Yes
Data Loss Processing	Yes	No	Yes
Data Availability	Low	High	High

4.2 安全性分析

本节对外部攻击、内部攻击和共谋攻击 3 方面进行安全性分析.

1) 外部攻击

所有数据通过无线网络进行传输,外部攻击者通过窃听获取感知数据是一种最常见的攻击方式,本文假设攻击者能够进行全网窃听.由于 ECPPDA 对传输感知数据进行加密,外部攻击者窃听到的数据为加密后的密文.密文数据为(C_a^i,C_b^i),其中 $C_b^i=d_i\times G+r_i\times P^{CK}$,攻击者必须计算出 $r_i\times P^{CK}=x_1\times C_a^i+x_2\times C_a^i+\cdots+x_i\times C_a^i$ 才能得到明文 d_i.由于攻击者缺少私钥$(x_1,x_2,\cdots,x_i)$,所以无法通过公共参数$(G,C_a,Y_1,\cdots,Y_i)$得出式子 $r_i\times P^{CK}$ 的计算结果[23],因此外部攻击者窃听到密文(C_a^i,C_b^i)后不能计算出对应的明文 d_i.此外,由于节点传输的数据为融合后的密文,所以攻击者还不能确定所得密文是由哪些节点融合的.因此,ECPPDA 能够抗窃听攻击.

2) 内部攻击

攻击者通过捕获群智感知中的参与者(节点)或服务提供商进行内部攻击.攻击者能够获得被捕获节点自身的一切数据,因此攻击不仅能够获得公共参数$(G,C_a,Y_1,\cdots,Y_i)$,还能得到被捕获节点n_i的私钥x_i.同理,攻击者获得其他节点明文,必需得到秘钥$(x_1,x_2,\cdots,x_i)$,所以攻击者不能通过秘钥x_i计算出$r_i\times P^{CK}=x_1\times C_a^i+\cdots+x_i\times C_a^i$.当内部攻击者为服务器时,攻击者只能得到每个簇的融合后明文,而不能推导出单个节点的明文数据.同时,服务器节点不能获得秘钥$(x_1,x_2,\cdots,x_i)$,从而也不能通过密文解密得到单个节点的明文.差分攻击即$\{n_1,n_2,\cdots,n_k\}$的2个子集S_1和S_2只有一个节点不同,这个节点的数据可以通过$Sum(S_1)-Sum(S_2)$推导出来.ECPPDA具有抵御差分攻击的能力,因为ECPPDA中的簇内节点都不相同,无法通过差分攻击获得节点感知数据.综上,ECPPDA能够抵抗内部攻击.

3) 共谋攻击

参与者与服务器或多个参与者共谋攻击其他节点以获取其敏感信息.由于每个节点n_j的秘钥x_j是独享的,因此共谋的多个参与者不能得到完整的秘钥$(x_1,x_2,\cdots,x_i)$,从而不能通过密文(C_a^j,C_b^j)计算得到节点n_j的明文d_j.因此ECPPDA可抵御共谋攻击.当簇内恶意节点(共谋者)数量为$k-1$时,诚实节点(目标节点)的数据会泄露,假设不诚实的节点概率为γ时,诚实节点泄密的概率和簇内成员节点个数有关,其概率为$\gamma^{k-1}\times(1-\gamma)\times k$,因此当$k$较大,其概率可忽略不计,且ECPPDA设置初始簇成员节点数量$k=\gamma\times m+2$时能避免这种极端情况的出现.

4.3 复杂度分析

群智感知中参与节点资源通常相对受限,而服务器资源比较丰富,所以本文主要考虑节点的资源消耗情况.ECPPDA算法主要由3个部分组成:初始化、数据传输、解密.初始化阶段节点经过计算得到公钥,然后传输公钥和分簇信息给服务器以完成节点分簇,所以初始阶段单个节点的时间与空间复杂度为$O(1)$,通信量为$O(|G|)$($|G|$表示基点G的点长),因此,群智感知网络在初始化阶段时间与空间复杂度为$O(n)$以及总通信量为$O(n|G|)$;数据传输阶段,每个节点需完成一次数据加密和密文融合,并将融合密文传输给下一节点,所以此该阶段单个节点的时间与空间复杂度分别为$O(1)$和$O(|G|)$,通信量为$O(|G|)$,该阶段网络总的时间与空间复杂度分别为$O(n)$和$O(n|G|)$,总通信量为$O(n|G|)$;解密阶段,节点通过服务器发送的聚合密文(C_a,C_b)计算得到D_i并传输给服务器,所以解密阶段单个节点的时间与空间复杂度分别为$O(1)$和$O(|G|)$,以及通信量为$O(|G|)$.综上分析,ECPPDA算法总的时间与空间复杂度分别为$O(n)$和$O(n|G|)$,群智感知网络总通信量为$O(n|G|)$.

5 实验与结果实验评估

实验环境为Win7 OS,Intel Core i5 CPU和16 GB内存,采用GeoLife项目[18,24]中的公开数据点来对本文提出的方案进行评估.该公开数据点是由不同的GPS记录器每间隔5～10 m或者2～5 s采集的.随机选择中间的70个数据点,经过相应的坐标放大变换转化为2维相对坐标.为了对算法的通信量和隐私性能进行比较,下面对PDAIF[22]算法和ECPPDA算法进行仿真实验对比.

图2给出了2种算法在簇内节点数目k不同时节点的通信量比较,由于PDAIF算法通过未来消息缓冲机制保证容错性,节点每次传输的是2份数据,且在构建簇时预先在服务器缓存B个备份数据,所以ECPPDA方案可以有效地降低节点的通信消耗.

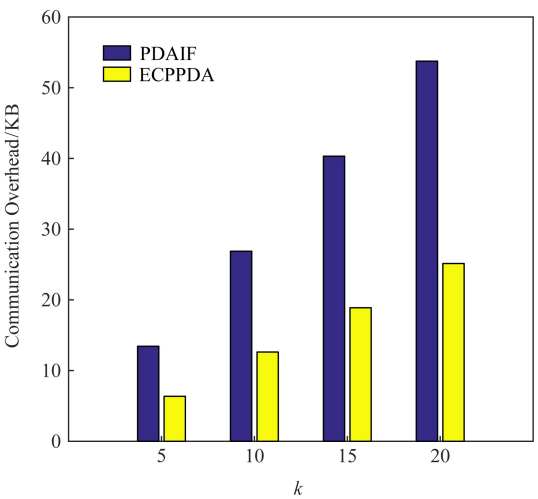

Fig. 2 Communication overhead with different number of nodes

图2 通信量比较

图3给出了在恶意节点概率为10%时,不同簇节点数目下任意节点的数据隐私泄露概率,即攻击者通过第三方服务器和恶意参与者进行共谋攻击,

解密单个节点数据的概率,由于 PDAIF 协议没有考虑共谋攻击,当簇内节点越多时,共谋攻击情况下单个节点数据隐私泄露的概率就越高.ECPPDA 方案通过簇内所有成员共同协作进行密文解密的方式来抵御共谋攻击,其次通过数据加密抵御外部窃听攻击来保证数据的机密性,所以 ECPPDA 方案可以有效地保护节点数据的隐私.

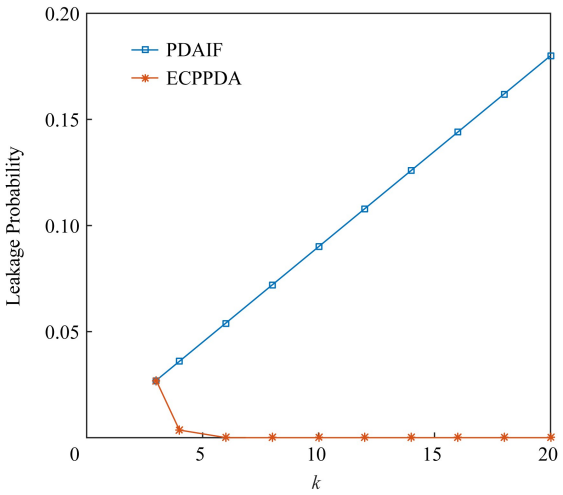

Fig. 3 The leakage probability of single node

图 3 单个节点泄密概率

图 4 给出了簇内成员节点数目不同时(恶意节点概率为 10%),簇内所有节点的感知数据发生隐私泄露的概率,即攻击者通过服务器和恶意参与者进行共谋攻击,获得簇内诚实节点的感知数据的概率,由于 PDAIF 算法没有考虑共谋攻击,仅与左右邻居节点交换参数,因此当攻击者与目标节点的左右邻节点进行共谋就能获得该节点的感知数据.

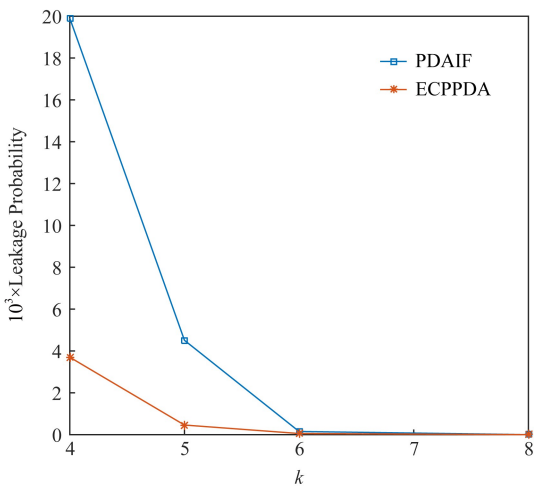

Fig. 4 The leakage probability of cluster

图 4 簇内节点泄露概率

ECPPDA 算法中每个节点数据的解密需要簇内所有节点协同合作,因此 ECPPDA 算法中节点数据发生隐私泄露的概率远远低于 PDAIF 算法,如图 4 所示.

由安全性分析可知只要簇内节点数量不小于 k 时,所有节点数据泄露概率可忽略不计,即使不满足节点数量要求,节点数据泄露的概率也较低,特别当簇内节点数量超过 6 时,节点数据发生隐私泄露概率几乎忽略不计.当簇内节点数为 2 时,单个节点数据泄露概率较高,因为只要有 1 个恶意节点就会导致数据泄露.如图 5 所示:

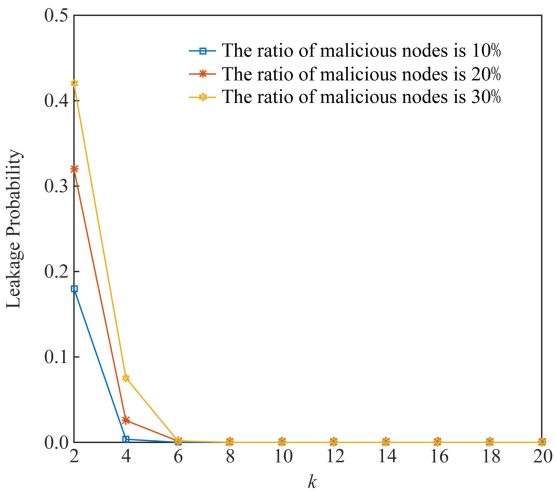

Fig. 5 The leakage probability of single node with different ratio of malicious nodes

图 5 恶意节点不同占比下单个节点泄密概率

图 6 给出簇内成员节点数初始值 k 不同情况下,新加入节点数 b 不同时需要更新簇公钥的节点

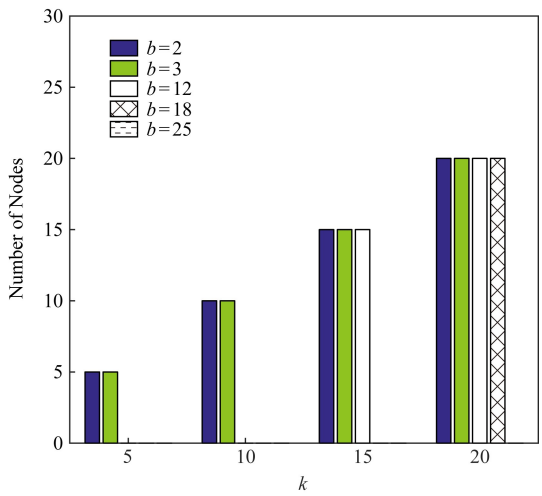

Fig. 6 Number of nodes that need to update the cluster public key with different number of added nodes

图 6 加入不同节点数时需要更新簇公钥的节点数

数量.由图 6 可知,当加入节点数量小于簇内成员节点数量时,需要更新的节点数量为簇内成员节点数,因为加入的节点会选择 1 个簇加入然后管理器需要更新该簇内原有节点的公钥.当加入节点数量大于簇内节点 k 值时,需要更新的节点数为 0,因为加入节点会选择建立新簇,因此不会影响到原有节点.

6 总 结

群智感知中设计具有高安全、高效率、高质量隐私保护的融合数据方案是一个挑战性的问题.本文提出了一种具有隐私保护的融合数据算法 ECPPDA.ECPPDA 通过服务器构造多个簇,然后簇节点利用簇公钥加密数据并在簇内融合传输,直到最后 1 个簇节点将融合数据上传至服务器,服务器通过与簇的所有成员协同合作簇融合密文进行解密,并对所有簇的数据进行融合得到任务区域内所需的融合数据.数据传输过程中采用超时重传机制来解决数据丢失问题.由于密文是通过多方协同合作进行解密,所以 ECPPDA 可以抵御共谋攻击.理论分析和实验表明 ECPPDA 算法能够保护每个参与者数据的隐私性,显著降低了参与者的通信开销,并保证了数据的实用性.

参 考 文 献

[1] Yu Ruiyun, Wang Pengfei, Bai Zhihong, et al. Participatory sensing: People-centric smart sensing and computing [J]. Journal of Computer Research and Development, 2017, 54(3): 457-473 (in Chinese)
(于瑞云, 王鹏飞, 白志宏, 等. 参与式感知: 以人为中心的智能感知与计算[J]. 计算机研究与发展, 2017, 54(3): 457-473)

[2] Burke J, Estrin D, Hansen M, et al. Participatory sensing [C] //Proc of the World Sensor Web Workshop at ACM Sensys. New York: ACM, 2006: 117-134

[3] Fan Xiaoyi, Liu Jiangchuan, Wang Zhi, et al. CrowdNavi: Demystifying last mile navigation with crowdsourced driving information [J]. IEEE Transactions on Industrial Informatics, 2017, 13(2): 771-781

[4] Yang Hongming, Deng Youjun, Qiu Jing, et al. Electric vehicle route selection and charging navigation strategy based on crowd sensing [J]. IEEE Transactions on Industrial Informatics, 2017, 13(5): 2214-2226

[5] Lin Lu, Li Jianxin, Chen Feng, et al. Road traffic speed prediction: A probabilistic model fusing multi-source data [J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(7): 1310-1323

[6] Jing Yao, Guo Bin, Chen Huihui, et al. Crowdtracker: Object tracking using mobile crowd sensing [J]. Journal of Computer Research and Development, 2019, 56(2): 328-337 (in Chinese)
(景瑶, 郭斌, 陈荟慧, 等. Crowdtracker: 一种基于移动群智感知的目标跟踪方法[J]. 计算机研究与发展, 2019, 56(2): 328-337)

[7] Maisonneuve N, Stevens M, Niessen M, et al. Noisetube: Measuring and mapping noise pollution with mobile phones [C] //Proc of Information Technologies in Environmental Engineering. Berlin: Springer, 2009: 215-228

[8] Meng Chuishi, Jiang Wenjun, Li Yaliang, et al. Truth discovery on crowd sensing of correlated entities [C] //Proc of the 13th ACM Conf on Embedded Networked Sensor Systems. New York: ACM, 2015: 169-182

[9] Castelluccia C, Chan A, Mykletun E, et al. Efficient and provably secure aggregation of encrypted data in wireless sensor networks [J]. ACM Transactions on Sensor Networks, 2009, 5(3): Article 20: 1-20

[10] Shi E, Chan T, Rieffel E, et. al. Privacy-preserving aggregation of time-series data [C] //Proc of the 18th Annual Network and Distributed System Security Symp. Berlin: Springer, 2011: 111-125

[11] Zeng Juru, Chen Hong, Peng Hui, et al. Privacy preservation in mobile participatory sensing [J]. Chinese Journal of Computers, 2016, 39(3): 595-614 (in Chinese)
(曾菊儒, 陈红, 彭辉, 等. 参与式感知隐私保护技术[J]. 计算机学报, 2016, 39(3): 595-614)

[12] Hu Ling, Hahabi C. Privacy assurance in mobile sensing networks: Go beyond trusted servers [C] //Proc of the 8th IEEE Int Conf on Pervasive Computing and Communications Workshops. Piscataway, NJ: IEEE, 2010: 613-619

[13] Qiu Fudong, Wu Fan, Chen Guihai. Privacy and quality preserving multimedia data aggregation for participatory sensing systems [J]. IEEE Transactions on Mobile Computing, 2015, 14(6): 1287-1300

[14] Dimitriou T, Krontiris I, Sabouri A. PEPPeR: A querier's privacy enhancing protocol for participatory sensing [C] //Proc of Int Conf on Security and Privacy in Mobile Information and Communication Systems. Berlin: Springer, 2012: 93-106

[15] Cornelius C, Kapadia A, Kotz D, et al. Anonysense: Privacy-aware people-centric sensing [C] //Proc of the 6th Int Conf on Mobile Systems, Applications and Services. New York: ACM, 2008: 211-224

[16] Peng Tao, Liu Qin, Meng Dacheng, et al. Collaborative trajectory privacy preserving scheme in location-based services [J]. Information Sciences, 2017, 387: 165-179

[17] Chen Jianwei, Ma Huadong, Zhao Dong, et al. Participant density-independent location privacy protection for data aggregation in mobile crowd-sensing [J]. Wireless Personal Communications, 2018, 98: 699-723

[18] Jung Taeho, Li Xiangyang, Wan Meng. Collusion-tolerable privacy-preserving sum and product calculation without secure channel [J]. IEEE Transactions on Dependable and Secure Computing, 2015, 12(1): 45-57

[19] He Wenbo, Liu Xue, Nguyen H, et al. PDA: Privacy-preserving data aggregation for information collection [J]. ACM Transactions on Sensor Networks, 2011, 8(1): Article 6: 1-22

[20] Ma Teng, Liu Yun, Zhang Zhenjiang. An energy-efficient reliable trust-based data aggregation protocol for wireless sensor networks [J]. International Journal of Control and Automation, 2015, 8(3): 305-318

[21] Balli M, Uludag S, Selcuk A, et al, Distributed multi-unit privacy assured bidding (PAB) for smart grid demand response programs [J]. IEEE Transactions on Smart Grid, 2017, 9(5): 4119-4127

[22] Chen Jianwei, Ma Huadong, Zhao Dong. Private data aggregation with integrity assurance and fault tolerance for mobile crowd-sensing [J]. Wireless Networks, 2017, 23(1): 131-144

[23] Lu Rongxing, Cao Zhenfu. Simple three-party key exchange protocol [J]. Computers & Security, 2007, 26(1): 94-97

[24] Zheng Yu, Zhang Lizhu, Xie Xing, et al. Mining interesting locations and travel sequences from GPS trajectories [C] // Proc of the 18th Int Conf on World Wide Web. New York: ACM, 2009: 791-800

Wang Taochun, born in 1979. PhD, professor. Member of CCF. His main research interests include crowd sensing, privacy preservation and WSNs.

Jin Xin, born in 1994. Master. His main research interest is security of the crowd sensing.

Lü Chengmei, born in 1995. Master candidate. Her main research interest is security of the crowd sensing.

Chen Fulong, born in 1978. PhD, professor. Senior member of CCF. His main research interests include embedded computing and pervasive computing, cyber-physical systems, high-performance computer architecture, and security of the Internet of things.

Zhao Chuanxin, born in 1978. PhD, professor. Member of CCF. His main research interests include wireless network, optimization, and machine learning.