# 轨迹隐私保护技术研究

霍 峥 孟小峰
(中国人民大学信息学院 北京 100872)

**摘 要** 随着移动设备和定位技术的发展,产生了大量的移动对象轨迹数据.轨迹数据含有丰富的时空信息,对其分析和挖掘可以支持多种与移动对象相关的应用.然而,针对轨迹数据的攻击性推理可能导致个人的兴趣爱好、行为模式、社会习惯等隐私信息暴露.另一方面,在基于位置的服务中,由于现有位置隐私保护技术并不能解决轨迹隐私泄露的问题,移动对象的个人隐私很可能通过实时运行轨迹而暴露.针对上述两种场景,轨迹隐私保护的研究提出了明确的要求:在轨迹数据发布中,隐私保护技术既要保护轨迹数据的隐私,又要保证数据有较高的可用性;在基于位置的服务中,隐私保护技术既要保护移动对象的实时轨迹隐私,又要保证用户获得较高的服务质量.该文针对上述两个问题分析了轨迹隐私保护中存在的挑战性问题,针对不同的隐私保护方法分析了现有的研究工作,介绍了当前该领域的研究热点,指明了未来的研究方向.

**关键词** 数据库应用;隐私保护;轨迹数据;数据发布;基于位置的服务
**中图法分类号** TP391 **DOI 号**: 10.3724/SP.J.1016.2011.00000

# A Survey of Trajectory Privacy-Preserving Techniques

HUO Zheng MENG Xiao-Feng
(*School of Information, Renmin University of China, Beijing* 100872)

**Abstract** As the high-up development of Location-Based Service (LBS) and location-aware devices, the amount of locations and trajectories of moving objects collected by service providers is continuously increasing. The collected trajectories with wealth spatio-temporal information will be published for novel applications. However, directly publishing trajectories may present serious threats to individuals' privacy, since trajectories enable intrusive inferences, which may reveal individual's habits, behavioral patterns, social customs, etc. On the other hand, individuals' real time trajectories may be exposed when they are requiring for location-based services even if their location privacy have been protected. Moreover, specific requirements for trajectory privacy-preserving methods are proposed based on different application scenarios: In trajectory data publishing scenario, privacy-preserving techniques must preserve data utility; in LBS scenario, privacy-preserving techniques must guarantee high quality of services that users acquired. All these requirements make trajectory privacy-preserving more challenging. According to the above problems, the key challenges in trajectory privacy-preserving are analyzed; recent research works on both trajectory data publishing and trajectory privacy in LBS are analyzed in this paper. At last, suggestions for future research works are put forward.

**Keywords** database applications; privacy-preserving; trajectory data; data publication; location-based service

收稿日期: 2011-07-30.最终修改稿收到日期:20011-09-15.本课题得到国家自然科学基金(60833005,61070055,91024032)、中国人民大学科学研究基金(10XNI018)、核高基重大专项(2010ZX01042-002-003)、高等学校博士学科点专项科研基金(200800020002)资助.
**霍 峥**,女,1982 年生,博士研究生,主要研究方向为轨迹隐私保护.**孟小峰**,男,1964 年生,教授,博士生导师,主要研究领域为 Web 数据管理、移动数据管理、XML 数据管理、云数据管理等. E-mail: xfmeng@ruc.edu.cn.

# 1 引 言

随着移动设备和定位技术的发展,近年来涌现出各种移动定位设备,如,车载导航仪、带有 GPS 功能的手机、PDA、平板电脑以及位置传感器等.移动定位设备的出现催生了多种基于位置数据的应用,这些应用大致可以分为两类:第 1 类是离线(offline)应用.位置服务提供商或其它机构收集、分析移动对象的轨迹数据或发布给第三方使用.对轨迹的分析和挖掘可以支持多种与移动对象相关的应用,如,优化道路网络及交通管理策略、分析用户的行为模式以支持商业决策等.第 2 类是在线(online)应用,指通过移动对象的实时位置向其提供相应的服务.比如,基于位置的服务(LBS)、移动对象的实时监控、感知位置的营销等.在上述两类应用中,移动对象的位置及轨迹隐私保护是一个至关重要的问题.

从 2003 年开始,研究者们对位置隐私保护技术展开了研究,并获得了丰富的研究成果[1-7].然而,近年来研究者们发现仅仅保护移动对象的位置隐私是不够的:一方面,位置隐私保护并不能解决离线应用中的轨迹隐私泄露问题;另一方面,保护了移动对象的位置隐私并不能保证移动对象的轨迹隐私不泄露.移动对象的原始轨迹暴露可能会导致个人的兴趣爱好、行为模式、生活习惯等隐私信息泄露.比如,通过对轨迹的分析,攻击者不仅能发现移动用户正在什么位置、过去访问过的位置,还可以分析出移动对象的家庭住址、工作地点,甚至可以从日常运行轨迹分析出该用户的行为模式和生活习惯等私密信息[8].在实际应用中,也出现了由于轨迹数据的暴露造成个人隐私泄露甚至人身安全受到威胁的例子.美国福克斯公司曾报道了有人利用 GPS 轨迹数据跟踪前女友以实施报复的案例.因此,轨迹隐私保护已经成为用户和研究者们迫切关注的问题.

本文第 2 节介绍轨迹隐私保护的概念及需要解决的关键问题;第 3 节介绍轨迹隐私保护的系统结构;第 4 节到第 6 节分别介绍 3 种主流的轨迹隐私保护技术;第 7 节对现有方法进行了分析和对比,并对轨迹隐私保护的各类技术进行总结和分析,评估各类技术的性能;最后展望未来研究工作.

# 2 轨迹隐私保护中的关键问题

## 2.1 基本概念

**轨迹**是指某个移动对象的位置信息按时间排序的序列.通常情况下,轨迹 $T$ 可以表示为 $T=\{qi,(x_1,y_1,t_1),(x_2,y_2,t_2),\cdots,(x_n,y_n,t_n)\}$.其中,$qi$ 表示该轨迹的标识符,它通常代表移动对象、个体或某种服务的用户.$(x_i,y_i,t_i)(1\leq i\leq n)$ 表示移动对象在 $t_i$ 时刻的位置为 $(x_i,y_i)$,也称为采样位置或采样点,$t_i$ 则被称为采样时间.移动对象已经停止运行后,收集到的轨迹数据是静态数据;若移动对象在运行中,那么轨迹就是增量更新的动态数据.

**隐私**是指个人、机构等实体不愿意被外部知晓的信息.比如,个人的行为模式、兴趣爱好、健康状况、公司的财务状况等.个人隐私即为数据所有者不愿意被披露的敏感信息,如个人的收入水平、健康状况、兴趣爱好等.由于人们对隐私的限定标准不同,对隐私的定义也有所差异.一般来说,任何可以确认特定个人的,但个人不愿意披露的信息都可以称为个人隐私.

**轨迹隐私**是一种特殊的个人隐私,它是指个人运行轨迹本身含有的敏感信息(如访问过的敏感位置),或者由运行轨迹推导出的其它个人信息(如家庭住址、工作地点、生活习惯、健康状况等).因此,轨迹隐私保护即要保证轨迹本身的敏感信息不泄露,又要防止攻击者通过轨迹推导出其它的个人信息.

## 2.2 轨迹隐私保护的场景

目前,关于轨迹隐私保护的研究工作主要解决在下述两种应用中的隐私问题.

### 2.2.1 数据发布中轨迹隐私保护

轨迹数据本身蕴含了丰富的时空信息,对轨迹数的分析和挖掘可以支持多种移动应用,因此,许多政府及科研机构都加大了对轨迹数据的研究力度.例如,美国政府利用移动用户的 GPS 轨迹数据分析基础交通设施的建设情况,用以更新和优化交通设施[9];社会学的研究者们通过分析人们的日常轨迹研究人类的行为模式;某些公司通过分析雇员的上下班轨迹以提高雇员工作效率等.然而,假如恶意攻击者在未经授权的情况下,计算推理获取与轨迹相关的其它个人信息,用户的个人隐私通过其轨迹完全暴露.数据发布中的轨迹隐私泄露情况大致可分为两类:

(1)由于轨迹上敏感或频繁访问位置的泄露而导致移动对象的隐私泄露.轨迹上的敏感或频繁访问的位置很可能暴露其个人兴趣爱好、健康状况、政治倾向等个人隐私,如,某人在某个时间段内频繁访问医院/诊所,攻击者可以推断出这个人近期患上了某种疾病;

(2) 由于移动对象的轨迹与外部知识的关联导致的隐私泄露. 比如，某人每天早上固定的时间从地点 $A$ 出发到地点 $B$，每天下午固定的时间段从地点 $B$ 到地点 $A$，通过挖掘分析，攻击者很容易做出判断：$A$ 是某人的家庭住址，$B$ 是其工作单位. 通过查找 $A$ 所在区域和 $B$ 所在区域的邮编、电话簿等公开内容，很容易确定某人的身份、姓名、工作地点、家庭住址等信息，某人的个人隐私通过其运行轨迹完全泄露.

在轨迹数据发布中，最简单的隐私保护方法是删除每条轨迹的准标识属性，即 QI(Quasi-identifier)属性. 然而，单纯的将 QI 属性移除并不能保护移动对象的轨迹隐私——攻击者通过背景知识(如受攻击者的博客、谈话记录或其它外部信息)与特定用户相匹配，从而推导出个体的隐私信息. 例如，在删除了 QI 属性的数据中，攻击者发现某个移动对象在某个时刻 $t_i$ 访问了地点 $L_1$ 和 $L_2$，在攻击者已知的背景知识中，小王曾在时刻 $t_i$ 左右分别访问过这两个位置，如果小王是在 $t_i$ 时刻唯一分别访问过 $L_1$ 和 $L_2$ 的移动对象，那么攻击者就可以断定该轨迹属于小王，继而从轨迹中发现小王访问过的其它位置. 可见，简单的删除移动对象的 QI 属性并不能起到隐私保护的目的.

2.2.2　LBS 中的轨迹隐私保护

用户在获取 LBS 服务时，需要提供自己的位置信息，为了保护移动对象的位置隐私，出现了位置隐私保护技术. 然而，保护了移动对象的位置隐私并不一定能保护移动对象的实时运行轨迹的隐私，攻击者极有可能通过其它手段获得移动对象的实时运行轨迹. 比如，利用位置 $k$-匿名(Location $k$-anonymity)模型对发出连续查询的用户进行位置隐私保护时，移动对象的匿名框(Cloaking region)位置和大小产生连续更新. 如果将移动对象发出 LBS 请求时各个时刻的匿名框连接起来，就可以得到移动对象大致的运行路线[10]. 这是由于移动对象在查询过程中生成的匿名框包含了不同的移动对象所造成的，单纯的延长匿名框的有效时间会造成服务质量的下降. 虽然，目前已有针对连续查询的位置隐私保护技术，然而，其查询有效期只有秒级，无法满足轨迹隐私保护的需求. 因此，在 LBS 中也需要轨迹隐私保护技术.

在上述两种场景中，轨迹隐私保护需要解决以下几个关键问题：

(1) 保护轨迹上的敏感/频繁访问位置信息不泄露；

(2) 保护个体和轨迹之间的关联关系不泄露，即，保证个体无法与某条轨迹相匹配.

(3) 防止由移动对象的相关参数限制(最大速度、路网等)而泄露移动对象轨迹隐私的问题.

### 2.3　轨迹隐私保护技术分类与度量标准

2.3.1　轨迹隐私保护技术的分类

轨迹隐私保护技术大致可以分为以下 3 类：

(1) 基于假数据的轨迹隐私保护技术. 它是指通过添加假轨迹对原始数据进行干扰，同时又要保证被干扰的轨迹数据的某些统计属性不发生严重失真[9].

(2) 基于泛化法的轨迹隐私保护技术. 该技术是指将轨迹上所有的采样点都泛化为对应的匿名区域[10,12-19]，以达到隐私保护的目的.

(3) 基于抑制法的轨迹隐私保护技术. 它是指根据具体情况有条件的发布轨迹数据，不发布轨迹上的某些敏感位置或频繁访问的位置以实现隐私保护[20-22].

假轨迹隐私保护方法简单、计算量小，但易造成假数据的存储量大及数据可用性降低等缺点；基于泛化法的轨迹隐私保护技术可以保证数据都是真实的，然而计算开销较大；基于抑制法的轨迹隐私保护技术限制发布某些敏感数据，实现简单，但信息丢失较大. 目前，基于泛化法的轨迹 $k$-匿名技术(Trajectory $k$-anonymity)在隐私保护度和数据可用性上取得了较好的平衡，是目前轨迹隐私保护的主流方法.

2.3.2　轨迹隐私保护度量标准

在轨迹数据发布中，由于发布后的数据要供第 3 方分析和使用，隐私保护技术要在保护轨迹隐私的同时有较高的数据可用性；在基于位置的服务中，隐私保护技术既要保护移动对象的轨迹隐私，又要保证移动用户获得较高的服务质量.

综合起来，轨迹隐私保护技术的度量标准有以下两个方面：

(1) 隐私保护度. 一般通过轨迹隐私的披露风险来反映，披露风险越小，隐私保护度越高. 披露风险是指在一定情况下，轨迹隐私泄露的概率. 披露风险与隐私保护算法的好坏和攻击者掌握的背景知识有很大的关联. 攻击者掌握的背景知识越多，披露风险越高. 在轨迹隐私保护中，攻击者掌握的背景知识可能是在空间中移动对象的分布情况、移动对象的运行速度、该区域的道路网络情况等.

(2) 数据质量/服务质量. 在轨迹数据发布中，

数据质量是指发布数据的可用性，数据的可用性越高，数据质量越好. 一般采用信息丢失率（又称为信息扭曲度）来衡量数据质量的好坏. 在基于位置的服务中，采用服务质量来衡量隐私保护算法的好坏，在相同的隐私保护度下，移动对象获得的服务质量越高则隐私保护技术越成熟. 一般情况下，服务质量由响应时间、查询结果的准确性来衡量.

## 3 轨迹隐私保护的系统架构

在数据发布的轨迹隐私保护中，大多数系统结构基于“先收集、再匿名、后发布”的原则，即由一个数据收集服务器收集轨迹数据，并将原始数据存储到轨迹数据库中. 然后由轨迹隐私保护服务器进行隐私保护处理，最后形成可发布的轨迹数据. 轨迹隐私保护服务器中有 3 个主要模块：数据预处理模块、隐私保护模块和可用性衡量模块，如图 1 所示. 数据预处理模块负责对收集到的轨迹数据进行等价类划分、轨迹同步等预处理操作；隐私保护模块负责对预处理后的数据进行隐私保护处理，最后由可用性衡量模块评估隐私处理后的数据可用性，最后将轨迹数据发布.

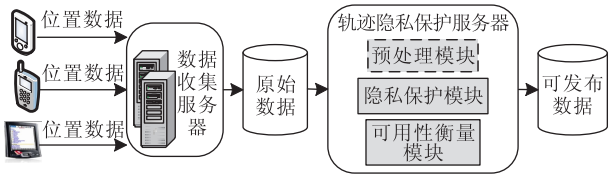


图 1 离线轨迹隐私保护体系结构

另一种架构在收集数据时直接进行隐私保护处理，采用“先匿名、再收集、后发布”的原则，将隐私保护算法放在数据收集之前，收集到的数据可以直接发布. 相比之下，采用前一种架构不要求较高的实时性，可以优先考虑算法的隐私保护度和数据的可用性，后者需要处理动态更新数据的隐私保护，难度更高. 然而，后者可以防止数据收集服务器得到原始数据，用户体验更好.

在基于位置的服务中，轨迹隐私保护系统结构有分布式点对点结构和中心服务器结构两种. 分布式点对点结构由客户端和服务提供商两个部件组成，客户端之间通过 P2P 协议通信，判断客户端之间的距离，通过彼此协作完成隐私保护过程. 中心服务器结构由客户端、服务提供商和匿名服务器三部分组成，如图 2 所示. 匿名服务器包含了隐私保护模块和结果处理模块. 隐私保护模块负责收集客户端的位置、对客户端进行隐私保护处理；结果处理模块负责接受服务提供商发回的候选结果，对候选结果求精，并将最终结果返回给客户端. 由于中心服务器结构具有容易实现、掌握全局数据等优点，已经成为目前最常用的系统结构.

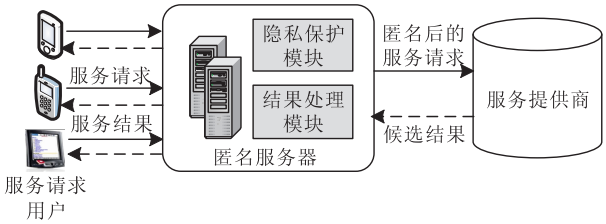


图 2 在线轨迹隐私保护体系结构

## 4 基于假数据的轨迹隐私保护技术

在上述度量标准和系统架构的基础上，本节开始对几种主流的轨迹隐私保护技术进行介绍和分析.

### 4.1 假轨迹方法概述

在位置隐私保护技术中，假位置是经常使用的一种简单有效的技术. 假位置即不发布真实位置，用假位置获得相应的服务[23]. 在轨迹数据隐私保护中，同样可以使用假轨迹方法. 假轨迹方法通过为每条轨迹产生一些假轨迹来降低披露风险. 例如，表 1 中存储了原始轨迹数据，移动对象 $O_1$，$O_2$，$O_3$ 在 $t_1$，$t_2$，$t_3$ 时刻的位置存储在数据库中，形成了 3 条轨迹.

**表 1 原始数据**

| *MOB* | $t_1$ | $t_2$ | $t_3$ |
|---|---|---|---|
| $O_1$ | (1,2) | (3,3) | (5,3) |
| $O_2$ | (2,3) | (2,7) | (3,8) |
| $O_3$ | (1,4) | (3,6) | (5,8) |

表 2 是对原始数据进行假数据扰动后的结果. $I_1$，$I_2$，$I_3$ 分别是移动对象 $O_1$，$O_2$，$O_3$ 的假名. $I_4$，$I_5$，$I_6$ 是生成的假轨迹的假名. 经过假轨迹扰动后的数据库中含有 6 条假轨迹，每条真实轨迹的披露风险降为 1/2. 简单的说，产生的假轨迹越多，披露风险就越低.

**表 2 用假数据法干扰后的轨迹数据库**

| *MOB* | $t_1$ | $t_2$ | $t_3$ |
|---|---|---|---|
| $I_1$ | (1,2) | (3,3) | (5,3) |
| $I_2$ | (2,3) | (2,7) | (3,8) |
| $I_3$ | (1,4) | (3,6) | (5,8) |
| $I_4$ | (1,1) | (2,2) | (3,3) |
| $I_5$ | (2,4) | (2,6) | (4,6) |
| $I_6$ | (1,3) | (2,5) | (3,7) |

一般来说，假轨迹方法要考虑以下几个方面：

(1) 假轨迹的数量. 假轨迹的数量越多，披露风

险越低，但是同时对真实数据产生的影响也越大，因此，假轨迹的数量通常根据用户的隐私需求选择折衷数值；

(2) 轨迹的空间关系. 从攻击者的角度看，从交叉点出发的轨迹易于混淆，因此，应尽可能产生相交的轨迹以降低披露风险；

(3) 假轨迹的运行模式. 假轨迹的运动模式要和真实轨迹的运动模式相近，不合常规的运行模式容易被攻击者识破.

### 4.2　假轨迹方法的实现

针对上述 3 种要求，出现了两种生成假轨迹的方法：随机模式生成法和旋转模式生成法.

(1) 随机生成法. 随机生成一条连接起点到终点、连续运行且运行模式一致的假轨迹.

(2) 旋转模式生成法. 以移动用户的真实轨迹为基础，以真实轨迹中的某些采样点为轴点进行旋转，旋转后的轨迹为生成的假轨迹. 旋转点的选择和旋转角度的确定需要和信息扭曲度进行关联权衡. 旋转模式生成法生成的假轨迹与真实用户的运动模式相同，并和真实轨迹有交点，难以被攻击者识破.

## 5　基于泛化法的轨迹隐私保护技术

在轨迹隐私保护中，最常用的方法是轨迹 $k$-匿名技术. $k$-匿名模型最早是由文献[24]提出的，主要应用在关系数据库的隐私保护中. 其核心思想是将 QI 属性泛化，使得单条记录无法和其它 $k-1$ 条记录区分开来. Marco Gruteser 最先将 $k$-匿名技术应用到位置隐私保护中，产生了位置 $k$-匿名模型：当移动对象在某一时刻的位置无法与其它 $k-1$ 个用户的位置相区别时，称此位置满足位置 $k$-匿名. 随后，$k$-匿名模型应用到轨迹隐私保护技术中，产生了轨迹 $k$-匿名，一般来说，$k$ 值越大则隐私保护效果越好，然而丢失的信息也越多.

给定若干条轨迹，对于任意一条轨迹 $T_i$，当且仅当在任意采样时刻 $t_i$，至少有 $k-1$ 条轨迹在相应的采样位置上与 $T_i$ 泛化为同一区域时，称这些轨迹满足轨迹 $k$-匿名，满足轨迹 $k$-匿名的轨迹被称为在同一个 $k$-匿名集中. 采样位置的泛化区域(又称匿名区域)可以是最小边界矩形(MBR)，也可以是最小边界圆形(MBC)，可以根据具体需求进行调整. 下面的例子展示了轨迹 $k$-匿名的概念.

**表 3　轨迹 3-匿名**

| $MOB$ | $t_1$ | $t_2$ | $t_3$ |
|---|---|---|---|
| $I_1$ | [(1,2),(2,4)] | [(2,3),(3,7)] | [(3,5),(3,8)] |
| $I_2$ | [(1,2),(2,4)] | [(2,3),(3,7)] | [(3,5),(3,8)] |
| $I_3$ | [(1,2),(2,4)] | [(2,3),(3,7)] | [(3,5),(3,8)] |

表 3 是对表 1 中的原始数据进行轨迹 3-匿名后的结果. 表 3 中的 $I_1$，$I_2$ 和 $I_3$ 分别是移动对象 $O_1$，$O_2$，$O_3$ 的假名. 3 个时刻的位置也泛化为 3 个移动对象的最小边界矩形. 匿名区域采用左下坐标和右上坐标来表示. 例如，[(1,2),(2,4)]表示左下角坐标为(1,2)，右上角坐标为(2,4)的最小边界矩形.

在数据发布中和基于位置的服务中均有关于轨迹 $k$-匿名技术的研究，两种场景下对轨迹 $k$-匿名的侧重点不同，下面分别介绍两个场景中的轨迹 $k$-匿名方法.

### 5.1　数据发布中的轨迹 $k$-匿名

在轨迹数据发布的隐私保护中，轨迹 $k$-匿名要将静态的轨迹数据库 $D$ 转换为 $D^*$，使得 $D^*$ 中的任意一条轨迹 $T_i^*$ 都属于某个轨迹 $k$-匿名集，且 $D^*$ 和 $D$ 之间的信息扭曲度最小. 在信息扭曲度最小的情况下达到轨迹 $k$-匿名是 NP-hard 问题，其中有以下几个关键的研究问题.

#### 5.1.1　QI 属性的识别

QI 属性又称为准标识符，它是指联合起来能唯一识别某个个体的多个属性的集合. 比如，邮编、生日、性别等联合起来是准标识符. 在关系数据隐私保护中，属性一般分为 QI 属性和敏感信息，隐私保护技术将 QI 属性泛化，使得发布的数据中每一条记录不能区分于其它 $k-1$ 条记录. 然而在轨迹数据中，QI 属性与敏感信息很难界定，轨迹上任何位置或位置的集合都有可能成为区分于其它轨迹的 QI 属性. 比如，在某个时刻，$T_i$ 是唯一一个经过了位置 $L_i$ 和 $L_j$ 的轨迹，那么 $L_i$ 和 $L_j$ 就可以作为 $T_i$ 的 QI 属性.

多数算法在保护轨迹数据隐私时，并不考虑 QI 属性与其它属性的区别，而是将整条轨迹上的任何一个采样点均做泛化处理[12-13]；也有的方法从动态 QI 属性的角度出发进行隐私保护[14]. 所谓动态 QI 属性是指，某条轨迹的 QI 属性在不同的时刻 $t_i$ 由不同的位置组成. 由于动态 QI 属性自身的特性，必须找到在所有时刻 $t_1,\cdots,t_n$ 上距离 $O$ 的聚集距离最小的 $k-1$ 个对象，并将这 $k$ 个对象匿名到一个匿名区域中以达到轨迹 $k$-匿名.

#### 5.1.2　轨迹 $k$-匿名集的形成

寻找轨迹 $k$-匿名集的原则是使得 $D^*$ 与 $D$ 的信

息扭曲度尽可能小，因此，匿名集中的 $k$ 条轨迹在“时空”上要尽可能地相近，即匿名集中的轨迹既要分布在相同的时间段内又要在空间距离上相近. 为了能达到时间相近的目的，大多数算法都采用了预处理的方式，将分布在相同时间段内的轨迹放入同一个等价类中. 然后，在同一个等价类中寻找空间距离相近的轨迹 $k$-匿名集. 寻找轨迹 $k$-匿名集的方法有两大类，一类是通过整条轨迹聚类找到距离相近的轨迹形成 $k$-匿名集[13]；另一类是通过某条轨迹上最近邻采样点位置找到轨迹 $k$-匿名集[12,14]. 不管用何种方式，为了到达较小的信息扭曲度，都必须遵循 $k$ 条轨迹之间距离尽可能小的原则.

5.1.3 轨迹距离的计算

轨迹聚类或寻找最近邻采样点均需要计算轨迹或者采样位置之间的距离. 目前，研究者们提出了多种轨迹距离的计算方法，比如欧几里德距离(Euclidean distance)、编辑距离(edit distance)、最长共同序列距离(longest common sequences distance)、对数距离(log cost metric)等. 目前，大多数方法采用欧几里德距离计算轨迹或采样点之间的距离，也有方法采用对数距离计算轨迹之间的距离. 选择何种距离计算函数和信息扭曲度的衡量函数有直接关系. 例如，如果信息扭曲度由数据库 $D$ 和 $D^*$ 之间的欧几里德距离衡量，那么，相应的轨迹距离也应采用欧式距离.

## 5.2 LBS 中的轨迹 $k$-匿名

基于位置的服务是指服务提供商根据移动用户的位置信息提供各种服务，比如，紧急救援服务、基于位置的娱乐信息服务、生活信息服务以及基于位置的广告服务等. 由于 LBS 服务与用户提出请求的位置有关，因此，使用基于位置的服务最大的隐私威胁就是位置隐私的泄露[25]，也就是说，暴露用户的位置以及获知位置后用户收到的与时空相关的推理攻击. 比如，用户不想让别人知道目前所在的位置(如酒吧)以及将要去的位置(如查询最近的宾馆等). 位置隐私保护技术的出现解决了这类问题，它可以保护移动用户在某个时刻的位置信息以及用户在发出连续查询时的位置信息. 然而，更严重的问题是：保护了用户的位置隐私并不一定能保护用户的轨迹隐私. 比如，通过位置隐私保护技术，移动对象在发出 LBS 请求的时刻均发布了一个匿名框，将这些匿名框连接起来，会暴露移动对象的大致轨迹.

在基于位置的服务中，轨迹 $k$-匿名与位置 $k$-匿名不同，轨迹 $k$-匿名要求任一条轨迹在起始点至终止点的所有采样位置都必须和相同的 $k-1$ 条轨迹匿名. 基于位置的服务中的轨迹 $k$-匿名与数据发布中的轨迹 $k$-匿名不同，待匿名的轨迹数据不是静态的，而是动态变化的. 因此，如何从轨迹起始时就能确定轨迹 $k$-匿名集是一个挑战性问题. 在基于位置的服务中，轨迹 $k$-匿名集的方法大致有以下几种.

5.2.1 基于轨迹划分的轨迹 $k$-匿名

将轨迹分片，对每个片段与其它轨迹的片段进行匿名可以解决将整条轨迹匿名带来的不确定性问题，文献[26-27]提出了轨迹分片匿名的方法. 分片方法的关键问题在于：如何确定轨迹片段的长度——轨迹片段太短，则无异于位置隐私保护，起不到轨迹隐私保护的作用；轨迹片段太长，则起不到划分的效果. 文献[26]中的轨迹划分的方法是将二维空间划分为大小相等的正方形“格”，根据用户的隐私需求可将一个或多个“格”定义为一个“划分”. 假如一条轨迹穿过不同的“划分”，“划分”的边界将这条轨迹分成若干个轨迹片段，然后再分别对处于不同“划分”的轨迹片段进行匿名. 在划分交界处的位置隐私保护也是需要关注的问题，文献[26]则采用了在边界位置延时发布匿名区域的策略.

5.2.2 基于历史轨迹的 $k$-匿名

多数匿名方法都是和当前时间段内的移动对象匿名，匿名是否成功很大程度上依赖于路网的稠密度. 如果路网过于稀疏，容易造成匿名框过大，从而影响服务质量；若在服务时间内达不到用户设定的隐私级别，则会造成匿名失败. 基于上述问题，有的方法提出了用历史数据和用户的运行轨迹匿名的方法. 历史数据匿名技术采用中心服务器模式，客户端和位置服务器之间有一个可信的匿名服务器，且匿名服务器中含有存储历史轨迹数据的数据库. 移动对象增量地向匿名服务器发送运行轨迹 $T_0=\{c_1, c_2, \cdots, c_n\}$，匿名服务器需要为 $T_0$ 产生匿名区域的序列 $T=\{C_1, C_2, \cdots, C_n\}$，使 $T$ 完全覆盖 $T_0$，且包含 $k-1$ 条历史轨迹. 该方法通过为每一条历史轨迹建立基于格的索引来获取距离 $T_0$ 最近的历史轨迹. 在基于格的索引中，先使用四分树将二维空间划分为大小不等的“格”，为每个“格”维护一张表，表中存储了经过该“格”的轨迹 $id$ 以及其它信息. 通过该索引可以找到与 $T_0$ 经过相同“格”的轨迹，并将这些轨迹存入集合 $B$ 中，如果 $B$ 中的轨迹数据不足 $k-1$ 个，则继续查找经过与 $T_0$ 相邻的格的轨迹放入 $B$ 中，直至 $B$ 中含有至少 $k-1$ 条轨迹为止. 由于 $B$ 中的轨迹和 $T_0$ 经过相同或相邻的“格”，距离 $T_0$ 较近. 形成

轨迹 $k$-匿名的轨迹从集合 $B$ 中选取. 这样就完成了轨迹 $k$-匿名.

## 6　基于抑制法的轨迹隐私保护技术

抑制法是指有选择地发布原始数据，抑制某些数据项，即不发布某些数据项. 表 4 和表 5 展示了通过抑制法进行轨迹隐私保护的例子. 表 4 中存储了坐标与语义位置之间的对应关系（该信息可以通过反向地址解析器和黄页相结合得到），假如攻击者获得该信息，就可以作为背景知识对发布的数据进行推理攻击.

**表 4　位置信息表**

| Location | Name |
|---|---|
| (1,2) | Clinic |
| (2,7) | Hotel |
| (5,8) | Bar |
| (3,9) | Shopping Mall |

表 5 是经过简单抑制之后发布的轨迹数据，可以看出，所有敏感位置信息都被限制发布，移动对象的隐私得到保护.

**表 5　用抑制法隐私保护的数据**

| *MOB* | $t_1$ | $t_2$ | $t_3$ |
|---|---|---|---|
| $O_1$ | — | (3,3) | (5,3) |
| $O_2$ | (2,3) | — | (3,8) |
| $O_3$ | (1,4) | (3,6) | — |

简单来说，抑制法有如下两个原则：

(1) 抑制敏感/频繁访问的位置信息；

(2) 抑制增大整条轨迹披露风险的位置信息.

如何找到需要抑制的位置信息以降低披露风险且尽可能地提高数据的可用性是抑制法需要解决的关键问题. 文献[20]根据攻击者掌握移动对象的部分轨迹的情况，提出了抑制某些信息来保护移动用户轨迹隐私的方法. 该方法要解决的问题是将轨迹数据库 $D$ 转换为 $D^*$，使得攻击者 $A$ 不能以高于 $P_{br}$ 的概率推导出轨迹上的位置属于某个移动对象. 假定轨迹 $T$ 上的位置 $p_j$ 来源于位置集合 $P$，不同的攻击者拥有不同的位置集合，攻击者 $A$ 的位置集合表示为 $P_A$，攻击者 $A$ 掌握的轨迹片段表示为 $T_A$. 因此，需要计算某个不属于 $P_A$ 的位置可能被 $A$ 推导出其所有者的概率，如果这个概率大于 $P_{br}$，则 $p_j$ 必须被抑制. 使用抑制法进行隐私保护时，如果抑制的数据太多，势必会严重影响数据的可用性.

文献[22]中采用了另一种抑制法进行隐私保护. 该方法根据某个区域访问对象的多少将地图上的区域分为敏感区域和非敏感区域，一旦移动对象进入敏感区域，将抑制或推迟其位置更新，以保护其轨迹隐私. 对于非敏感区域，算法并不限制移动对象的位置更新.

抑制法简单有效，能处理攻击者持有部分轨迹数据的情况. 在保证数据可用性的前提下，抑制法是一种效率较高的方法. 然而，上面提到的方法仅适用于了解攻击者拥有某种特定背景知识的情形，当隐私保护方不能确切地知道攻击者的背景知识时，这种方法就不再适用.

## 7　性能指标与分析比较

### 7.1　性能指标

在轨迹数据发布的隐私保护中，需要衡量隐私后的数据可用性. 轨迹数据库 $D$ 经过隐私保护之后转化为可发布的轨迹数据库 $D^*$. 计算 $D$ 和 $D^*$ 之间的信息扭曲度有如下几种方式.

(1) 用原始数据库 $D$ 与发布数据库 $D^*$ 之间的距离来衡量信息扭曲度. 两个轨迹数据库之间的距离是指 $D$ 和 $D^*$ 中对应的轨迹之间的欧式距离之和. 对于某条轨迹 $T\subset D$，发布之后转化为轨迹 $T^*\subset D^*$，$T$ 到 $T^*$ 的距离就是 $T$ 到 $T^*$ 之间对应采样点的欧式距离之和. 由于这种衡量方法要求每条轨迹 $T$ 都有对应的发布轨迹 $T^*$，一般用于抑制法的信息扭曲度计算中.

(2) 采用分辨率来衡量信息扭曲. 根据分辨的内容不同，分辨率方法又可以分为按采样位置的分辨率与按整条轨迹的分辨率两种. 采样位置分辨率是指在某个时刻可以完全确定移动对象的位置的概率，文献[14]中采用了采样位置分辨率衡量方法，它将信息扭曲度定义为移动对象的采样位置分辨率在 $D^*$ 中的降低程度. 文献[13]中采用了整条轨迹分辨率的方法，该方法假定数据库 $D$ 经过匿名后形成的匿名集为 $P=\{p_1,p_2,\cdots,p_n\}$，其中 $p_n$ 为删除的轨迹集合，轨迹的分辨率取决于 $p_i$ 的半径大小与 $p_n$ 中包含的轨迹数目. 用分辨率的方法衡量信息扭曲度的方法多用在轨迹 $k$-匿名中.

提高发布数据的可用性大致有以下几种策略：(1) 采样点的匿名区域要尽可能地小；(2) 删除的轨迹或者采样点的数目要尽可能地少；这两条策略可以直接降低信息扭曲度；(3) 由于大部分轨迹分析和挖掘算法都是针对原子轨迹进行的，因此，发布原子轨迹比发布匿名区域的可用性高，尽可能地发布

原子轨迹也是提高数据可用性的可行方法.

### 7.2 分析比较

本文对目前国内外在轨迹隐私保护研究中提出的各类方法进行了分类和对比,列举了各类方法的主要优点、主要缺点以及代表性技术,如表 6 所示.总的来说,3 类方法各有优缺点:假数据方法计算开销小,实现简单,但是算法移植性较差、数据可用性/服务质量较差;而泛化的方法虽然算法移植性以及数据可用性/服务质量有较高的提升,实现代价却也大大提高;抑制法实现简单且隐私保护度较高,然而数据失真严重.设计隐私保护方法时要根据隐私保护的需求以及针对的攻击模型出发,设计合适的隐私保护算法.

**表 6 各类隐私保护方法对比**

| 方法 | 主要优点 | 主要缺点 | 代表技术 |
|---|---|---|---|
| 假数据法 | 计算开销小;实现简单 | 数据失真严重;算法移植性较差 | Dummy[9]<br>Path protection[15] |
| 泛化法 | 算法移植性好;数据较真实;实现简单 | 实现最优化轨迹匿名开销较大;<br>有隐私泄露风险 | $(k,\delta)$-anonymity[13]<br>Anonymity-reconstruction[12]<br>Split-generalization[26]<br>History data anonymity[10]<br>Extreme Union[14]<br>Symmetric Anonymization[14] |
| 抑制法 | 实现简单;隐私保护度较高 | 数据失真严重 | Suppression-based[20]<br>Location tracking[22] |

作为新兴的研究热点,轨迹隐私保护的主要研究方向分为数据发布中的轨迹隐私保护和基于位置服务中的轨迹隐私保护,研究者分别提出了多种轨迹隐私保护方法.这些方法主要是从传统关系数据隐私保护方法向时空方向拓展而得到的.本文总结并分析了目前的轨迹数据隐私技术,对各种技术进行了分类,指出了各种技术的典型应用并对比分析了它们的隐私保护度以及数据可用性/服务质量等衡量标准.隐私保护度分为"低"、"中"、"高"3 种级别,数据可用性/服务质量分为"差"、"一般"、"好"3 种级别.表 7 中列出了目前主要的轨迹隐私保护算法.从表 7 中可以看出,隐私保护度主要取决于添加的假数据数量、抑制的敏感数据数量、匿名集中含有的轨迹数据;数据可用性/服务质量或者是根据发布数据库 $D^*$ 与原始数据库 $D$ 之间的欧式距离来衡量;或是根据整条轨迹从 $D^*$ 中区分出来的概率来衡量;也可以是从用某个时刻的匿名区域分辨精确位置的概率来衡量;少数针对数据挖掘的工作则是根据在 $D^*$ 上进行数据挖掘的效率来衡量.综合比较,泛化法在隐私保护度与数据可用性/服务质量之间获得了较好的平衡,因此,泛化法是目前隐私保护技术中的主流方法,作为代表技术之一的轨迹 $k$-匿名也成为轨迹隐私保护的主流技术.抑制法和假数据方法在某些特定攻击模型下,或对隐私保护度要求不高的情况下可以采用.

**表 7 各种隐私保护方法对比**

| 方法名称 | 典型应用 | 方法分类 | 隐私保护度 | 数据可用性/服务质量 |
|---|---|---|---|---|
| Dummy[9] | 轨迹数据发布 | 假数据法 | 低;主要取决于假轨迹数目 | 差;数据可用性由 $D^*$ 和 $D$ 之间的欧式距离衡量; |
| Suppression-based[22] | | 抑制法 | 中;主要取决于抑制信息的多寡与敏感程度 | |
| (k,δ)-anonymity[13] | | 泛化法 | 高;主要取决于同一个匿名集中轨迹的个数 | 好;数据可用性由轨迹 $T$ 从 $D^*$ 中区分出来的概率表示 |
| Anonymity-reconstruction[12] | | | | 好;数据可用性由删除的位置数量衡量 |
| Extreme Union[14] | | | | 好;数据可用性用 *MOB* 在某时刻的位置分辨率的降低来衡量 |
| Symmetric Anonymization[14] | | | | |
| Path protection[18] | 基于位置的服务 | 假数据法 | 低;主要取决于产生的假出发点、目的地的个数 | 差;服务质量由产生的假出发点、目的地的数量衡量,数量越多服务质量越低 |
| Location tracking[22] | | 抑制法 | 中;主要取决于抑制的敏感位置多寡 | 一般;服务质量由 *MOB* 在某时刻位置的分辨率衡量 |
| KAT[10] | | 泛化法 | 高;主要取决于同一个匿名集中轨迹的个数 | 一般;服务质量用在每个时刻产生的匿名框的面积衡量 |
| Split-generalization[26] | | | | 好;在隐私保护后的用户数据上进行数据挖掘,根据数据挖掘效率衡量 |
| Dummy-generalization[19] | | 混合法 | 高;主要取决于产生的假轨迹的数目以及同一个匿名集中的轨迹数目 | 一般;服务质量由产生的假轨迹片段、结点之间交换的轨迹片段数量衡量 |

## 8　未来工作展望

目前，轨迹数据的隐私保护还是一个新的研究领域，很多挑战性的问题有待解决：

(1) 轨迹匿名新技术的研究

在轨迹隐私保护技术中，最广泛使用的模型是轨迹 $k$-匿名模型. 然而，该模型并不能完全保证轨迹隐私不泄露. 轨迹 $k$-匿名模型保证匿名框内至少覆盖 $k$ 条轨迹的相应采样点，即攻击者无法获知 $k$ 个移动对象的真实轨迹. 假如 $k$ 条轨迹的相似度极高，则可能造成轨迹隐私泄露. 极端情况下，$k$ 条轨迹重叠在一条线上(或重叠于同一点上)，或者都经过同一个敏感区域，则轨迹隐私泄露. 文献[28]中，作者借鉴数据发布中隐私保护的 $l$-差异性($l$-diversity)思想，提出了保证轨迹差异性的方法，即保证在同一匿名集中的 $k$ 条轨迹之间有一定的差异性以免造成隐私泄露，作者采用每条轨迹的最小边界矩形的面积来定义轨迹 $l$-差异性：当 $k$ 条轨迹形成的 MBR 的面积大于某个阈值时，认为满足差异性标准. 我们认为，仅采用 MBR 的面积之差来衡量差异性是不足的，具体原因有以下两点：(1) $k$ 条轨迹形成的 MBR 面积主要取决于轨迹运行的曲线，它并不能真实反映 $k$ 条轨迹之间的距离，满足 MBR 的阈值标准并不能防止 $k$ 条轨迹不重合或者距离太近；(2) 该方法无法处理几条轨迹同时经过敏感位置的问题，即使两条轨迹的 MBR 的面积满足差异性标准，仍会暴露移动对象的隐私. 我们认为，可以采用两条轨迹之间形成的面积大小来衡量两条轨迹之间的差异性，如果距离值太小，这两条轨迹不宜匿名在同一匿名集中. 然而，简单的规定最小距离势必会带来数据可用性的下降. 因此，如何定义匿名轨迹的空间差异性，并使用新的轨迹匿名模型以防止差异性不足导致轨迹隐私泄露是一个具有挑战性的问题.

此外，在轨迹数据发布的隐私保护中，现有的轨迹数据隐私保护方法在防止攻击时，都假设攻击者拥有不同的背景知识(不同的背景知识导致不同的攻击模型)，然后根据不同的攻击模型设计出相应的模型和算法. 一旦攻击者拥有新的背景知识，算法将不再适用，移动对象的隐私也极有可能泄露. 然而，提前获知攻击者拥有的背景知识也是不现实的问题. 文献[29]中定义了多种攻击模式，并分析了各种攻击模式下隐私泄露的风险，如果能设计一种隐私保护算法能同时在多种攻击模型下保护数据隐私，则必会降低数据的可用性. 因此，如何分析在各种背景知识下的披露风险，并设计出相应的隐私保护策略也是保护移动对象轨迹隐私的重要研究方向.

(2) 针对语义轨迹的隐私保护

近年来，语义轨迹成为热门的研究点[30-31]，语义轨迹是指通过对轨迹数据的分析和挖掘，将轨迹上的简单位置变成有语义信息的地点，从而可以支持交通优化、热门区域推荐等应用[32-33]. 然而，从轨迹隐私保护的角度出发，相对于用 GPS 记录表示的单纯位置，攻击者更关注的是移动对象曾经访问过哪些地方、将要去哪些地方，而不单单是经过了哪些位置. 因此，运行过程中移动对象“访问过哪些地方”是更值得我们保护的. 另一方面，整条轨迹匿名由于会造成数据可用性下降，也不利于各种应用分析. 因此，可以从语义轨迹的角度出发，针对移动对象访问过的位置进行隐私保护.

我们认为，可以将轨迹上的停留点抽取出来，使用地理反向编译器将其转换为地址，然后，将停留点转换为泛化的区域，用泛化的区域对地图进行划分，并将移动对象的轨迹转换为(运行点，停留点)的序列，运行点不做处理，将停留点用泛化区域来替代，从而阻止攻击者获知移动对象曾经访问过的具体位置. 然而，如何生成泛化区域，泛化后的区域中应该包含哪种位置点才能更好地保护移动对象的隐私是需要继续研究的问题.

(3) 面向隐私保护的轨迹数据收集

近来，几大知名移动手机生产厂商面临用户的信任危机——一些智能手机或其上的第 3 方应用被指控不顾及用户隐私而强行收集用户的位置数据，并定期将这些数据发送给生产厂商的数据中心用以分析移动用户的行为模式或者发布给其它机构进行商业应用. 目前的轨迹隐私保护技术大都是将数据收集并进行隐私保护处理之后再发布给其它机构使用. 由于没有第 3 方监管，假如数据收集者在没有对数据进行隐私保护处理的情况下分析数据，会给用户隐私带来极大的威胁. 因此，对于用户来说，一种能避免任何机构获得原始数据的隐私保护方法会更加安全. 面向隐私保护的轨迹数据收集能够在收集数据的同时进行隐私保护的处理，数据收集者也无法得到最原始的数据. 然而，面向隐私保护的数据收集不可避免的需要用户之间的协作，这样就存在了用户是否可信、是否有恶意攻击者假冒用户来窃取用户的隐私等诸多问题. 如何辨别可信用户、半可信

用户、恶意用户以及如何在 3 种用户同时存在的情况下设计隐私保护策略都是颇具挑战性的问题.

(4)轨迹数据外包中的隐私保护

某些收集轨迹数据的部门(如位置服务提供商、交通监管部门等)在收集了大量的移动对象轨迹数据,又无力管理海量数据时,迫切需要将轨迹数据外包给第 3 方(如云数据库系统)以有效地降低数据管理的成本.查询用户直接向第 3 方提出查询,第 3 方将满足条件的数据发送给查询者.由于大部分轨迹数据包含了移动对象的个人信息,在轨迹数据外包时,需要对其进行隐私保护处理以防移动对象的隐私泄露给第 3 方.传统的数据外包中隐私保护方式多以数据加密为基础,然而,由于轨迹数据自身的特点,传统的数据外包的隐私保护方式不能完全适用.比如,很多加密算法并不支持基于距离的查询,因此,为轨迹数据设计新的加密模式,使其能够支持基于距离的查询(如 $k$NN 等)是未来值得继续研究的问题.

(5)新应用中的位置与轨迹隐私保护

随着社交网络以及移动网络技术的发展,产生了一种新的应用——移动社交网络.移动社交网络是一种基于地理位置的社交网络,在移动社交网络中,用户之间可以根据自己的地理位置信息和朋友之间分享生活经历,建立联系.典型的应用包括,用户之间可以共享旅游路线;通过挖掘多个用户的轨迹与位置信息,发现热门的地点,进行旅游推荐;此外,还可以根据用户相似的行为模式进行朋友推荐等等.然而,在移动社交网络中,位置和轨迹隐私保护是迫切需要解决的问题.用户使用移动社会网络社交时,可以根据用户自身的需求或环境的变动使用不同的隐私保护级别.比如,当用户在某个公共场所使用移动社会网络时(比如,在某个公园时,该用户发布照片),使用较低的隐私保护级别,该用户所有的朋友都可以获取这个资源;相反,在比较私密的场所发布资源时,为了避免其它用户知道该用户在私密场所,发布的资源就要经过一定的处理.此外,资源发布时间之间的联系也是隐私保护需要考虑的问题,资源发布的时间差和移动对象的运行速度,可以推测某个用户在某个时间段内可以到达或者不能到达某处[34].最后,由于在移动社会网络中,即使用户可以控制自己发布内容的情况,但是却无法控制好友发布的内容中涉及到自己,这也给用户的位置或轨迹隐私保护带来了很大的挑战.总之,在新应用中,尤其是移动社交网络中,用户的位置和轨迹隐私保护是迫切需要解决的问题.

## 9 结束语

随着移动定位设备、无线传感网络的发展以及基于位置的服务的发展,人们在方便地获取基于位置的服务的同时,也存在着严重的隐私泄露的风险.目前研究者们已经在位置隐私保护方法方面做了大量的工作,在新兴的轨迹隐私保护方面也有一些研究工作.本文对最近几年来国际上在该领域的主要研究成果进行了回顾与总结,综述了在数据发布中与基于位置的服务中的隐私保护技术研究的现状,对各种方法进行了分析和对比,指出仍然存在的问题和将来可能的解决办法.总地来说,轨迹隐私保护技术的研究仍然处于起步的阶段,仍然有大量关键的问题还需要做深入细致的研究.

## 参 考 文 献


[1] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking//Proceedings of the 1st International Conference on Mobile Systems, Applications, and Services (MobiSys 2003). San Fransisco, 2003: 31-42

[2] Mokbel M F, Chow C Y, Aref W G. The newcasper: Query processing for location services withoutcompromising privacy//Proceedings of the 32nd Conference of Very Large Databases (VLDB 2006). Seoul, 2006: 763-774

[3] Bamba B, Liu L. Supporting anonymous location queries in mobile environments with privacy grid//Proceeding of the 17th International Conference on World Wide Web (WWW 2008). Beijing, 2008: 237-246

[4] Pan X, Meng X, Xu J. Distortion-based anonymity for continuous queries in location-based mobile services//Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS 2009). Washington, 2009: 256-265

[5] Pan X, Xiao Z, Meng X. Survey of location privacy-preserving. Journal of Frontiers of Computer Science and Technology, 2007, 1(3): 268-281(in Chinese)
(潘晓,肖珍,孟小峰. 位置隐私研究综述. 计算机科学与探索, 2007, 1(3): 268-281)

[6] Krumm J. A survey of computational location privacy. Personal and Ubiquitous Computing, 2009, 13(6): 391-399

[7] Bettini C, Wang S X, Jajodia S. Protecting privacy against location-based personal identification//Proceedings of the 2nd VLDB workshop on Secure Data Management (SDM2005). Trondheim, 2005: 185-199

[8] Krumm J. Inference attacks on location tracks//Proceedings of the 5th International Conference on Pervasive Computing (PERVASIVE 2007). Toronto, 2007: 127-143

[9] Luper D, Cameron D, Miller J A, Arabnia H R. Spatial and temporal target association through semantic analysis and GPS data mining//Proceedings of the 2007 International Conference on Information & Knowledge Engineering (IKE 2007). Las Vegas, 2007: 251-257

[10] Xu T, Cai Y. Exploring historical location data for anonymity preservation in location-based services//Proceedings of the 27th Conference on Computer Communications (INFOCOM 2008). Phoenix, 2008: 547-555

[11] You T, Peng W, Lee W. Protecting moving trajectories with dummies//Proceedings of the 2007 International Conference on Mobile Data Management (MDM 2007). Mannheim, 2007: 278-282

[12] Nergiz M E, Atzori M, Saygin Y, Baris G. Towards trajectory anonymization: A generalization-based approach. IEEE Transactions on Data Privacy, 2009, 2(1): 47-75

[13] Abul O, Bonchi F, Nanni M. Never walk alone: Uncertainty for anonymity in moving objects databases//Proceedings of the 2008 IEEE 24th International Conference on Data Engineering (ICDE 2008). Cancun, 2008: 376-385

[14] Yarovoy R, Bonchi F, Lakshmanan L, Wang H W. Anonymizing moving objects: How to hide a MOB in a crowd ? //Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology (EDBT 2009). Saint-Petersburg, 2009: 72-83

[15] Mohammed N, Fung B M, Debbabi M. Walking in the crowd: Anonymizing trajectory data for pattern Analysis//Proceeding of the 18th ACM Conference on Information and Knowledge (CIKM 2009). Hong Kong, 2009: 1441-1444

[16] Hoh B, Gruteser M, Xiong H, Alrabady A. Preserving privacy in GPS traces via uncertainty-aware path cloaking//Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS 2007). Alexandria, 2007: 161-171

[17] Gkoulalas A, Verykios V S, Mokbel M F. Identifying unsafe routes for network-based trajectory privacy//Proceedings of the 9th SIAM International Conference on Data Mining (SDM 2009). Nevada, 2009: 942-953

[18] Lee K, Lee W, Leong H V, Zheng B. Navigational path privacy protection//Proceeding of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009). Hong Kong, 2009: 691-700

[19] Gidofalvi G, Huang X, Bach P T. Privacy preserving trajectory collection//Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS 2008). Irvine, 2008: 46

[20] Terrovitis M, Mamoulis N. Privacy preserving in the publication of trajectories//Proceedings of the 9th International Conference on Mobile Data Management (MDM 2008). Beijing, 2008: 65-72

[21] Abul O, Atzori M, Bonchi F, Giannotti F. Hiding sensitive trajectory patterns//Proceedings of the 7th IEEE International Conference on Data Mining Workshops (ICDMW 2007). Singapore, 2007: 693-698

[22] Gruteser M, Liu X. Protecting privacy in continuous location-tracking applications. IEEE Security and Privacy, 2004, 2(2): 28-34

[23] Kido H, Yanagisawa Y, Satoh T. An anonymous communication technique using dummies for location based services//Proceedings of the 21st International Conference on Data Engineering Workshops (ICDE worshops 2005). Tokyo, 2005: 1248

[24] Sweeney L. K-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10(5): 557-570

[25] Mokbel M F. Privacy in location-based services: State-of-the-art- and research directions//Proceedings of the 2007 International Conference on Mobile Data Management (MDM 2007). Mannheim, 2007: 228

[26] Gidofalvi G, Huang X, Pedersen T B. Privacy-preserving data mining on moving object trajectories//Proceedings of the 2007 International Conference on Mobile Data Management (MDM 2007). Mannheim, 2007: 60-68

[27] Shin H, Vaidya J, Atluri V, Choi S. Ensuring privacy and security for LBS through trajectory partitioning//Proceedings of the 11th International Conference on Mobile Data Management (MDM 2010). Missouri, 2010: 224-226

[28] Machanavajjhala A, Gehrke J, Kifer D. L-diversity: Privacy beyond k-anonymity//Proceedings of the 22nd International Conference on Data Engineering (ICDE 2006). Atlanta, 2006: 24

[29] Chris Y T Ma, David K Y Yau, Nung Kwan Yip, Nageswara S V Rao. Privacy vulnerability of published anonymous mobility traces//Proceedings of the 16th Annual International Conference on Mobile Computing and Networking (MOBICOM 2010). Chicago, 2010: 185-196

[30] Yan Z. Towards semantic trajectory data analysis: A conceptual and computational approach//Proceedings of the VLDB 2009 Ph. D. Workshop. Lyon, 2009: 1-6

[31] Yan Z, Spremic L, Chakraborty D, Parent C, Spaccapietra S, Aberer K. Automatic construction and multi-level visualization of semantic trajectories//Proceedings of the 18th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems (GIS 2010). San Jose, 2010: 524-525

[32] Cao X, Cong G, Jensen C S. Mining significant semantic locations from GPS data//Proceedings of the VLDB Endowment, 2010, 3(1): 1009-1020

[33] Zheng Y, Zhang L, Xie X, Ma W. Mining interesting locations and travel sequences from GPS trajectories//Proceedings of the 18th International Conference on World Wide Web (WWW 2009). Madrid, 2009: 791-800

[34] Freni D, Vicente C R, Mascetti S, Bettini C, Jensen C S. Preserving location and absence privacy in geo-social networks//Proceedings of the 19th ACM Conference on Information and Knowledge Management (CIKM 2010). Toronto, 2010: 309-318

**HUO Zheng**, born in 1982, Ph. D. candidate. Her research interests are trajectory privacy-preserving and mobile data management.

**MENG Xiao-Feng**, born in 1964, professor and Ph. D supervisor. His research interests include Web data management, mobile data management, native XML database and cloud data management.

**Background**

This research is partially supported by the grants from the Natural Science Foundation of China (Nos. 60833005, 61070055, 91024032), the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (No. 10XNI018), National Science and Technology Major Project of Key Electronic Devices, High-end General-purpose Chips and Fundamental Software Products (No. 2010ZX01042-002-003), Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 200800020002).

Recent years, with the development of location-aware devices, more and more locations and traces of moving objects are collected and then published for novel applications. As an example, analyzing trajectories of passengers in an area may help people making commercial decisions, such as where to build a restaurant, another example can be seen in traffic control systems, analyzing trajectories of vehicles in a city may help government to optimize traffic control strategy. Although publishing trajectories is beneficial for mobility-related decision making processes, it may represent serious threats to individuals privacy, since trajectories contain rich spatio-temporal information, which may reveal individual's habits, health condition, social customs, etc. But simply remove identifiers cannot help, adversaries may re-identify a user through background knowledge linkage.

In order to protect trajectory privacy, more and more attentions have been given to this area. In the past years, researchers proposed several techniques in trajectory privacy-preserving. The content of this paper mainly provides a summary for previous works and helps researchers pay attention to the interesting issues need to be addressed.

Trajectory privacy-preserving is a rather young research area that has received lots of concerns recent years. Before this research, we have studied location privacy-preserving since 2006, several key problems are solved, Related research findings are published in DASFAA, CIKM, ACM SIGSPATIAL GIS and TKDE.