

轨迹数据的差分隐私保护方法综述

孙新越¹, 张伟哲¹, 何慧¹, 杨任宇²

(1. 哈尔滨工业大学计算学部网络空间安全学院, 哈尔滨 黑龙江 150001; 2. 北京航空航天大学软件学院, 北京 100191)

摘要: 随着传感器和定位技术的快速发展, 用户的智能移动设备生成、存储并分享了海量的轨迹数据。这些数据蕴含有价值的个人时空移动特征, 商业公司和政府部门可以利用这些特征为用户和社会提供高效、便捷的服务。然而, 个人的轨迹数据是私密且敏感的, 不合理的数据使用不仅会泄露用户的家庭和工作地址, 还会暴露他们的健康状况、经济情况。这种隐私担忧导致人们不愿分享自己的轨迹数据, 阻碍了位置服务和应用的发展。为了解决这种隐私担忧问题, 差分隐私技术被提出, 它不仅能够为用户的敏感数据提供严格可证明的隐私保证, 还能保留数据中有价值的信息。具体地, 综述了差分隐私技术在保护轨迹数据隐私方面的研究进展; 从隐私模型、应用场景和扰动方式等角度对面向轨迹数据的差分隐私方法进行了重点分析; 对差分隐私在轨迹数据隐私保护方面的发展方向进行了展望。

关键词: 数据隐私; 位置隐私; 差分隐私; 轨迹数据

中图分类号: TP309

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2025027

Survey of differentially private methods for trajectory data

SUN Xinyue¹, ZHANG Weizhe¹, HE Hui¹, YANG Renyu²

1. School of Cyberspace Science, Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China

2. School of Software, Beihang University, Beijing 100191, China

Abstract: With the rapid development of sensor and positioning technologies, vast amounts of trajectory data were generated, stored, and shared by users' smart mobile devices. These data contained valuable personal spatiotemporal mobility features, which could be leveraged by businesses and government agencies to provide efficient and convenient services to users and society. However, individual trajectory data were private and sensitive, and improper data usage could not only expose users' home and work addresses but also reveal their health status and economic conditions. These privacy concerns led to reluctance in sharing trajectory data, hindering the development of location-based services and applications. To address this issue, a promising solution—differential privacy (DP) technology—

收稿日期: 2025-01-01; 修回日期: 2025-04-30

通信作者: 张伟哲, wzzhang@hit.edu.cn

基金项目: 国家自然科学基金(62402137); 国家资助博士后研究人员计划(GZC20242205); 国家自然科学基金联合基金(U22A2036); 深圳市高校稳定支持计划(GXWD20220817124251002); 深圳市科技计划项目(GXWD20231130110352002)

Foundation Items: The National Natural Science Foundation of China(62402137), The Postdoctoral Fellowship Program of CPSF(GZC20242205), The Joint Funds of the NSFC(U22A2036), The Shenzhen Colleges and Universities Stable Support Program(GXWD20220817124251002), The Shenzhen Stable Supporting Program(GXWD20220817124251002)

引用格式: 孙新越, 张伟哲, 何慧, 等. 轨迹数据的差分隐私保护方法综述[J]. 网络与信息安全学报, 2025, 11(3): 1-18.

Citation Format: SUN X Y, ZHANG W Z, HE H, et al. Survey of differentially private methods for trajectory data[J]. Chinese Journal of Network and Information Security, 2025, 11(3): 1-18.

was proposed. DP could offer rigorous, provable privacy guarantees for users' sensitive data while preserving valuable information. The research progress of DP in protecting trajectory data privacy was reviewed. The DP methods for trajectory data were analyzed from perspectives such as privacy models, application scenarios, and perturbation mechanisms. Finally, an outlook on the future development of DP for trajectory data privacy protection was provided.

Keywords: data privacy, location privacy, differential privacy, trajectory data

0 引言

在万物互联的时代,海量的轨迹数据由具备传感技术和定位能力的个人智能移动设备(如智能手机、智能手表和新能源汽车)产生。这些轨迹数据记录了用户在地理空间中的移动特征,蕴含了大量有价值的信息,如人群分布、用户社交关系等。位置服务的供应商通过分析和挖掘轨迹数据中的信息,为用户提供更优服务。例如,通过分析不同时间段人群的流动模式,网约车公司可以合理地调度网约车以提升车辆的运行效率^[1];城市管理部门可以制定出更加有效的交通灯策略以缓解城市的交通拥堵状况^[2]。

尽管轨迹数据有效地支持了位置服务供应商提供更优质的服务,但其天生的敏感属性也引起了广泛的隐私担忧。未经授权的轨迹数据泄露可能会暴露用户的敏感信息,如家庭住址^[3-4]。更糟糕的是,攻击者可以通过分析用户的轨迹数据推断出更深层的敏感信息,如健康状况和经济状况等^[5]。最近的一项调查显示,90%的受访者对自己的轨迹数据可能被泄露给广告商和陌生人表示担忧^[6]。针对日益严重的数据隐私问题,世界多个国家和组织已经制定了个人数据隐私保护法律。例如,欧盟在2018年正式施行了《通用数据保护法》^[7],限定了任何组织侵害个人数据隐私所应该承担的法律责任和义务。我国第十三届全国人民代表大会常务委员会第三十次会议通过了《中华人民共和国个人信息保护法》^[8],以保护个人信息权益和规范个人信息处理活动。因此,在解决用户隐私担忧的前提下收集、管理和使用轨迹数据成为了势在必行的需求。

事实上,用户数据隐私风险一直是研究人员的关注重点,研究人员提出了众多的解决方案。根据隐私保护机制来划分,现有隐私保护技术可

分为3类:基于匿名化的隐私保护技术、基于数据加密的隐私保护技术和基于差分隐私的隐私保护技术。基于匿名化的隐私保护技术(如k-匿名技术^[9-13]和t-closeness技术^[14]等)通常对用户数据进行匿名处理后才发布。遗憾的是,这些技术已经被证明会导致用户隐私泄露^[15-17]。基于数据加密的隐私保护技术采用加密协议保护数据隐私,主要依赖零知识证明、同态加密和安全多方计算等密码学技术^[18-20]。这类技术要求数据收集者仅收集统计信息,以确保用户的隐私不被泄露。然而,参与数据加密的用户需承担巨大计算开销,这限制了该技术的广泛应用。基于差分隐私的隐私保护技术是通过扰动用户的原始数据进行扰动实现隐私保护的,并通过设计噪声的分布确保扰动后数据在统计分析上的表现近似原始数据^[21-23]。基于差分隐私的隐私保护技术凭借可证明的隐私保护定义和抵抗任意背景知识攻击的能力,已成为数据隐私保护的事实标准并被工业界广泛采用。例如,苹果公司使用该技术收集用户常用的表情符号^[24-25];谷歌公司使用此技术收集用户浏览记录^[26]。因此,本文聚焦差分隐私技术在轨迹数据隐私保护中的应用。

近年来,新的基于差分隐私的轨迹数据隐私保护研究成果不断涌现于国际会议和期刊上。表1对现有涉及轨迹数据差分隐私保护综述^[27-29]的文献进行了总结,从表1可以看出,这些工作要么仅综合分析了广义的隐私保护技术(如k-anonymity^[9-10,12-13]和数据加密^[15-18])而缺乏对差分隐私技术应用的深入分析,要么仅限于讨论在中心化差分隐私(centralized differential privacy, CDP)模型下的轨迹数据发布问题。对此,本文立足于差分隐私模型,对轨迹数据的隐私保护研究进行回顾与梳理,旨在帮助研究人员对该领域的研究进展、主流方法和未来研究方向等有系统性的了解。

本文的主要贡献如下。

表1 涉及轨迹数据差分隐私保护的综述文献的总结

Table 1 Summary of existing literature on differential privacy protection of trajectory data

文献	优点	缺点
文献[27]	覆盖了位置隐私保护研究的各个方面	缺乏对差分隐私技术的深入讨论和分析
文献[28]	聚焦基于CDP的轨迹数据发布	缺乏对本地化差分隐私在轨迹数据隐私保护方面的讨论和分析
文献[29]	从理论和实验层面讨论了基于CDP的轨迹数据发布	

1) 根据差分隐私模型、应用场景和扰动方式3个维度，制定了轨迹数据差分隐私保护方法的分类原则。

2) 基于制定的分类原则，全面回顾了近年来在轨迹数据差分隐私保护方面的重要研究工作。

3) 总结了现有方案中的固有挑战，并展望了未来的研究方向。

1 轨迹数据隐私保护问题概述

本节首先概述了轨迹数据的隐私保护问题，然后探讨用户的隐私担忧和敌手模型。

1.1 背景概述

轨迹数据从采集到管理与分析，再到共享与发布的完整过程通常涉及3个角色，分别是用户、数据收集者和数据消费者，如图1所示。用户为轨迹数据的拥有者，用户将自身的数据发送给数据收集者。数据收集者负责收集和管理用户的轨迹数据，并以提供查询结果和数据发布等方式将数据分享给数据消费者使用。数据消费者使用数据收集者提供的数据库执行统计和分析任务。

实际上，用户的轨迹数据在采集、管理和分析过程中存在多种隐私泄露的威胁。具体来

说，在数据采集阶段，用户的轨迹数据在传输过程中可能被网络攻击者拦截和窃取，甚至攻击者可能会入侵或者劫持数据收集者，使得数据收集者不可信，从而导致用户上传的轨迹数据被攻击者直接获取。在数据管理与分析阶段，不可信的数据收集者可能会故意泄露用户的轨迹数据，甚至从其管理的轨迹数据中挖掘出用户的深层次敏感信息，如健康状况。在数据共享与发布阶段，恶意的数据消费者可以从数据收集者共享的数据和对外提供的查询服务中识别出真实用户并恢复其个人信息。因此，攻击者可能在轨迹数据的收集、管理与分析、共享与发布的任何阶段攻击或窃取用户的隐私信息。

1.2 用户的隐私担忧

由于轨迹数据含有大量的个人敏感信息，因此轨迹数据的采集、管理与分析、共享与发布过程引发了广泛的隐私担忧。首先，轨迹数据本身就是敏感信息，一旦攻击者获取用户的轨迹数据，那么用户的私人信息（如家庭住址）可以轻易地被攻击者获取。其次，攻击者通过分析用户的轨迹数据可以挖掘出深层次敏感信息，如宗教信仰、健康状况、经济状况等^[5]。例如，攻击者

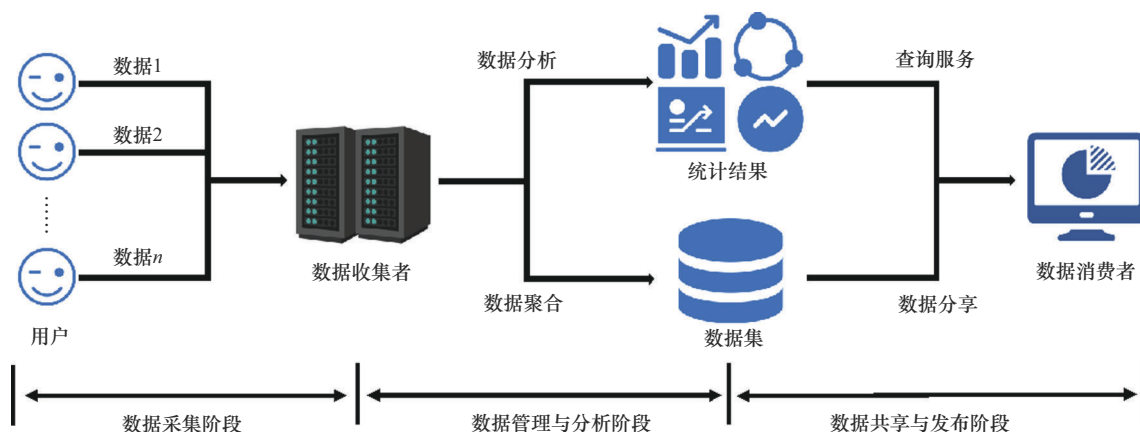


图1 轨迹数据的采集、管理与分析、共享与发布的完整流程

Figure 1 Complete process of trajectory data acquisition, management and analysis, sharing and release

会从用户经常访问医院的轨迹数据中推断出其健康状况。

这些隐私担忧导致用户不愿意分享自身的轨迹数据给数据收集者, 阻碍了数据驱动应用的发展。例如, 为了避免泄露个人隐私信息, 用户可能会选择不再使用定位服务, 或者提供虚假的位置信息, 使得数据收集者无法获得有效的用户轨迹数据。这种行为不仅影响了基于位置的服务和应用的正常运行, 还限制了相关领域的研究进展。因此, 必须在解决这些用户隐私担忧的前提下收集、管理和使用轨迹数据。

1.3 敌手模型

在轨迹数据的采集、管理与分析、共享与发布过程中, 数据收集者和数据消费者都可能成为隐私攻击者。因此, 研究人员通常考虑图2所示的两种敌手模型。

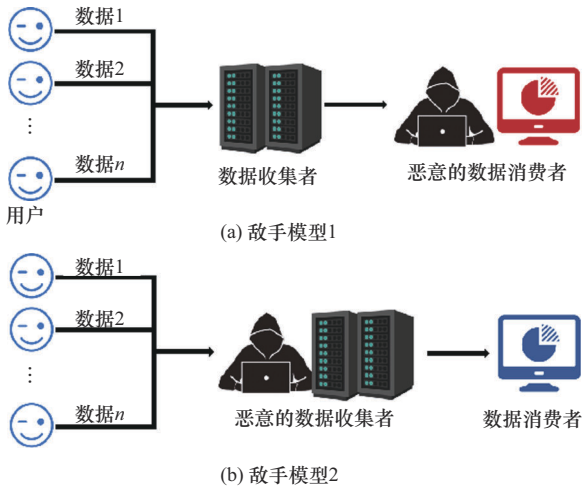


图2 敌手模型
Figure 2 Adversary model

1) 敌手模型1: 恶意的数据消费者。在这种模型中, 数据收集者被认为是可信的, 而数据消费者为恶意的攻击者。因此, 用户会将自己的真实轨迹数据上传给数据收集者, 而数据收集者有责任在分享数据和提供查询的过程中防范恶意的数据消费者窃取用户的敏感信息。

2) 敌手模型2: 恶意的数据收集者。在这种模型中, 数据收集者被认为是恶意的攻击者, 它可以直接获得用户上传的轨迹数据并窃取其隐私信息。因此, 用户只能依靠自己来保护其敏感信息。

2 差分隐私概述

本节首先概述了差分隐私及其隐私模型的处理机制; 然后从差分隐私模型、应用场景及扰动方式3个维度出发, 提出了面向轨迹数据的差分隐私保护方法的分类原则; 最后分析和比较了基于不同隐私模型的隐私保护方法。

2.1 中心化差分隐私

CDP被设计用于对抗敌手模型1, 帮助数据收集者将针对其数据集的精确查询结果转换为特定分布, 从而阻止恶意的数据消费者从查询结果中推断出用户的原始数据。具体来说, CDP的定义依赖于相邻数据集。

定义1 相邻数据集。给定数据集 D 和 D' , 当且仅当 $D \subset D'$ 且 $|D'| = |D| + 1$ (反之亦然) 时, 它们为相邻数据集。

定义2 ϵ -CDP。随机算法 A 满足 ϵ -CDP的条件是, 对于任意相邻数据集 D 和 D' , 以及 A 的所有可能的输出 $R \subseteq \text{Range}(A)$, 下列不等式成立。

$$\Pr[A(D) \in R] \leq e^\epsilon \cdot \Pr[A(D') \in R] \quad (1)$$

其中, $\text{Range}(A)$ 表示随机算法 A 所有可能的输出集合, $\Pr[\cdot]$ 表示事件发生的概率。 ϵ 表示隐私预算, 它控制了整个机制的隐私保护等级。更小的 ϵ 意味着更高的隐私保护等级。

CDP定义表明, 通过观察算法 A 的任意输出 R , 具有任意背景知识的攻击者无法分辨出该算法的输入数据集是原始数据集 D 还是其任意的相邻数据集 D' , 从而保护了原始数据集的隐私。然而, CDP定义仅指出了隐私保护的目标, 并没有给出具体的实现机制。研究人员对实现CDP的机制进行了大量研究, 其中被广泛使用的机制是拉普拉斯机制 (Laplace mechanism, LM)^[22]和指数机制 (exponential mechanism, EM)^[23]。

定义3 拉普拉斯机制。令 $\text{Lap}(\Delta f/\epsilon)$ 表示从均值为0且尺度参数为 $\Delta f/\epsilon$ 的拉普拉斯分布中采样的随机变量。对于任意的函数 $f(\cdot)$, 随机算法 $A(f, D)$ 满足 ϵ -CDP的条件为

$$A(f, D) = f(D) + \text{Lap}(\Delta f/\epsilon) \quad (2)$$

其中, Δf 是 $f(\cdot)$ 在任意两个相邻数据集 D 和 D' 上的敏感度, 它的定义如下所示。

定义4 全局敏感度。令 $f: D \rightarrow \mathbb{R}^m$ 表示一个

函数，该函数将数据集 D 映射到 m 维实数组成的向量，则 f 的全局敏感度定义如下。

$$\Delta f = \max_{D, D'} |f(D) - f(D')| \quad (3)$$

式中， $|\cdot|$ 表示相邻数据集 D 和 D' 上查询结果的 L_1 范数。本质上，拉普拉斯机制通过在查询函数的输出上添加拉普拉斯噪声来满足 ϵ -CDP，噪声的量级由查询函数的全局敏感度和隐私预算 ϵ 共同决定。

定义5 指数机制。对于隐私预算 ϵ 和质量函数 $u: (D \times R) \rightarrow R$ ，以正比于 $e^{u(D,r)/(2\Delta u)}$ 的概率输出 $r \in R$ ，随机算法 $A(D)$ 满足 ϵ -CDP，即

$$\Pr[A(D) = r] = \frac{e^{u(D,r)/(2\Delta u)}}{\sum_{r' \in R} e^{u(D,r')/(2\Delta u)}} \quad (4)$$

其中， Δu 表示 $u(\cdot)$ 在任意两个相邻数据集 D 和 D' 的敏感度，其定义如下所示。

定义6 质量函数的全局敏感度。质量函数 $u: (D \times R) \rightarrow R$ 的全局敏感度的定义如下。

$$\Delta u = \max_{D, D', r \in R} |u(D, r) - u(D', r)| \quad (5)$$

其中， $|\cdot|$ 表示相邻数据集 D 和 D' 上查询结果的 L_1 范数。目前，拉普拉斯机制和指数机制已经被广泛应用于发布轨迹数据^[30-70]。

2.2 本地化差分隐私

CDP 建立在一个假设之上，即存在一个可信的数据收集者集中管理所有用户的数据。然而，当数据收集者遭受恶意攻击或者被黑客劫持时，所有用户的原始数据就会被直接暴露给攻击者。本地化差分隐私（local differential privacy, LDP）^[71-73]被提出来解决这个问题。LDP 假设数据收集者是不可信的，因此用户只能相信自己。为此，每个用户独立地在本地设备上扰动自身的数据以保护其隐私信息，然后将扰动后的数据发送给数据收集者。LDP 确保了数据收集者只能获得用户扰动后的数据，因此可以用于应对恶意的数据收集者（敌手模型2）。

LDP 的数学定义如下所示。

定义7 ϵ -LDP。随机算法 A 满足 ϵ -LDP的条件是，对于任意两个用户的数据 x 和 x' ，以及 A 所有可能的输出 $R \subseteq \text{Range}(A)$ ，下列不等式成立。

$$\Pr[A(x) \in R] \leq e^\epsilon \cdot \Pr[A(x') \in R] \quad (6)$$

其中， ϵ 表示隐私预算。 $\text{Range}(A)$ 表示随机算法 A 的输出域。LDP的定义保证了攻击者无法通过观察满足 ϵ -LDP的算法 A 的输出 R ，高概率地推测出用户的真实数据是 x 还是 x' ，从而保护用户的数据不会泄露。LDP被广泛用于收集用户数据的统计信息。最经典的LDP机制是通用随机响应（general randomized response, GRR）^[74-75]。GRR能够评估固定数据域中任意给定值的频率，它的扰动函数的定义如下。

$$\Pr[\text{GRR}(x) = y] =$$

$$\begin{cases} p = \frac{e^\epsilon}{|X| - 1 + e^\epsilon}, & y = x \\ q = \frac{1}{|X| - 1 + e^\epsilon}, & y = X \setminus x \end{cases} \quad (7)$$

利用该扰动函数，每个用户将自己的私有项目 $x \in X$ 以概率 p 向不受信任的数据收集者发送真实值 x ；以概率 q 发送随机的值 $x' \in X \setminus x$ ，其中 $|X|$ 是数据域 X 中项目的数量。显然，上述扰动函数满足 ϵ -LDP，因为两个概率的比值为 e^ϵ ，即 $p/q = e^\epsilon$ 。为了评估 $x \in X$ 的频率，数据收集者首先统计 x 被用户报告的次数并将其表示为 $c(x)$ ，然后通过如下方式取 x 的评估计数结果。

$$\tilde{c}(x) = \frac{c(x) - n \times q}{p - q} \quad (8)$$

其中， $\tilde{c}(x)$ 是 $c(x)$ 的评估计数， n 代表参与的用户总数量。研究^[76-77]表明， $c(x)$ 是真实计数的无偏估计，并且 $c(x)$ 的方差如下。

$$\text{Var}[c(x)] = n \cdot \frac{|X| - 2 + e^\epsilon}{(e^\epsilon - 1)^2} \quad (9)$$

目前，LDP已经被广泛应用于收集数据^[78-112]。

2.3 轨迹数据差分隐私保护方法分类原则

差分隐私保护技术可以帮助数据收集者以提供查询和数据发布等数据共享方式，将用户的轨迹数据分享给数据消费者使用，同时为数据的隐私性和实用性提供强有力的理论保证。

图3展示了面向轨迹数据的差分隐私保护方法的分类。首先，根据敌手模型可以分为基于CDP和基于LDP的轨迹数据隐私保护方法；然后，根据应用场景和数据扰动方式进一步细分。

根据应用场景，基于CDP的轨迹数据隐私

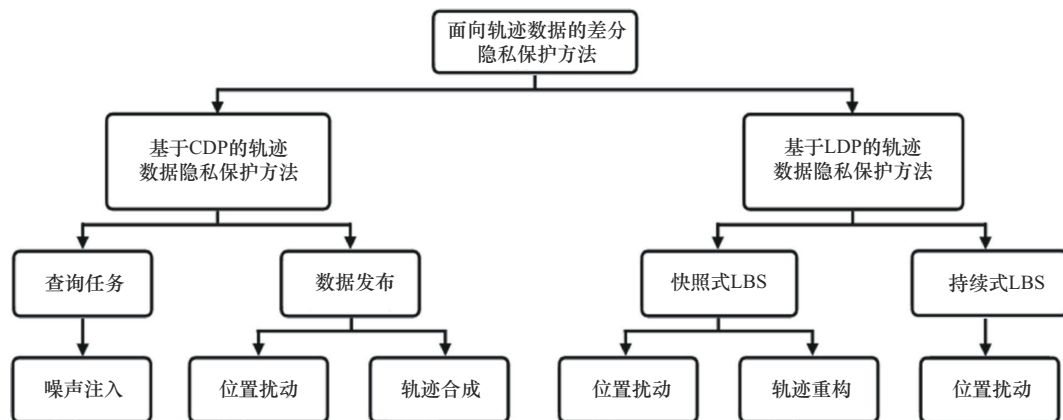


图3 面向轨迹数据的差分隐私保护方法的分类

Figure 3 Classification of differential privacy protection methods for trajectory data

保护方法可以分为查询任务和数据发布两类。在不同的应用场景下，这些方法的扰动方式有所不同。

在查询任务应用场景下，数据收集者在保护用户隐私的前提下，基于其管理的轨迹数据集为数据消费者提供查询结果。这种查询是通过数据加噪的扰动方式实现隐私保护的。具体来说，数据收集者从其管理的真实轨迹数据集中获得精确的查询结果，然后利用差分隐私机制（如拉普拉斯机制、指数机制）向这些结果注入服从某种分布的噪声值来隐藏真实数据中的敏感信息，并将带有噪声的查询结果提供给数据消费者。例如，Cormode等^[42]利用拉普拉斯机制为城市不同区域的人口密度分布注入噪声，并将这种带有噪声的人口密度分布提供给基于位置的服务提供商。

在数据发布应用场景下，数据收集者向数据消费者共享其管理的轨迹数据集，同时保证用户的隐私信息不被窃取。这种隐私保护的轨迹数据发布通常是通过位置扰动和轨迹合成实现的。位置扰动是直接利用差分隐私机制来扰动用户轨迹中的原始位置，然后对外发布扰动后的数据。轨迹合成则是从原始轨迹数据上获得某些统计参数，为这些参数注入满足差分隐私的噪声，然后在这些含有噪声的特征上重新构建出一个合成轨迹数据集。Pelletier等^[103]为蒙特利尔交通公司开发了基于CDP的轨迹数据发布方案，对外提供兼顾隐私保护和高实用性的轨迹数据。

面向轨迹数据的差分隐私保护研究工作早期聚焦于提供查询，后来逐渐扩展到数据发布。这

是因为数据发布允许数据消费者直接在发布的轨迹数据上进行数据分析，既不需要修改现有的数据分析算法，又无须担心隐私问题。

根据应用场景，基于LDP的轨迹数据隐私保护方法可以分为快照式LBS（location based services）和持续式LBS两类，其中快照式LBS根据扰动方式可进一步分为位置扰动和轨迹重构两类。

在快照式LBS应用场景下，用户只需向数据收集者（LBS的服务器）发送一次轨迹数据。例如，用户将他们在指定时间段内的轨迹数据上传给交通管理部门，以支持交通管理部门改善公共服务（如优化或定制公交路线、地铁路线等）。为了保证自身的隐私信息不被窃取，用户需要在本地对其轨迹数据进行扰动，这种扰动通常是通过位置扰动和轨迹重构实现的。位置扰动指用户使用LDP机制对自身轨迹的每个位置进行扰动，并上传扰动后的数据给数据收集者。轨迹重构包含下列两种数据收集方式。

1) 用户首先使用LDP机制对自身轨迹的重叠子片段进行扰动，然后将这些扰动后的片段重构为轨迹数据，并将重构轨迹发送给数据收集者。

2) 用户直接发送扰动后的重叠子片段，数据收集者接收这些子片段，并将其重组为轨迹数据集。

Cunningham等^[93]提出了基于LDP的轨迹扰动方法，允许用户在本地图动自己的轨迹数据，并将扰动后的数据发送给数据收集者。

在持续式LBS应用场景下，用户连续地向数据收集者报告他们当前的位置，同时保证自身的隐私信息不泄露。例如，社交网络服务公司

Foursquare 和 Gowalla 会持续收集和发送用户的位置数据。为了保护隐私, 用户需要持续地对其当前的移动位置数据进行扰动, 并将扰动后的位置数据上传给数据收集者。Wang 等^[92]在用户和数据收集者之间引入了一个可信的代理层, 用于持续收集用户的轨迹数据并利用拉普拉斯机制对所收集的数据进行扰动, 最后将扰动结果上传给数据收集者。

2.4 基于 CDP 和 LDP 的轨迹数据隐私保护方法的分析与比较

(1) 基于 CDP 和 LDP 的轨迹数据隐私保护方法的优缺点

基于 CDP 的轨迹数据隐私保护方法的优点: 执行 CDP 操作的数据收集者通常拥有全局数据, 其可以利用全局数据的统计特性在保护隐私的同时对外提供高精度的轨迹特征。基于 CDP 的轨迹数据隐私保护方法的缺点: 要求用户完全信任数据收集者, 一旦数据收集者是不可信的, 所有用户的轨迹数据都有可能泄露。

基于 LDP 的轨迹数据隐私保护方法的优点: 由用户侧完成轨迹数据的扰动, 因此用户无须信任数据收集者。换句话说, 即使数据收集者是不可信的, 用户的轨迹数据隐私依然能够得到保护。基于 LDP 的轨迹数据隐私保护方法的缺点: 每个用户独立进行扰动, 噪声的累积效应会降低所收集轨迹数据的整体可用性。

(2) 应用场景及其效果

基于 CDP 的轨迹数据隐私保护方法适用于需要对大量轨迹数据进行分析的交通管理和城市规划等场景。例如, 数据收集者可以通过 CDP 为交通管理部门提供大量的隐私保护轨迹数据, 交通管理部门从这些数据中可以获取精确的交通流量信息, 从而优化公交路线, 缓解交通拥堵。

基于 LDP 的轨迹数据隐私保护方法适用于收集高度敏感的轨迹数据的应用 (例如, 提供个人定位服务的应用)。因为这类方法要求用户在本地上扰动其轨迹数据, 并且仅上传扰动后的数据, 这样, 即使服务器被攻击, 用户的真实位置也不会被泄露。

(3) 两种方法的比较

基于 CDP 的轨迹数据隐私保护方法采用集

中化处理模式, 能够在提供高精度数据的同时采用复杂的隐私保护机制实现隐私保护, 适用于需要精细数据分析的场景。然而, 这类方法要求数据收集者高度可信。

相比之下, 基于 LDP 的轨迹数据隐私保护方法在数据源头完成扰动, 有效降低了对数据收集者的信任要求, 适用于高隐私要求和去中心化的应用场景。然而, 这类方法对所收集数据的精度要求相对较低, 适用于较简单的轨迹数据分析任务。

3 基于 CDP 的轨迹数据隐私保护方法

本节梳理了基于 CDP 的轨迹数据隐私保护方法的研究路线, 并从应用和数据扰动两个层面对现有研究进行评述和总结。已有的基于 CDP 的轨迹数据隐私保护方法总结见表 2。

3.1 面向查询任务的轨迹数据隐私保护方法

在保护隐私的前提下回答关于轨迹数据集的查询任务, 常用的实现方法是将地理空间划分为不重叠的单元格, 统计每个单元格中 GPS 位置点的数量或密度, 利用拉普拉斯机制向这些统计结果注入噪声以满足差分隐私的要求, 通过组合这些含噪声的统计数据, 数据收集者就可以回答任意空间范围内的密度、计数或直方图查询。然而, 查询结果的准确性不仅依赖于差分隐私机制的隐私预算, 还与空间划分方式密切相关。空间划分不合理会导致大多数单元格中的实际计数过少, 从而使噪声完全掩盖真实计数, 显著降低查询结果的准确性。

为了解决这一问题, Cormode 等^[42]提出了隐私空间分解 (private spatial decompositions, PSD) 方法, 将地理空间递归地分成更小的单元格, 并对每个单元格中的位置数据进行统计。为确保每个单元格内的计数符合差分隐私的要求并保持统计数据的精度, PSD 利用指数机制来确定位置点的中位数, 从而在满足 ϵ -CDP 的前提下找到树结构 (如 kd-树、二叉树和四叉树) 的分割点。随后, PSD 依据隐私树结构将地理空间细分为相应的单元格, 并向每个单元格的计数数据注入噪声。这样, 数据消费者可以对这些带有噪声的统计结果执行密度查询, 而无须担心泄露原始数据。

Qardaji 等^[43]提出了一种两级自适应网格

表2 已有的基于CDP的轨迹数据隐私保护方法总结

Table 2 Summary of existing methods of trajectory data privacy protection based on CDP

应用场景	扰动方式	扰动算法	隐私模型	特点
提供查询	噪声注入	PSD ^[42]	CDP	可以回答关于原始轨迹数据集的计数和密度查询
		AG ^[43]	CDP	利用约束推理来设计自适应网格方法,以提高查询的准确性
		PrivTree ^[44]	CDP	采用分层分解方式,消除了对预定义的递归深度的依赖
位置扰动		PL ^[45]	GI	将受控的随机噪声添加到用户的位置,以实现地理不可区分性
		OptQL ^[46]	GI	采用线性编程技术实现服务质量损失最小化
		TOM ^[47-48]	GI	只需较少的隐私预算,就能扰动轨迹中的每个位置
		AGENT ^[51]	GI	使用R-tree重复利用生成位置,从而减少对隐私预算的消耗
		PIM ^[52-53]	δ -location set	通过实现 δ -location set来保护轨迹中的每个真实位置
		PIVE ^[36]	GI	利用用户定义的推理误差阈值来确定保护位置集,并生成伪定位
		AGM ^[54]	GI	根据被混淆地点的相关程度来调整噪声
		PriSTE ^[55-59]	ST-event privacy	将位置隐私机制转换为时空事件隐私机制
		GL ^[60]	CDP	注入拉普拉斯噪声扰乱轨迹中重要位置的局部或整体频率分布
数据发布		SDS ^[61]	CDP	采用可变长度的N-gram模型提取轨迹数据集的基本信息
		Hybrid-PRG ^[62]	CDP	使用一个混合粒度的前缀树结构提取轨迹数据集的信息
		DPT ^[63-64]	CDP	使用分层参考系统模型对轨迹进行离散,并采用自适应机制构建前缀树
		TGM ^[67]	CDP	首先利用EM扰动原始轨迹中的每个位置,然后利用LM恢复被扰动位置的方向
		DPTD ^[68]	CDP	首先对原始位置进行聚类以离散化空间,然后构建一个带有噪声的前缀树用来发布差分隐私的轨迹数据
		AdaTrace ^[65]	CDP	提取轨迹数据的差分隐私特征,并利用提取的特征合成轨迹数据
		DP-Star ^[66]	CDP	使用归一化算法对输入轨迹进行预处理,以捕捉准确的移动性特征
		SPRT ^[69]	CDP	利用现实世界的常识(道路网络)来生成差分隐私的轨迹
		PriTrace ^[70]	CDP	使用一阶和二阶马尔可夫链模型,并选择DPT和AdaTrace中的有用信息进行生成
		PUTS ^[100]	CDP	设计了一种从路径到轨迹的双层数据合成方法,用来发布隐私保护的轨迹数据
数据合成		ESPS ^[104]	CDP	设计了一种层次图模型,用于在轨迹数据发布过程中保留语义信息
		SOK ^[105]	CDP	设计了轨迹数据保护效果的标准化评估工具
		PureGL ^[106]	CDP	将特殊的拉普拉斯噪声注入重要位置的局部和全局频率分布,以保护隐私
		DSG-BTra ^[107]	CDP	评估用户行为的差分隐私敏感性,以量化用户行为面临的隐私风险
		DPTraj-PM ^[108]	CDP	结合前缀树和m阶马尔可夫过程进行轨迹建模,并为建模特征注入噪声
		DPST ^[109]	CDP	将原始轨迹数据建模为后缀树结构,并将噪声注入建模的后缀树中

(adaptive grids, AG) 划分方法,通过选择合适的分区粒度来构建划分空间的网格,从而在噪声误差和非均匀性误差之间取得平衡,提高查询精确度。具体来说,首先,在空间域上构建一个粗粒度的网格,并统计每个单元格内带有拉普拉斯噪声的位置点数量;然后,对计数密集的单元格进行细粒度的划分,对计数稀疏的单元格则保持粗粒度的划分。特别地,一级网格由 $m_1 \times m_1$ 个

均匀网格构成, m_1 可以计算为

$$m_1 = \max(10, 1/4 \lceil N\epsilon/c \rceil) \quad (10)$$

其中, N 是位置点的数量, ϵ 是 AG 的总体隐私预算, c 是一个取决于数据集的常数(本文实验 $c=10$), N' 是注入噪声的位置点总计数。

Zhang 等^[44]认为 PSD 在空间划分过程中过于依赖预定义的递归深度。预定义的递归深度值太小会导致单元格划分粒度过粗,预定义的递归深

度值过大(引入的噪声过多)会导致单元格划分粒度过细。为此,他们设计了一种新的空间分解算法 PrivTree,采用分层方式消除了对预定义的递归深度的依赖。在分层分解背景下,PrivTree 可以使用 $O(1)$ 级别的噪声来发布 0/1 值序列,而不必考虑该序列的敏感度。基于此,PrivTree 能够通过恒定噪声决定单元格是否需要进一步分解,从而在细粒度单元格分解时不必担心递归深度。真实数据的实验结果表明,PrivTree 在直方图查询性能方面显著优于现有方法。

3.2 面向数据发布的轨迹数据隐私保护方法

3.2.1 位置扰动

尽管数据加噪的方法允许数据收集者在保护隐私的前提下提供查询服务,但这通常需要根据不同的查询设置不同的隐私保护算法,给数据消费者带来了使用上的不便。对此,一种有效的解决方案是对轨迹中的每个位置进行扰动,并发布这些扰动后的数据。此时,数据消费者可以直接在发布的轨迹数据上进行分析,无须修改现有的数据分析算法。Andres 等^[45]提出了一种针对位置数据的差分隐私概念——地理不可区分性(geo-indistinguishability, GI)。该概念定义了一个直观的隐私保护方法:在以用户为中心、半径为 r 的圆形区域内保护用户的位置,隐私保护水平与扰动距离相关。根据 GI 的定义,Andres 等^[45]提出了一种平面拉普拉斯(planar Laplace, PL)机制,即利用拉普拉斯分布的特定平面版本来扰动用户的位置,从而实现地理不可区分性。然而,GI 在保护用户真实位置的同时,会导致数据消费者获取的数据质量下降。为此,Bordenabe 等^[46]引入了“质量损失”概念,旨在 GI 提供的隐私保护与数据质量损失之间寻求平衡。

然而,对轨迹中的每个位置进行扰动会消耗隐私预算,进而引入大量噪声,导致轨迹数据的准确度显著下降。为此,Chatzikokolakis 等^[47-48]提出了轨迹混淆机制(trajec-tory obfuscation mechanism, TOM),旨在以较少的隐私预算实现每个位置数据的扰动。TOM 包含预测函数、噪声机制和测试机制。首先使用预测函数根据之前的子轨迹预测下一个位置,然后应用测试机制

判断预测位置是否在用户当前位置的阈值内。如果在阈值内,那么报告预测位置;否则,应用噪声机制生成一个新位置。TOM 仅在测试或应用噪声机制时消耗隐私预算,并设计了隐私预算管理器等以进一步节省预算。Ma 等^[51]提出了自适应扰动机制 AGENT。AGENT 利用测试机制和 R-Tree 计数实现对抗扰动位置的重复使用,从而减少隐私预算的消耗。不同于 TOM,当 AGENT 发现符合测试条件的扰动位置时,会直接报告该位置。

此外,由于轨迹中的位置具有时空关联性,攻击者通过分析一系列扰动后的位置数据可能推断出原始轨迹,从而将特定的运动行为与特定用户关联起来^[47-50]。Shao 等^[49]利用位置关联性从扰动后的轨迹数据中恢复用户的真实位置。为此,Xiao 等^[52-53]提出了隐私定义 δ -location set,这一定义结合了 CDP 和位置间的相关性。他们还提出了对应的位置扰动机制 PIM,旨在最大化扰动数据的可用性。PIM 将隐马尔可夫模型和差分隐私技术相结合,用来保护轨迹中的每个位置,防止推理攻击。Yu 等^[36]提出了动态位置隐私框架 PIVE,旨在同时实现地理不可区分性和抵御推理攻击。PIVE 首先根据用户定义的推理错误阈值和先验知识确定一个位置子集作为保护位置集合,然后使用指数机制对用户的真实位置进行扰动,使其落在保护位置集合中的一个位置。Al-Dhubhani 等^[54]提出了自适应隐私保护机制 AGM,该机制根据用户位置与之前扰动位置的相关性调整噪声量,以确保扰动后的轨迹数据能够抵御推理攻击。

此外,Cao 等^[55-59]系统分析了 CDP 在轨迹位置间时间相关性所引发的潜在隐私损失,即时间性隐私泄露(temporal privacy leakage, TPL),并指出 TPL 可能随着轨迹长度的增加而加剧。为了解决这一问题,他们设计了一种高效算法来计算 PriSTE,并开发了相应的数据扰动机制,以将任何现有的 CDP 机制转化为针对 TPL 的保护机制。Jin 等^[60]提出了隐私保护轨迹发布方案 GL,它通过注入拉普拉斯噪声来扰乱轨迹中重要位置的局部/整体频率分布,并使用这些分布去修改用户的原始轨迹。

3.2.2 轨迹合成

上述方法主要通过噪声注入的扰动机制来解决连续位置扰动时轨迹数据可能遭受的隐私泄露问题,但这些方法仍然面临身份和记录链接攻击的风险。为了弥补不足,最新的研究思路是采用数据合成的方式来保护用户的轨迹数据。这类方法的主要思想是统计真实数据的总体统计特征,然后向这些特征注入满足 ϵ -CDP的噪声,并使用这些特征重新构建合成轨迹数据。显然,合成的轨迹数据并不直接依赖于真实数据集中的任何特定轨迹。这确保了输出轨迹与任何真实轨迹(输入)之间的不可链接性,从而可以有效抵御身份和记录链接攻击。因此,合成的轨迹数据在统计学上与真实数据相似,但它们不会泄露参与合成过程的任何个体的信息。本质上,这种方式是对满足 ϵ -CDP的查询结果进行二次加工,即在含有噪声的查询结果上重新构建用户的数据。

Chen等^[61-62]提出了一种序列数据扰动方法SDS。该方法利用自然语言处理中成熟的 N -gram模型来统计原始数据的特征并注入噪声,然后使用这些带噪声的 N -gram模型重新构建合成轨迹数据。此外,他们提出了一种从带噪声的 N -gram模型中合成轨迹数据集的方法。He等^[63-64]提出了一种差分隐私轨迹数据发布方案DPT。该方案首先采用层次参考系统模型来精确捕捉轨迹中相邻位置的相关性和速度特征,然后在每个参考系统中构建前缀树并向其注入拉普拉斯噪声,最后从带有噪声的树中重新构建合成轨迹数据。Gursoy等^[65]提出了一种轨迹数据合成方案AdaTrace。该方案通过特征提取、概要学习、隐私保护和效用保护的噪声注入4个环节,从真实轨迹数据中构建生成模型,生成保护隐私的合成轨迹数据。基于这一流程,Gursoy等^[66]进一步扩展了该研究工作,通过对输入数据进行归一化处理,以生成具有更高可用性的合成轨迹数据。Wang等^[70]提出了一种隐私保护轨迹合成方法PrivTrace,该方法结合了DPT和AdaTrace的优点。首先,PrivTrace结合了一阶和二阶马尔可夫链模型,以更好地平衡数据精度和引入的噪声。其次,PrivTrace优化了AdaTrace中的行程分布,以在较少的隐私预算下保留原始轨迹数据中的行

程特征。Ghane等^[67]提出了混合指数机制和拉普拉斯机制的隐私轨迹发布方法TGM。TGM首先使用指数机制扰动原始轨迹中的每一个位置,然后利用拉普拉斯机制恢复被扰动位置的方向,从而提高发布数据的可用性。Cai等^[68]提出了隐私轨迹数据库发布算法DPTD,首先对原始位置进行聚类以离散化空间,然后构建一个带有噪声的前缀树以发布差分隐私的轨迹数据。该算法尽管基于马尔可夫模型,且仅在前缀树的奇数层添加噪声以减少误差,但是树结构较深仍可能会引入过多的噪声。

Sun等^[69]提出了一种使用差分隐私合成实际轨迹数据的方法SPRT,其核心思想是将公开地理空间的结构整合到隐私轨迹合成过程中,从而使SPRT能够更精确地捕捉移动特征。他们通过构建的地理感知网格来划分轨迹数据集覆盖的地理空间,这不仅可以捕捉细粒度的用户移动,还可以避免不切实际的用户移动出现,从而大幅度提升了提取的差分私有特征的准确率。另外,他们基于现实世界的交通规则定义了合成轨迹的移动限制,进一步提升了合成数据的可用性。在上述工作的基础上,Sun等^[100]提出了一种基于城市交通网络结构的隐私保护轨迹数据合成方法PUTS。该方法直接使用城市路网离散化地理空间,并从路径和位置两个维度提取原始数据中的差分隐私轨迹特征。为了合成接近真实的轨迹数据,PUTS给出了一个从路径到轨迹的双层数据合成方法,首先利用路径特征合成差分私有的路径,然后利用位置特征在合成的路径上重构位置点。因此,PUTS发布的隐私保护轨迹数据能够完全替代真实轨迹,并应用于现实世界的场景中,其数据质量好,与真实数据非常接近。

Du等^[104]基于拉普拉斯机制和马尔可夫链理论开发了一种新的隐私轨迹合成算法ESPS,并在ESPS算法中设计了一种层次图模型,以在轨迹数据发布过程中保留语义信息,并提出一种关联分数的度量方法来衡量两个位置之间的关系,从而更好地捕捉轨迹的地理特征。

Miranda等^[105]提出了保护性掩码度量框架SOK来为轨迹数据库提供差分隐私保护。为了实现这一目标,他们系统性地总结了轨迹数据库的

效用评价指标,并深入分析了针对轨迹数据的差分隐私粒度概念。Jin等^[106]提出了一种基于频率的随机化模型PureGL,用于发布隐私保护的轨迹数据。他们首先引入两种随机化机制为轨迹数据中重要的局部和全局频率分布注入特定的拉普拉斯噪声;然后,将轨迹重构任务表述为kNN搜索问题,并设计了两个分层索引以在带有噪声的频率分布上高效地重构轨迹数据。

Qiu等^[107]提出了一种为用户生成差分语义泛化行为轨迹的方法DSG-BTra。该方法的原理是,首先形式化用户的社会流动性以探索潜在的用户行为模式,并评估用户行为模式的隐私敏感度;然后,他们使用用户敏感性量化强度概括用户的行为语义,并基于此为用户参与的移动服务生成差分语义泛化行为轨迹。Wang等^[108]提出了在CDP下合成轨迹数据的方法DPTraj-PM。基于个人轨迹主要由初始轨迹段和下一个位置点决定的假设,该方法首先将原始轨迹数据离散化为邻近单元,然后结合前缀树结构和 m 阶马尔可夫过程对其进行建模,最后为建模后的特征注入差分隐私噪声,并根据噪声规模合成隐私保护的轨迹数据集。Tang等^[109]提出一种分布式预测后缀树构建解决方案DPST,用于在多方环境下发布差分

隐私的合成序列数据集。DPST的原理是,首先将原始轨迹数据建模为后缀树结构,然后将拉普拉斯噪声注入建模的后缀树结构中,从而确保在隐私保护的前提下生成高质量的合成轨迹数据。

4 基于LDP的轨迹数据隐私保护方法

本节从应用和数据扰动两个层面评述基于LDP的轨迹数据隐私保护方法。已有的基于LDP的轨迹数据保护方法总结见表3。

4.1 快照式LBS的轨迹数据隐私保护方法

4.1.1 位置扰动

在快照式LBS中,每个用户需要向数据收集者报告其在一定时间内的轨迹。为了保护用户隐私,一种直接的方法是对用户轨迹中的每个位置进行独立扰动。面向频率估计的一些LDP协议(如GRR)^[76-77,80-81]可以用于收集用户的轨迹数据,并通过自适应调整来评估各位置的频率。以经典的GRR为例,数据收集者可以将地理空间划分为网格,并要求每个用户将其轨迹中的每个位置编码为所在小区的编号。接着,用户使用GRR的扰动算法对编码数据进行扰动,并将处理后的数据上传给数据收集者。数据收集者在收集到用户上传的扰动数据后,可以使用GRR的聚合算法从这些数据中评估各位置的访问频率。

表3 已有的基于LDP的轨迹数据隐私保护方法总结

Table 2 Summary of existing methods of trajectory data privacy protection based on LDP

应用场景	扰动方式	算法	隐私模型	特点
快照式LBS	位置扰动	GRR, OUE ^[76]	LDP	将用户的轨迹编码为一些数值,然后对编码后的数值进行扰动
		PM, HM ^[82]	LDP	通过限制每个地点的经纬度输入范围来降低扰动数据的误差
		SM ^[83-84]	LDP	设计了一个二维联合概率性分布,并从中输出一个扰动的位置
		LE ^[78-79]	Local d-privacy	以一种与距离有关的方式为不同的输入和输出分配不同的扰动概率
		PCE ^[85]	PLDP	允许用户指定其安全区域(用户位置的扰动域)和隐私预算
	轨迹重构	OBE ^[110]	LDP	提出了一种优化的单一编码协议,用于LDP下的比特向量编码
		EM ^[15]	LDP	将所有轨迹之间的时空距离用来设定指数机制的概率分布
		NGram ^[93]	LDP	使用轨迹的重叠片段,以高效且保护隐私的方式捕捉移动模式
		ATP ^[101]	LDP	自适应地限制每个位置的扰动区域,以获得准确的移动模式
		LDPTTrace ^[102]	LDP	在LDP框架下收集轨迹特征,并使用这些特征合成轨迹数据
持续式LBS	位置扰动	PRIMA ^[111]	LDP	采用自定义隐私预算分配策略处理轨迹长度不均问题
		DADP ^[92]	w-event DP	在用户和不被信任的服务器之间引入了一个新的多重代理层级
		LDP-IDS ^[98]	w-event DP	提出了一个新颖的隐私预算分割方法
		RetraSyn ^[112]	LDP	设计了动态移动更新机制,确保实时合成的轨迹数据能够反映用户的真实行为和移动模式

Tire 等^[110]提出了一种 LDP 位置分布收集解决方案 OBE。该方案的收集过程包含分区、敏感度计算、用户端噪声响应计算及服务器端估计 4 个环节。具体来说,利用新设计的一种兼顾采样误差和噪声误差的高级分区策略进行分区,结合查询集图模型和顶点着色技术计算分区的敏感度,利用新设计的一种支持任意位向量编码的 LDP 协议完成用户端噪声响应计算和服务器端估计。

然而,OBE 容易引入大量噪声,导致从收集的数据中得到的频率估计被噪声覆盖。以 GRR 为例,每个真实位置被保留下来的概率为 $e^{\epsilon/k}/(d-1+e^{\epsilon/k})$,其中 d 表示网格中单元格的总数, k 为用户轨迹的长度, ϵ 为 LDP 机制的隐私预算。假设有一个 5×5 的网格,LDP 机制的隐私预算为 1,用户的轨迹长度为 5,那么每个真实位置扰动后被保留的概率仅为 4.84%。目前,解决这个问题的思路主要有两种。

1) 改进扰动机制:将轨迹数据中的每个位置视为包含经度和纬度的二维数据,从而摆脱网格限制,直接对位置的经纬度进行扰动。基于这一思路,Wang 等^[82]设计了分片机制 (piecewise mechanism, PM) 和混合机制 (hybrid mechanism, HM),在 LDP 下收集每个维度的数据,并通过限定多维数据中每个维度的输入范围来减少扰动后数据的误差。然而,当使用这些机制对用户位置进行扰动时,由于每个维度的值都是独立扰动的,扰动后的位置常常在一个维度上与原始位置相似,但在另一个维度上却有较大偏差。为了解决这些问题,Hong 等^[83-84]提出一种方形机制 (square mechanism, SM)。该机制通过从一个二维联合概率分布中输出一个扰动的位置来实现隐私保护,靠近原始位置的方形区域内的概率密度较高,其他区域的概率密度则较低。

2) 放缩隐私模型:设计 LDP 模型的广义 (放缩) 版本,以寻求隐私和效用之间更好的平衡。具体来说,通过放缩 LDP 的定义以控制数据域中不均匀的扰动概率,从而提升真实信息被报告的几率。Chatzikokolakis 等^[78]基于距离度量提出了一种本地 d-privacy 的隐私模型。该模型通过引入距离度量,将隐私保护与输入对之间的距离相关联,从而提供优越的效用,尤其是在范围

查询中。Alvim 等^[94]进一步扩展了这一概念,将其应用于各种度量和分布式架构。对于轨迹数据,通常选择欧氏距离作为距离度量。基于本地 d-privacy 的概念,Gu 等^[79]提出了一种扰动机制 LE,旨在有效支持涉及个体信息的范围查询和总体信息的频率估计。该机制在随机响应过程中根据输入和输出的距离分配不同的扰动概率,既保留了随机响应在频率估计中的优势,又实现了范围查询的有效性。

Chen 等^[85]提出了个性化本地差分隐私 (personalized local differential privacy, PLDP),允许不同的用户有不同的隐私期望或偏好。他们为 PLDP 设计了一个高效的实现协议 PCE,并从理论上分析了其实用性、隐私性和复杂性。在 PCE 的基础上,他们进一步提出了隐私位置数据聚合框架 PSDA,该框架允许不受信任的服务器在满足每个用户的 ϵ -PLDP 要求的同时,准确了解用户在空间域的分布情况。同样,Gu 等^[96]提出了一种输入鉴别型 LDP 模型,为每个数据点分配其独特的隐私级别。这一模型更好地反映了不同数据点在隐私保护需求上的差异性。Acharya 等^[95]提出了上下文感知本地差分隐私 (context-aware LDP) 模型,该模型允许在每对元素之间设置不同的敏感度级别,以整合上下文信息。这个模型的定义使得用户可以根据背景信息为每对位置设定不同的敏感度,从而在隐私保护和数据效用之间找到更好的平衡。

4.1.2 轨迹重构

除了位置扰动外,一些研究人员选择对用户的轨迹数据进行整体扰动。具体来说,每条轨迹可以被视为高维空间中的一个点。通过枚举所有可能的轨迹,计算这些高维点与真实轨迹之间的距离,并基于该距离确定指数机制中的概率分布。用户可以根据这个概率分布扰动自己的轨迹,然后将处理后的轨迹数据上传给不可信的数据收集者。然而,这种方法要求实例化一个包含所有可能轨迹的数据域,并且该数据域的尺寸会随着轨迹长度的增加呈指数增长。在实际应用中,这种大规模数据域会导致指数机制在计算和存储方面的不可行性。

为了解决这个问题,Cunningham 等^[93]提出

了本地轨迹扰动方法NGRAM。该方法不再将轨迹视为高维点，而是通过使用轨迹的重叠片段来捕捉时空模式，从而实现高效的隐私保护计算。他们设计了分层分解机制以缩减轨迹片段数据域，并创建了基于时空距离的质量函数，用于构建指数机制的扰动概率密度。然而，由于LDP的严格隐私定义，数据的扰动空间仍然过大，因此输出轨迹中包含大量错误信息。Zhang等^[101]提出了一种仅依赖轨迹位置的扰动机制ATP，通过自适应地限制每个位置的扰动区域，以获取准确的用户移动模式。在同一时期，Du等^[102]开发了LDPTrace，其目标是在本地化差分隐私框架下收集轨迹特征，并使用这些特征来合成轨迹数据。LDPTrace允许用户独立地扰动并上传自己的移动模式特征。随后，它会汇总所有用户上传的移动模式特征，并将这些特征整合为马尔可夫模型用于合成轨迹数据。Guner等^[111]提出了基于LDP的离散移动特征提取方法PRIMA。在PRIMA中，每个用户使用LDP协议在其本地设备上对自己的轨迹数据进行编码和扰动，上传扰动后的数据给数据收集者以供其提取移动特征。特别地，PRIMA中还设计了自定义隐私预算分配策略以实现不均匀的隐私预算分配，以便在轨迹长度不均匀的情况下提升扰动后数据的效用。

4.2 持续式LBS的轨迹数据隐私保护方法

在持续式位置服务中，用户除了直接上传完整的轨迹外，还需持续上传其轨迹数据。例如，在众包服务中，工人需要持续报告自己的位置。Hu等^[112]提出了实时轨迹合成框架RetraSyn，以实时地从用户轨迹流中隐私提取的移动模式上合成轨迹数据。RetraSyn利用LDP在用户端对敏感数据进行扰动，从而在共享或分析之前保护用户隐私。特别地，他们设计了动态移动更新机制，确保实时合成的轨迹数据能够反映用户的真实行为和移动模式，从而提高数据的真实性和实用性。

尽管上述方法可以扰动用户轨迹中的位置，但由于轨迹数据长度不确定，难以确定每次扰动所需的隐私预算。特别是当用户报告的轨迹数据长度无限时，这种方法将不再适用。

针对无限流数据的差分隐私保护问题，Kellaris等^[88]提出了一种窗口级差分隐私模型w-

event DP，旨在平衡无限流数据的隐私性和可用性。w-event DP依赖于相邻数据流的定义。

定义9 相邻数据流。令 w 为正整数，两个数据流前缀 S_t, S'_t 为相邻的条件为

(1) 对于每个 $S_t[i], S'_t[i]$ ，如果 $i \in [t]$ 且 $S_t[i] \neq S'_t[i]$ ，则认为 $S_t[i], S'_t[i]$ 是相邻的。

(2) 对于每个 $S_t[i_1], S_t[i_2], S'_t[i_1], S'_t[i_2]$ 且满足 $i_1 < i_2$ ， $S_t[i_1] \neq S'_t[i_1], S'_t[i_2] \neq S'_t[i_2]$ ， $i_2 - i_1 + 1 \leq w$ 。

w-event DP的正式定义如下。

定义10 w-event DP。假设算法 A 输入一个任意长度的数据流前缀 S_t 。Range(A)为 A 的所有可能输出的集合， A 满足w-event DP，当且仅当对于所有w-相邻数据流 S_t, S'_t ，所有可能的输出 $R \subseteq \text{Range}(A)$ 和 t ，下列不等式成立。

$$\Pr[A(S_t) \in R] \leq e^\epsilon \cdot \Pr[A(S'_t) \in R]$$

在遵循w-event DP的算法中，任意连续的 w 时间窗口内的敏感数据得到保护，防止攻击者通过获取连续时间的数据发布推断敏感信息。这个隐私定义类似于中心化差分隐私，依赖于可信的数据收集者，这导致基于w-event DP的相关工作无法用于本地式架构。

为了解决这个问题，Wang等^[92]基于分布式代理的w-event DP保护框架DADP，在用户和不可信任的数据收集者之间引入了一个新的多重代理层级。用户可以随机选择一个代理，并通过匿名连接技术将签到信息上传给该代理，而不是直接上传给不可信任的数据收集者。每个代理对收到的位置数据进行汇总，并使用拉普拉斯机制对汇总的统计数据数据进行本地扰动。然后，来自所有代理的扰动统计数据被进一步组合在一起，形成整个扰动统计数据以供发布。Ren等^[98]提出了一种新的基于w-event DP的隐私保护方法LDP-IDS，可以为无限流提供 ϵ -LDP的隐私保护。他们为LDP在无限流上的隐私预算划分开发了一个框架，该框架包含了两种根据非确定性流的稀疏性自适应分配隐私预算的算法。此外，他们为LDP-IDS提出了一个基于人群划分的方法，以实现更高的效用和更低的通信开销。

5 结束语

在大数据时代, 个人的轨迹数据隐私问题日益受到重视。差分隐私作为数据隐私保护的事实标准, 将成为推动数据科学进一步发展的关键技术。

尽管目前基于差分隐私的轨迹数据隐私保护方法在部分真实轨迹数据上表现出色, 但现有的研究依然存在很大的局限性, 限制了这些方法的进一步应用。

如何更好地设计隐私模型和优化隐私机制, 在保护用户数据隐私的同时, 将用户的数据分享给数据消费者以推动数据科学的进一步发展是一个极具挑战性的问题。本文综述了面向轨迹数据的差分隐私保护方法, 从不同层面分析了现有方法的研究方向, 并探讨了轨迹数据隐私保护中可尝试和实施的研究方向, 希望能为工业界和学术界的相关研究提供启发。

未来的研究工作将聚焦于超大规模的隐私保护轨迹数据发布、高质量隐私保护轨迹数据的采集。以及针对隐私保护轨迹数据采集的投毒攻击。

(1) 超大规模的隐私保护轨迹数据发布

海量的轨迹数据被用于训练更有效的深度学习模型, 以促进智能交通、自动驾驶等领域的发展。然而, 隐私保护问题限制了大规模的轨迹数据共享, 阻碍了这些领域的进一步发展。现有基于 CDP 的轨迹数据发布方法仍难以发布超大规模的数据。这是因为现有方法仍局限于使用简单的线性模型来提取原始数据的特征和合成隐私保护的轨迹, 从而无法支持超大规模的 (PB 级) 轨迹数据发布。

目前, 极具前景的解决方案是利用日渐成熟的深度学习模型来实现大规模隐私保护轨迹数据的发布。它能够挖掘和保留原始数据中的深层次特征, 提升发布数据的效用。然而, 将差分隐私技术应用于深度学习仍然面临着巨大挑战。

当前的基于差分隐私的深度学习技术面临着隐私保护能力不足的问题。现有研究通常使用高斯机制对深度学习模型中的梯度注入噪声, 从而保证攻击者不能从深度学习模型中推断出用户的隐私信息。然而, 高斯机制只能实现 (ϵ, δ) -CDP (一种放缩后的 CDP), 这意味着隐私泄露

的可能性仍然存在, 且与 δ 值成正比。换句话说, 这类方法为用户提供的数据隐私保护能力是不足的。

当前的基于差分隐私的深度学习技术面临着效用受限问题。现有研究工作通过直接对深度学习模型的梯度注入数值型噪声来实现隐私保护, 然后使用梯度修剪控制注入噪声的规模。然而, 这种噪声注入会覆盖全部的梯度参数, 大量噪声的注入导致深度学习模型难以收敛。

(2) 高质量隐私保护轨迹数据的采集

近年来, 用户隐私数据泄露事件层出不穷, 用户越来越不愿意分享自己的数据。尽管 LDP 技术可以在保护用户隐私的同时收集用户数据, 但它在收集轨迹数据时仍然面临着巨大挑战。

轨迹数据是一种时空相关、高维度的数据, 现有的 LDP 机制^[106-108, 110, 112]为了保护隐私会分割隐私预算 (即引入更剧烈的扰动), 扰动数据的每个维度, 从而导致收集的轨迹数据中含大量的错误信息。

尽管一些方法^[75, 94-95]试图通过考虑地理空间距离等因素来调整每个位置的扰动概率, 以提高保留真实信息的可能性, 但由于位置域的广泛性, 这些方法仍有很大概率将位置扰动到与原始位置无语义关系的其他位置。

(3) 数据投毒攻击挑战

轨迹数据采集本质上是一种分布式设定, 因此非常容易遭受数据投毒攻击。攻击者可以通过控制一部分正常用户 (这部分用户被称为恶意用户) 向数据收集者发送精心制作的恶意数据, 从而操纵数据收集的成果。以百度地图为例, 攻击者可以控制十几个使用百度地图 App 的设备在同一地点使用导航功能并保持不动, 导致百度地图显示该地点交通拥堵, 从而误导其他用户。当这种投毒攻击针对基于 LDP 的轨迹数据采集时, 处理起来非常困难。

LDP 扰动算法会向用户的轨迹数据注入噪声, 攻击者可以利用这些噪声掩护恶意数据, 使得数据收集者更加难以识别恶意数据。

攻击者可以在轨迹数据的几个维度随机注入虚假信息, 这也会增加数据收集者处理恶意数据的难度。

参考文献:

- [1] WANG Y, YIN H, CHEN H, et al. Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 1227-1235.
- [2] WEI H, ZHENG G, YAO H, et al. Intellilight: a reinforcement learning approach for intelligent traffic light control[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2018: 2496-2505.
- [3] SUN X, YE Q, HU H, et al. Generating location traces with semantic-constrained local differential privacy[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 9850 - 9865.
- [4] 贾靖宇, 谭畅, 刘哲伟, 等. 差分隐私的隐私观与隐私目标[J]. 网络与信息安全学报, 2023, 9(5): 82-91.
JIA J Y, TAN C, LIU Z W, et al. Privacy view and target of differential privacy[J]. Chinese Journal of Network and Information Security, 2023, 9(5): 82-91.
- [5] WANG H, GAO C, WANG G, et al. De-anonymization of mobility trajectories: dissecting the gaps between theory and practice[C]// Proceedings of the 25th Network and Distributed System Security Symposium. 2018.
- [6] EDDY N. Location-based applications popular, despite privacy concerns: ISACA[EB]. 2012.
- [7] General Data Protection Regulation (GDPR)[EB]. 2016.
- [8] 中华人民共和国个人信息保护法[EB]. 2021.
Personal information protection protection law of the People's Republic of China[EB]. 2021.
- [9] SAMARATI P, SWEENEY L. k -anonymity: a model for protecting privacy[C]// Proceedings of the IEEE Symposium on Research in Security and Privacy (S&P). 1998.
- [10] GEDIK B, LIU L. Location privacy in mobile systems: a personalized anonymization model[C]// Proceedings of the IEEE International Conference on Distributed Computing Systems. 2005: 620-629.
- [11] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]// Proceedings of the International Conference on Mobile Systems, Applications and Services. 2003: 31-42.
- [12] KALNIS P, GHINITA G, MOURATIDIS K, et al. Preventing location-based identity inference in anonymous spatial queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(12): 1719-1733.
- [13] MOKBEL M F, CHOW C Y, AREF W G. The new casper: query processing for location services without compromising privacy[C]// Proceedings of the International Conference on Very Large Data Bases. 2006: 763-774.
- [14] LI N H, LI T C, VENKATASUBRAMANIAN S. T-closeness: privacy beyond k -anonymity and l -diversity[C]// Proceedings of the IEEE International Conference on Data Engineering. 2007: 106-115.
- [15] DOURIEZ M, DORAISWAMY H, FREIRE J, et al. Anonymizing nyc taxi data: does it matter[C]// Proceeding of the IEEE International Conference on Data Science and Advanced Analytics. 2016: 140-148.
- [16] MA C Y, YAU D K, YIP N K, et al. Privacy vulnerability of published anonymous mobility traces[J]. IEEE/ACM Transactions on Networking, 2013(3), 21: 720-733.
- [17] ZANG H, BOLOT J. Anonymization of location data does not work: a large-scale measurement study[C]// Proceedings of the International Conference on Mobile Computing and Networking. 2011: 145-156.
- [18] KHOSHGOZARAN A, SHAHABI C. Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy[C]// Proceedings of the International Symposium on Spatial and Temporal Databases. 2007: 239-257.
- [19] MARIAS G F, DELAKOURIDIS C, KAZATZOPOULOS L, et al. Location privacy through secret sharing techniques[C]// Proceedings of the International Symposium on a World of Wireless Mobile and Multimedia Networks, Taormina-Giardini Naxos. 2005: 614-620.
- [20] GHINITA G, KALNIS P, KHOSHGOZARAN A, et al. Private queries in location based services: anonymizers are not necessary[C]// Proceedings of the 2008 ACM SIGMOD International Conference on Management of data. 2008: 121-132.
- [21] DWORK C. Differential privacy[C]// Proceedings of the International Colloquium on Automata, Languages and Programming. 2006: 1-12.
- [22] DWORK C, MCSHERRY F, NISSIM K, et al. Calibrating noise to sensitivity in private data analysis[C]// Proceedings of the Theory of Cryptography Conference. 2006: 1-12.
- [23] DWORK C, ROTH A. The algorithmic foundations of differential privacy[J]. Foundations and Trends in Theoretical Computer Science, 2014, 9: 211-407.
- [24] TANG J, KOROLOVA A, BAI X, et al. Privacy loss in apple's implementation of differential privacy on MACOS 10.12[J]. arXiv. preprint arXiv: 1709. 02753. 2017.
- [25] Apple. Differential privacy team, learning with privacy at scale[EB]. 2017.
- [26] ERLINGSSON Ú, PIHUR V, KOROLOVA A. Rappor: Randomized aggregatable privacy-preserving ordinal response[C]// Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. 2014: 1054-1067.
- [27] JIANG H, LI J, ZHAO P, et al. Location privacy-preserving mechanisms in location-based services: a comprehensive survey[J]. ACM Computing Surveys, 2021, 54(1): 1-36.
- [28] JIN F, HUA W, FRANCIJA M, et al. A survey and experimental study on privacy-preserving trajectory data publishing[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(6): 5577-5596.
- [29] SUN S, MA S, SONG J H, et al. Experiments and analyses of anonymization mechanisms for trajectory data publishing[J]. Journal of

- Computer Science and Technology, 2022, 37(5): 1026-1048.
- [30] WANG H, WANG E, YANG Y, et al. Privacy-preserving online task assignment in spatial crowdsourcing: a graphbased approach [C]// Proceedings of IEEE INFOCOM. 2022: 570-579.
- [31] ZHANG P, HU C, LI H, et al. ShiftRoute: achieving location privacy for map services on smartphones[J]. IEEE Transactions on Vehicular Technology, 2018, 67(5): 4527-4538.
- [32] ZHENG Y. Trajectory data mining: an overview[J]. ACM Transactions on Interactive Intelligent Systems, 2015, 6(3): 1-41.
- [33] Didi Chuxing[EB]. 2024.
- [34] GEDIK B, LIU L. Protecting location privacy with personalized k -anonymity: architecture and algorithms[J]. IEEE Transactions on Mobile Computing, 2007, 7(1): 1-18.
- [35] VU K, ZHENG R, GAO J. Efficient algorithms for k -anonymous location privacy in participatory sensing[C]//Proceedings of IEEE INFOCOM. 2012: 2399-2407.
- [36] YU L, LIU L, PU C. Dynamic differential location privacy with personalized error bounds[C]//Proceedings of the IEEE International Conference on Distributed Computing Systems. 2017.
- [37] SHOKRI R, THEODORAKOPOULOS G, PAPADIMITRATOS P, et al. Hiding in the mobile crowd: Location privacy through collaboration[J]. IEEE Transactions on Dependable and Secure Computing, 2014, 11(3): 266-279.
- [38] MCKENNA R, SHELDON D, MIKLAU G. Graphical-model based estimation and inference for differential privacy[C]//Proceedings of the International Conference on Machine Learning. 2019: 4435-4444.
- [39] ZHANG J, CORMODE G, PROCOPIUC C M, et al. Privbayes: private data release via bayesian networks[J]. ACM Transactions on Database Systems, 2017, 42(4): 1-41.
- [40] ZHANG Z, WANG T, LI N, et al. PrivSyn: differentially private data synthesis[C]//Proceedings of the 30th USENIX Security Symposium (USENIX Security 21). 2021: 929-946.
- [41] MCSHERRY F D. Privacy integrated queries: an extensible platform for privacy-preserving data analysis[C]//Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. 2009: 19-30.
- [42] CORMODE G, PROCOPIUC C, SRIVASTAVA D, et al. Differentially private spatial decompositions[C]//Proceedings of the 28th International Conference on Data Engineering, Arlington. 2012: 20-31.
- [43] QARDAJI W, YANG W, LI N. Differentially private grids for geospatial data[C]//Proceedings of the 29th International Conference on Data Engineering, Brisbane. 2013: 757-768.
- [44] ZHANG J, XIAO X, XIE X. Privtree: a differentially private algorithm for hierarchical decompositions[C]//Proceedings of the 2016 International Conference on Management of Data. 2016: 155-170.
- [45] ANDRES M E, BORDENABE N E, CHATZIKOKOLAKIS K, et al. Geo-indistinguishability: differential privacy for location-based systems[C]//Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security. 2013: 901-914.
- [46] BORDENABE N E, CHATZIKOKOLAKIS K, PALAMIDESSI C. Optimal geo-indistinguishable mechanisms for location privacy[C]// Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. 2014: 251-262.
- [47] CHATZIKOKOLAKIS K, PALAMIDESSI C, STRONATI M. A predictive differentially-private mechanism for mobility traces[C]// Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium. 2014: 21-41.
- [48] CHATZIKOKOLAKIS K, PALAMIDESSI C, STRONATI M. Constructing elastic distinguishability metrics for location privacy[C]// Proceedings on Privacy Enhancing Technologies. 2015: 156-170.
- [49] SHAO M, LI J, YAN Q, et al. Structured sparsity model based trajectory tracking using private location data release[J]. IEEE Transactions on Dependable and Secure Computing, 2020, 18(6): 2983-2995.
- [50] OLTEANU A M, HUGUENIN K, SHOKRI R, et al. Quantifying interdependent privacy risks with location data[J]. IEEE Transactions on Mobile Computing, 2016, 16: 829-842.
- [51] MA X, MA J, LI H, et al. AGENT: an adaptive geo-indistinguishable mechanism for continuous location-based service[J]. Peer-to-Peer Networking and Applications, 2018, 11(3): 473-485.
- [52] XIAO Y, XIONG L. Protecting locations with differential privacy under temporal correlations[C]//Proceedings of the 2015 ACM SIGSAC Conference on Computer and Communications Security. 2015: 1298-1309.
- [53] XIAO Y, XIONG L, ZHANG S, et al. Loclok: location cloaking with differential privacy via hidden Markov Model[C]//Proceedings of the VLDB Endowment. 2017: 1901-1904.
- [54] AL-DHUBHANI R, CAZALAS J M. An adaptive geo-indistinguishability mechanism for continuous LBS queries[J]. Wireless Networks, 2018, 24(8): 3221-3239.
- [55] CAO Y, YOSHIKAWA M, XIAO Y, et al. Quantifying differential privacy under temporal correlations[C]//Proceedings of the 33rd International Conference on Data Engineering. 2017: 821-832.
- [56] CAO Y, XIONG L, YOSHIKAWA M, et al. ConTPL: controlling temporal privacy leakage in differentially private continuous data release[C]//Proceedings of the VLDB Endowment. 2018: 2090-2093.
- [57] CAO Y, YOSHIKAWA M, XIAO Y, et al. Quantifying differential privacy in continuous data release under temporal correlations[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(7): 1281-1295.
- [58] CAO Y, XIAO Y, XIONG L, et al. PriSTE: from location privacy to spatiotemporal event privacy[C]//Proceedings of the 35th International Conference on Data Engineering. 2019: 1606-1609.
- [59] CAO Y, XIAO Y, XIONG L, et al. Protecting spatiotemporal event privacy in continuous location-based services[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(8): 3141-3154.
- [60] JIN F, HUA W, RUAN B, et al. Frequency-based randomization for guaranteeing differential privacy in spatial trajectories[C]//Proceedings of the 38th International Conference on Data Engineering. 2022: 1727-1739.
- [61] CHEN R, ACS G, CASTELLUCCIA C. Differentially private se-

- quential data publication via variable-length n -grams[C]//Proceedings of the 2012 ACM Conference on Computer and Communications Security. 2012: 638-649.
- [62] CHEN R, FUNG B C M, DESAI B C, et al. Differentially private transit data publication: a case study on the montreal transportation system[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2012: 213-221.
- [63] HE X, CORMODE G, MACHANAVAJJHALA A, et al. DPT: differentially private trajectory synthesis using hierarchical reference systems[C]//Proceedings of the VLDB Endowment. 2015: 1154-1165.
- [64] HE X, RAVAL N, MACHANAVAJJHALA A. A demonstration of VisDPT: visual exploration of differentially private trajectories[C]//Proceedings of the VLDB Endowment. 2016: 1489-1492.
- [65] GURSOY M E, LIU L, TRUEX S, et al. Utility-aware synthesis of differentially private and attack-resilient location traces[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. 2018: 196-211.
- [66] GURSOY M E, LIU L, TRUEX S, et al. Differentially private and utility preserving publication of trajectory data[J]. IEEE Transactions on Mobile Computing, 2019, 18(10): 2315-2329.
- [67] GHANE S, KULIK L, RAMAMOCHANARAO K, et al. TGM: a generative mechanism for publishing trajectories with differential privacy[J]. IEEE Internet of Things Journal, 2019, 7(4): 2611 - 2621.
- [68] CAI S, LYU X, LI X, et al. A trajectory released scheme for the internet of vehicles based on differential privacy[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(9): 16534-16547.
- [69] SUN X, YE Q, HU H, et al. Synthesizing realistic trajectory data with differential privacy[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(5): 5502-5515.
- [70] WANG H, ZHANG Z, WANG T, et al. PrivTrace: differentially private trajectory synthesis by adaptive Markov model[C]//Proceedings of the 32th USENIX Security Symposium. 2023: 1649-1666.
- [71] RASKHODNIKOVA S, SMITH A, LEE H K, et al. What can we learn privately[C]//Proceedings of the 49th Annual Symposium on Foundations of Computer Science. 2008: 531-540.
- [72] KASIVISWANATHAN S P, LEE H K, NISSIM K, et al. What can we learn privately[J]. SIAM Journal on Computing, 2011, 40(3): 793-826.
- [73] DUCHI J C, JORDAN M I, WAINWRIGHT M J. Local privacy and statistical minimax rates[C]//Proceedings of the 54th Annual Symposium on Foundations of Computer Science. 2013: 429-438.
- [74] BASSILY R, SMITH A. Local, private, efficient protocols for succinct histograms[C]//Proceedings of the Forty-seventh Annual ACM symposium on Theory of Computing. 2015: 127-135.
- [75] WARNER S L. Randomized response: a survey technique for eliminating evasive answer bias[J]. Journal of the American Statistical Association, 1965, 60(309): 63-69.
- [76] WANG T, BLOCKI J, LI N, et al. Locally differentially private protocols for frequency estimation[C]//Proceedings of the 26th USENIX Security Symposium. 2017: 729-745.
- [77] FANTI G, PIHUR V, ERLINGSSON Ú. Building a RAPPOR with the unknown: privacy-preserving learning of associations and data dictionaries[C]//Proceedings on Privacy Enhancing Technologies. 2016: 41-61.
- [78] CHATZIKOKOLAKIS K, ANDRES M E, BORDENABE N E, et al. Broadening the scope of differential privacy using metrics[C]//Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium. 2013: 82-102.
- [79] GU X, LI M, CAO Y, et al. Supporting both range queries and frequency estimation with local differential privacy[C]//Proceedings of the Conference on Communications and Network Security. 2019: 124-132.
- [80] AGRAWAL R, SRIKANT R, THOMAS D. Privacy preserving OLAP[C]//Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data. 2005: 251-262.
- [81] HUANG Z, DU W. OptRR: optimizing randomized response schemes for privacy-preserving data mining[C]//Proceedings of the 24th International Conference on Data Engineering. 2008: 705-714.
- [82] WANG N, XIAO X, YANG Y, et al. Collecting and analyzing multi-dimensional data with local differential privacy[C]//Proceedings of the 35th International Conference on Data Engineering. 2019: 638-649.
- [83] HONG D, JUNG W, SHIM K. Collecting geospatial data with local differential privacy for personalized services[C]//Proceedings of the 37th International Conference on Data Engineering. 2021: 2237-2242.
- [84] HONG D, JUNG W, SHIM K. Collecting geospatial data under local differential privacy with improving frequency estimation[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(7): 6739-6751.
- [85] CHEN R, LI H, QIN A K, et al. Private spatial data aggregation in the local setting[C]//Proceedings of the 32nd International Conference on Data Engineering. 2016: 289-300.
- [86] NIE Y W, YANG W, HUANG L, et al. A utility-optimized framework for personalized private histogram estimation[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31: 655 - 669.
- [87] BINDSCHAEDLER V, SHOKRI R. Synthesizing plausible privacy-preserving location traces[C]//Proceedings of the 2016 IEEE Symposium on Security and Privacy. 2016: 546-563.
- [88] KELLARIS G, PAPADOPOULOS S, XIAO X, et al. Differentially private event sequences over infinite streams[C]//Proceedings of the VLDB Endowment. 2014: 1155-1166.
- [89] WANG Q, ZHANG Y, LU X, et al. RescueDP: real-time spatio-temporal crowd-sourced data publishing with differential privacy[C]//Proceedings of the 35th Annual IEEE International Conference on Computer Communications. 2016: 1-9.
- [90] WANG Q, ZHANG Y, LU X, et al. Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy[J]. IEEE Transactions on Dependable and Secure Computing, 2016, 15(4): 591-606.
- [91] MA Z, ZHANG T, LIU X, et al. Real-time privacy-preserving data

- release over vehicle trajectory[J]. IEEE Transactions on Vehicular Technology, 2019, 68(8): 8091-8102.
- [92] WANG Z, PANG X, CHEN Y, et al. Privacy-preserving crowd-sourced statistical data publishing with an untrusted server[J]. IEEE Transactions on Mobile Computing, 2019, 18(6): 1356-1367.
- [93] CUNNINGHAM T, CORMODE G, FERHATOSMANOGLU H, et al. Real-world trajectory sharing with local differential privacy[C]//Proceedings of the VLDB Endowment. 2021: 2283-2295.
- [94] ALVIM M, CHATZIKOKOLAKIS K, PALAMIDESSI C, et al. Local differential privacy on metric spaces: optimizing the trade-off with utility[C]//Proceedings of the 31st Computer Security Foundations Symposium. 2018: 262-267.
- [95] ACHARYA J, BONAWITZ K, KAIROUZ P, et al. Context aware local differential privacy[C]//Proceedings of the International Conference on Machine Learning. 2020: 52-62.
- [96] GU X, LI M, XIONG L, et al. Providing input-discriminative protection for local differential privacy[C]//Proceedings of the 36th International Conference on Data Engineering. 2020: 505-516.
- [97] JIANG B, LI M, TANDON R. Context-aware data aggregation with localized information privacy[C]//Proceedings of the Conference on Communications and Network Security. 2018: 1-9.
- [98] REN X, SHI L, YU W, et al. LDP-IDS: local differential privacy for infinite data streams[C]//Proceedings of the 2022 International Conference on Management of Data. 2022: 1064-1077.
- [99] EUNJOON CHO, MYERS S A, LESKOVEC J, et al. Friendship and mobility: user movement in location-based social networks[C]//Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2021: 1082-1090.
- [100] SUN X, YE Q, HU H, et al. PUTS: privacy-preserving and utility-enhancing framework for trajectory synthesization[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(1): 296-310.
- [101] ZHANG Y, YE Q, CHEN R, et al. Trajectory data collection with local differential privacy[C]//Proceedings of the VLDB Endowment. 2023: 2591-2604.
- [102] DU Y, HU Y, ZHANG Z, et al. LDPTrace: locally differentially private trajectory synthesis[C]//Proceedings of the VLDB Endowment. 2023: 1897-1909.
- [103] PELLETIER M, SAUNIER N, N Y J L, et al. Differentially private analysis of transportation data[J]. Privacy in Dynamical Systems, 2020: 131-155.
- [104] DU X, ZHU H, ZHENG Y, et al. A semantic-preserving scheme to trajectory synthesis using differential privacy[J]. IEEE Internet of Things Journal, 2023, 10(15): 13784-13797.
- [105] MIRANDA-PASCUAL À, GUERRA-BALBOA P, PARRA-ARNAU J, et al. SoK: differentially private publication of trajectory data[C]//Proceedings on Privacy Enhancing Technologies. 2023: 496-516.
- [106] JIN F, HUA W, LI L, et al. Efficient frequency-based randomization for spatial trajectories under differential privacy[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(6): 2430-2444.
- [107] QIU G, TANG G, LI C, et al. DSG-btra: Differentially semantic-generalized behavioral trajectory for privacy-preserving mobile internet services[J]. IEEE Internet of Things Journal, 2024, 11(7): 13029-13038.
- [108] WANG N, KANKANHALLI M. DPTraj-PM: differentially private trajectory synthesis using prefix tree and Markov process[J]. arXiv preprint arXiv: 2404.14106, 2024.
- [109] TANG P, CHEN R, SU S, et al. Multi-party sequential data publishing under differential privacy[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(9): 9562-9577.
- [110] TIRE E, GURSOY M E. Answering spatial density queries under local differential privacy[J]. IEEE Internet of Things Journal, 2024, 11(10): 17419-17436.
- [111] GUNER E, GURSOY M E. Learning markov chain models from sequential data under local differential privacy[C]//European Symposium on Research in Computer Security. 2023: 359-379.
- [112] HU Y, DU Y, ZHANG Z, et al. Real-time trajectory synthesis with local differential privacy[C]//Proceedings of the 40th International Conference on Data Engineering. 2024: 1685-1698.

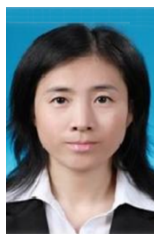
[作者简介]



孙新越 (1992-), 男, 黑龙江双鸭山人, 哈尔滨工业大学助理研究员, 主要研究方向为数据安全、数据隐私保护。



张伟哲 (1976-), 男, 黑龙江哈尔滨人, 哈尔滨工业大学教授、博士生导师, 主要研究方向为网络空间安全、并行与分布式系统。



何慧 (1974-), 女, 黑龙江哈尔滨人, 哈尔滨工业大学教授、博士生导师, 主要研究方向为网络主动防御技术、移动网络安全、云计算、迁移学习。



杨任宇 (1989-), 男, 北京人, 北京航空航天大学副教授、博士生导师, 主要研究方向为分布式系统可靠性、软件可靠性、智能计算系统。