计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202330877
61(9): 2156−2169, 2024

面向实时位置的隐私保护优化与加速求解算法

董 恺 王立夫 凌 振
(东南大学计算机科学与工程学院 南京 211189)
(dk@seu.edu.cn)

Optimizing and Accelerating Privacy Protection Algorithm for Real-Time Location

Dong Kai, Wang Lifu, and Ling Zhen
(*School of Computer Science and Engineering*, *Southeast University*, *Nanjing* 211189)

Abstract Existing electric vehicle API platforms (e.g., SmartCar), use access control mechanisms to preserve users' privacy. To preserve location privacy and meanwhile enable functionality of untrustworthy location-based services, a location privacy preserving mechanism (LPPM) can be used to generate a random pseudo-location as a reported location based on a user's true location. Existing techniques solve an optimization problem on a discrete grid, to construct an optimal LPPM which achieves the highest privacy bounded by minimum tolerable utility, or vice versa. However, they cannot be applied to real-time electric vehicle scenarios since the running time required to generate an optimal LPPM is too long (which can be several days). Another problem deals with optimality of constructed LPPMs. We reveal unexpected cases (anomaly) when the optimal LPPM constructed on a fine grid with superior granularity is worse than that on a coarse one with inferior granularity. We introduce granularity independence as a formal treatment, and propose an optimal LPPM named Divide-and-Coin which can be performed on the fly. Divide-and-Coin improves the running time from at least $O(n^{2.055})$ to $O(\log n)$, where n is the number of reported locations. Our experimental results show that Divide-and-Coin generates an optimal building-level reported location from a city-level area within one second.

Key words location privacy; optimal location privacy preserving mechanism; real-time location API; privacy-utility anomaly; granularity independence

摘 要 现有的电动汽车API平台(如SmartCar)使用访问控制机制来保护用户的隐私.为了在启用不可信位置服务功能的同时保护位置隐私,位置隐私保护机制(LPPM)根据用户的真实位置生成一个随机的伪位置作为报告位置.现有技术通过在离散网格上解决一个最优化问题构建一个最佳的LPPM,该最佳LPPM实现了在最低可容忍效用限制下的最高隐私,反之亦然.然而,它们很难直接应用于电动汽车等实时场景,因为生成最佳LPPM所需的运行时间太长(可能需要数天).另一个问题涉及构建出的LPPMs的最佳性.揭示了一些意外情况(异常),即在粒度更高的细网格上构建的最佳LPPM效用比在粒

收稿日期: 2023-10-31; **修回日期**: 2024-05-20
基金项目: 国家自然科学基金项目(62072098, 62022024, 62072103, 62072102, 62132009, 62061146001); 江苏省重点研发项目(BE2022065-5, BE2022680); 江苏省网络与信息安全重点实验室项目(BM2003201); 计算机网络和信息集成教育部重点实验室项目(93K-9); 软件新技术与产业化协同创新中心项目
This work was supported by the National Natural Science Foundation of China (62072098, 62022024, 62072103, 62072102, 62132009, 61061146001), the Jiangsu Key Research and Development Program (BE2022065-5, BE2022680), the Project of Jiangsu Provincial Key Laboratory of Network and Information Security (BM2003201), the Project of Key Laboratory of Computer Network and Information Integration of Ministry of Education of China (93K-9), and the Project of Collaborative Innovation Center of Novel Software Technology and Industrialization.
通信作者: 凌振(zhenling@seu.edu.cn)

度较低的粗网格上差. 引入了粒度独立性作为有效解决方法，提出了一个名为 Divide-and-Coin 的最佳 LPPM，其可以实时执行.Divide-and-Coin 将生成最佳报告位置的运行时间从至少 $O(n^{2.055})$ 缩短到 $O(\log n)$，其中 n 是报告位置的数量. 实验结果显示，Divide-and-Coin 可以在 1 s 内从城市级区域生成建筑级别的最佳报告位置.

关键词 位置隐私；最佳位置隐私保护机制；实时位置 API；隐私-效用异常；粒度独立性

中图法分类号 TP309

随着电动汽车(electric vehicle, EV)的发展，针对电动汽车用户的各种第三方位置服务(location-based services, LBS)已经日益流行. 这些 LBS 为电动汽车用户提供个性化的功能，如智能充电[1-2]、天气预报[3]、智能停车[4] 和道路辅助[5] 等.

LBS 通过调用电动汽车 API 平台提供的实时位置 API 来获取电动汽车用户的位置信息，如图 1 所示. 一个示例平台是 SmartCar[6]，它支持包括特斯拉、大众等在内的多种电动汽车制造商. API 平台直接与内置在电动汽车中的蜂窝调制解调器通信，并向第三方服务提供包括位置 API 在内的实时 API. 该平台始终使用访问控制机制来保护用户隐私，如 SmartCar 基于 OAuth2.0 授权访问. 用户必须完全信任第三方服务才能使用个性化功能，这带来了重要的位置隐私隐患[7-8].

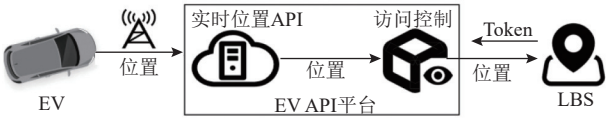


Fig. 1 Real-time EV location API

图 1 实时电动汽车位置 API

现有的研究仅关注位置隐私和 LBS 效用之间的权衡. 为了防止揭示用户的真实位置，位置隐私保护机制(location privacy preserving mechanism, LPPM)生成一个随机的伪位置作为报告位置，它可以形式化为给定真实位置情况下报告位置的概率密度函数(probability density function, PDF). 一个最佳 LPPM 通过解决一个最优化问题，从而实现最低可容忍的 LBS 效用限制下的最大化位置隐私，反之亦然.

现有的技术[9-15] 无法直接应用于电动汽车的实时位置 API. 第 1 个问题涉及到运行时间. 在现有技术中，求解最佳 LPPM 需要为每对真实位置和伪位置设置一个变量，例如，在给定真实位置的情况下报告伪位置的概率，这会引起巨大的计算开销. 即使候选的报告位置数量限制在不到 100 个，构建最佳 LPPM 所需的时间也超过 1 h[9]，甚至数天[11]. 同时由于电动汽车用户的历史位置(即用户位置分布)持续更新，最佳 LPPM 也需要定期重建. 第 2 个问题涉及到现有所谓“最佳”LPPMs 的最佳性. 现有的最新技术总是基于这样一个共同假设，即感兴趣区域内的所有位置都处在离散的网格上[9-13,16]. 然而，社会并没有充分了解这一假设的最佳性和实用性. 假设有一个细网格和一个粗网格，粗网格中的每个单元格可以通过合并细网格中的若干单元格获得. 文献 [11] 指出，一个在粗网格上最佳的 LPPM 不一定适用于细网格. 因此，许多方法[11,13] 便基于网格越精细，隐私和效用之间的权衡就会越好的期望，这一期望可以从整体趋势中得到证实[9,12-13,16]. 然而我们发现了一些意外的隐私-效用(PU)异常现象，即在精细网格上构建的最佳 LPPM 实现的隐私或效用可能不如在粗网格上，这与“最佳”的含义相矛盾，我们将这种意外现象称为 PU 异常.

本文的目标是实时构建最佳 LPPMs. 我们深入研究了 PU 异常，并引入了粒度独立性的概念作为有效解决方法. 基于这一概念，我们提出了一个实际而新颖的机制 Divide-and-Coin. 与现有技术总是计算所有位置上的完整概率分布不同，Divide-and-Coin 直接生成一个报告位置. 它将输入的位置区域递归地划分为子区域，计算出这些子区域的最佳报告概率，并根据这些概率随机选择一个子区域作为新的输入区域. 如果选择的区域符合用户设置的大小约束要求，其将作为计算出的报告位置.

我们证明了 Divide-and-Coin 是最佳的，即生成报告位置区域的概率等于最佳 LPPM 在该区域报告一个位置的累积概率. 此外，由于 Divide-and-Coin 不需要推导所有位置上的完整概率密度函数，与现有技术相比，它在运行时间上有了巨大的改进，复杂度从至少 $O(n^{2.055})$ 减少到 $O(\log n)$，其中 n 是报告位置的数量. 即使用户配置文件随时间变化，我们提出的机制仍然能够即时生成报告位置. 我们的实验结果显示，Divide-and-Coin 可以在 1 s 内从城市级区域($n=2^{217}=131\,072$)生成建筑级别的最佳报告位置.

总结一下，本文的主要贡献有3点：

1)做到能够实时构建最佳LPPM. 与现有技术相比，Divide-and-Coin将生成最佳报告位置的运行时间从至少$O(n^{2.055})$缩短到$O(\log n)$.

2)揭示并研究了最佳LPPMs的PU异常. 据我们所知，所有现有旨在构建最佳LPPMs的技术都受到了这一异常的影响.

3)引入了粒度独立性的概念，并作为对PU异常的有效解决方法.

1 系统模型

在本节中，我们描述本文考虑的系统模型，这与大多数现有的研究[9,12-13,16]是一致的. 我们还介绍了本文中使用的符号释义，总结在表1中.

Table 1 Notation Interpretation

表1 符号释义

符号	释义
$r,r',\hat{r}$	真实位置，报告位置，估计位置
$\mathcal{R},\mathcal{R}'$	真实位置集合，报告位置集合
$f(r',r)$	LPPM（给定r报告r'的概率）
$\pi(r)$	用户位于r的先验概率
$P(f,\pi),P^{\min},P^{\text{opt}}$	隐私，最小可忍受隐私，最佳LPPM隐私
$Q_{\text{loss}}(f,\pi),Q_{\text{loss}}^{\max},Q_{\text{loss}}^{\text{opt}}$	质量损失，最大可忍受质量损失，最佳LPPM质量损失
ω,ω'	输入给LPPM的位置粒度，由LPPM输出的位置粒度
$\mathcal{A},c_{\omega}^{\mathcal{A}},C_{\omega}^{\mathcal{A}}$	连续区域，粒度为ω的单元格，粒度为ω的单元格集合
$r_{\omega},r'_{\omega'},\hat{r}_{\omega}$	真实位置，报告位置，估计位置（对应粒度ω下）
$\mathcal{R}_{\omega},\mathcal{R}'_{\omega'}$	真实位置集合，估计位置集合（对应粒度ω下）
$f^{\text{tr}}_{\omega'_i,\omega_i}\left(r'_{\omega'_j},r_{\omega_j}\right)$	由粒度ω'_i,ω_i构建的LPPM再转换到粒度ω'_j,ω_j得到的等价LPPM

1.1 位置隐私保护机制

我们考虑瞬时LBSs[7]，用户可以零星地访问这些服务，因此同一用户的2个被揭示的位置是相互条件独立的[16]. 在不可信LBS提供商的情况下，攻击者可以获得该提供商持有的任何信息，包括用户的报告位置.

报告位置可以是用户的真实位置，也可以是由LPPM生成的伪位置. LPPM基于用户的真实位置$r\in\mathcal{R}$计算出恰当的报告位置$r'\in\mathcal{R}'$，从而实现用户的位置隐私和LBS效用之间的平衡. 报告位置集可以与真实位置集相同(即$\mathcal{R}=\mathcal{R}'$)，也可以不同(即$\mathcal{R}\neq\mathcal{R}'$)[17]. LPPM可以形式化为在给定真实位置$r$的情况下报告位置$r'$的概率密度函数$f(r',r)$，满足

$$\sum_{r'} f(r',r)=1,\ \forall r, \tag{1}$$

$$f(r',r)\geqslant 0,\ \forall r,r'. \tag{2}$$

1.2 攻击者掌握的信息

攻击者试图根据他的知识推断目标用户的真实位置r. 我们假设攻击者知道用户的报告位置r'和LPPM所计算出的概率密度函数f. 我们还假设他对用户的历史位置有一些了解，因此攻击者可以通过统计分析用户的历史信息来获取用户位置分布. 我们使用$\pi(r)$来表示用户位置分布，其代表用户在任何给定位置$r\in R$处的概率密度函数，其可被视为用户在位置r的先验概率.

1.3 理论上的最佳LPPM

为了评估LPPM，我们需要定义2个重要的度量标准. 使用Q_{loss}表示质量损失，即LPPM概率密度函数f的效用的对立面；使用P表示位置隐私. 当确定了这2个度量标准，一个最佳的LPPM便可以通过解决一个最优化问题来构建，它可以在最大可容忍的质量损失阈值下实现最高的隐私保护$Q_{\text{loss}}^{\max}$：

$$f^{\text{opt}}\equiv\arg\max_{f} P(\pi,f), \tag{3}$$

$$\text{s.t. (1)(2) and } Q_{\text{loss}}(\pi,f)\leqslant Q_{\text{loss}}^{\max}. \tag{4}$$

或者在最小可容忍的隐私阈值$P^{\min}$内实现最低的质量损失

$$f^{\text{opt}}\equiv\arg\min_{f} Q_{\text{loss}}(\pi,f), \tag{5}$$

$$\text{s.t. (1)(2) and } P(\pi,f)\geqslant P^{\min}. \tag{6}$$

此优化问题在大量文献[9−13,16,18−19]中被广泛使用，这些文献提出了不同的质量损失Q_{loss}度量标准和隐私P度量标准.

1.4 现实中的近似最佳LPPM

在现实世界中，报告位置可以是连续位置平面中的任何点(形式化为$\mathcal{R}^2$). 然而，出于计算原因，在构建最佳的LPPM时，$\mathcal{R}'$必须是可数集. 为了从$\mathcal{R}^2$生成可数集$\mathcal{R}'$，现有的研究始终考虑的是一个感兴趣的区域$\mathcal{A}$，该区域的大小有限，并通过指定单元格的边长将该区域映射到一个网格，并让$\mathcal{R}$或$\mathcal{R}'$中的任何位置都成为单元格的中心[9,11-13,16-17,20-27].

2 PU异常

现有方法[11,13]总是基于一个期望，即网格越精细，隐私与效用之间的权衡就能越好. 这一期望已经通

过模拟和实验得到了确认，但仅在总体趋势中成立.

在本节中，我们揭示了现有最佳 LPPMs 的 PU 异常，即通过更精细网格实现的“最佳”LPPM 的隐私（或质量损失）比通过较粗网格实现的隐私更少（或质量损失更大）. 我们将这种意外行为称为隐私-效用异常（PU 异常）.

2.1 参考技术

为了揭示和研究这一异常现象，我们参考了 2 种已经被社区广泛接受的代表性技术：基于估计距离误差构建最佳失真隐私机制的 EDE 程序[16]，以及基于地理不可区分性构建最佳机制的 GEO 程序[9]. 质量损失度量定义为报告位置 r' 和真实位置 r 的平均质量损失，其在 EDE 和 GEO 程序中都定义为

$$Q_{\text{loss}}\left(\pi,f,d_Q\right)\triangleq\sum_{r,r'}\pi(r)f(r',r)d_Q(r',r), \tag{7}$$

其中 $d_Q(r',r)$ 是 r' 和 r 的欧几里得距离. 然而，这些技术中的隐私度量定义并不相同.

在 EDE 程序中，隐私被定义为用户的无条件期望隐私:

$$P(\pi,f,d_P)\triangleq\sum_{r'}\min_{\hat{r}}\sum_{r}\pi(r)f(r',r)d_Q(\hat{r},r). \tag{8}$$

其中 $d_Q(\hat{r},r)$ 衡量了当真实位置为 r 时，攻击者猜测为 $\hat{r}$ 的损失，并以 $\hat{r}$ 和 r 之间的欧氏距离来构建.

在 GEO 程序中，隐私被定义为

$$P(f,d_P)\triangleq\min_{r,r',\hat{r}}\left(\frac{d_P(\hat{r},r)}{\ln|f(r',r)/f(r',\hat{r})|}\right), \tag{9}$$

其中 d_P 也以欧氏距离构建. 式(6)中的最小可容忍隐私 $P^{\min}$ 与最大隐私预算 ϵ 成反比，即 $P^{\min}=1/\epsilon$. 因此，式(6)中的约束条件等价于

$$f(r',r)\leqslant \mathrm{e}^{\epsilon d_P(r',r)}f(r',\hat{r}). \tag{10}$$

2.2 构建 LPPMs

对于每种技术，我们构建不同网格的最佳 LPPM，并比较它们所实现的隐私和效用的度量值. 为了区分不同的网格，我们使用参数 ω 来表示网格的粒度，一个基于粒度为 ω' 的报告位置集 $\mathcal{R}'$ 以及粒度为 ω 的真实位置集 $\mathcal{R}$ 构建的 LPPM 可以形式化为一个概率密度函数 $f^{\text{opt}}_{\omega',\omega}\left(r'_{\omega'},r_{\omega}\right)$. 为了简化计算，我们令 $\omega'=\omega$.

我们考虑一个非常简单的用户位置分布 π_0，如图 2 最右侧所示，其中 $\pi_0(r_i)=\pi_0(r_j)=0.5$. 我们还给出了一组粒度，其中任何 2 个粒度都相互构成上下级关系，如图 2 所示. 图 2 展示了不同粒度下用户的位置分布，$\pi_{\omega_{1\times2}}$，$\pi_{\omega_{2\times4}}$，$\pi_{\omega_{4\times8}}$，$\pi_{\omega_{8\times16}}$ 是基于粒度 $\omega_{1\times2}$，$\omega_{2\times4}$，$\omega_{4\times8}$，$\omega_{8\times16}$ 并由 π_0 转变的. 对于每个粒度，根据 π_0 生成一个新位置分布并将 π_0 域中的每个位置转换为该粒度下网格中的一个单元格的中心，例如，对于 $\omega_{1\times2}$，r_i=(0.29, 0.29) 被转换为 r_i=(0.5, 0.5). 下面给出 EDE LPPMs 和 GEO LPPMs 的定义.

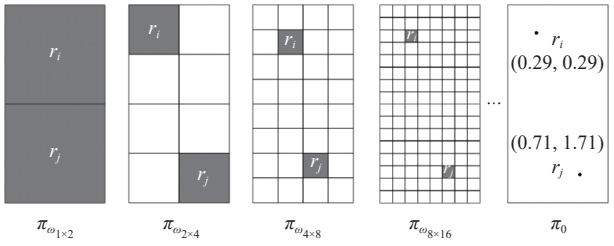


Fig. 2 User profiles with different granularities

图 2 不同粒度下的用户位置分布

1）EDE LPPMs. 对于给定的位置分布 π 和特定粒度 ω，以及给定的 $Q^{\max}_{\text{loss}}$ 上限，可以通过解决式(3)中定义的优化问题来构建一个最佳 LPPM，并同时满足式(1)(2)(4)(7)(8). 我们将 $Q^{\max}_{\text{loss}}$ 设置为 $\omega_{1\times2}$ 粒度下单元格边长的 0.5 倍，以构建图 2 中的所有用户位置分布的最佳 LPPM.

2）GEO LPPMs. 对于给定的位置分布 π 和特定粒度 ω，以及给定的 $P^{\min}$ 下限，可以通过解决式(5)中定义的优化问题来构建一个最佳 LPPM，并同时满足式(1)(2)(7)(9)(10). 我们将 $P^{\min}$ 设置为 1，以构建图 2 中的所有用户位置分布的最佳 LPPM.

2.3 PU 异常

确定一个最佳 LPPM $f^{\text{opt}}_{\omega',\omega_i}$ 是否在另一对粒度 (ω'_j,ω_j) 下仍然是最佳的，是一项不容易解决的任务. 一个直观的想法是再构建一个最佳 LPPM $f^{\text{opt}}_{\omega'_j,\omega_j}$，并比较这 2 个最佳 LPPM 所实现的隐私和效用的度量值. 然而，这 2 个 LPPM 的隐私和效用度量值不能直接进行比较，因为这些值的物理含义是由位置集合所决定的，不同粒度下的位置集合所决定的物理含义完全不同. 因此，以不同粒度构建的 LPPM 应该被转换到相同的粒度下进行比较.

我们现在讨论在什么情况下以及如何将某些粒度构建的 LPPM 进行转换. 假设有一个细网格和一个粗网格，粗网格中的每个单元格可以通过合并细网格中固定数量的单元格来获得. 基于条件概率，任何在细网格上构建的 LPPM 可以转换成在粗网格上的 LPPM，最终可以与直接在粗网格上构建的任何 LPPM 进行比较. 对于给定的区域 $\mathcal{A}$，粒度 ω 确定了如何将该区域映射到一个网格. 我们使用 $C^{\mathcal{A}}_{\omega}$ 来表示网格内所有单元格的集合，ω'_j 是相对于 ω'_i 的上级粒度，而 ω_j 是相对于 ω_i 的上级粒度. 对于任何的 $c'^{\mathcal{A}}_{\omega'_j}\in C'^{\mathcal{A}}_{\omega'_j}$ 和 $c^{\mathcal{A}}_{\omega_j}\in C^{\mathcal{A}}_{\omega_j}$，我们总是可以找到一些单元格，$c'^{\mathcal{A}}_{\omega'_i,\alpha}\in C'^{\mathcal{A}}_{\omega'_i}$

和$c^{\mathcal{A}}_{\omega_i,\beta}\in C^{\mathcal{A}}_{\omega_i}$, 满足:

$$c'^{\mathcal{A}}_{\omega'_j}=\bigcup_{1\leqslant\alpha\leqslant u,\ \alpha\in\mathbb{N}_+} c'^{\mathcal{A}}_{\omega'_i,\alpha},\ c^{\mathcal{A}}_{\omega_j}=\bigcup_{1\leqslant\beta\leqslant v,\ \beta\in\mathbb{N}_+} c^{\mathcal{A}}_{\omega_i,\beta}. \tag{11}$$

即对于任何一个存在下级粒度的单元格集合总能由有限个其下级粒度的单元格组合而成. 此外, $o(r')$表示r'被报告的观察事件, 而$a(r)$表示用户位于r的实际事件. 对于任何由下级粒度构建的原始概率密度函数$f^{\mathrm{opt}}_{\omega'_i,\omega_i}\left(r'_{\omega'_i},r_{\omega_i}\right)$, 我们可以基于条件概率计算其在上级粒度下的转换概率密度函数$f^{\mathrm{tr}}_{\omega'_i,\omega_i}\left(r'_{\omega'_j},r_{\omega_j}\right)$:

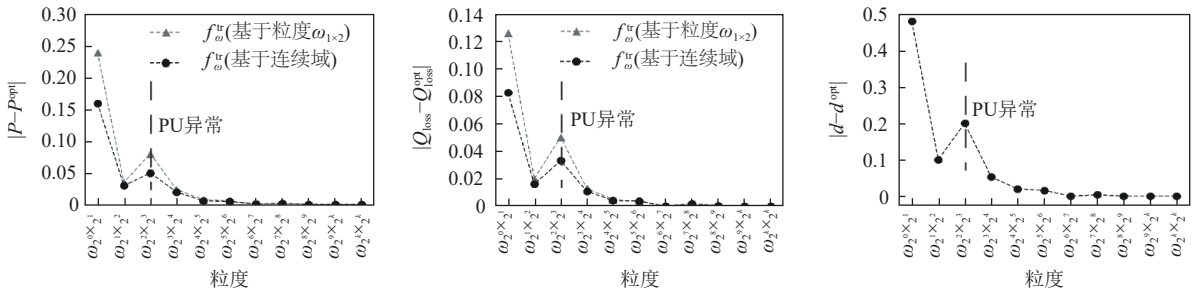
$$f^{\mathrm{tr}}_{\omega'_i,\omega_i}\left(r'_{\omega'_j},r_{\omega_j}\right)=Pr\left\{o\left(r'_{\omega'_j}\right)|a\left(r_{w_j}\right)\right\}=\frac{\sum\limits_{\alpha,\beta}f^{\mathrm{opt}}_{\omega'_i,\omega_i}\left(r'_{\omega'_i,\alpha},r_{\omega_i,\beta}\right)\pi\left(r_{\omega_i,\beta}\right)}{\sum\limits_{\beta}\pi\left(r_{\omega_i,\beta}\right)}, \tag{12}$$

其中需要注意的是, 带有不同下标的$r'_{(\cdot)}$和$r_{(\cdot)}$是单元格$c'^{\mathcal{A}}_{(\cdot)}$和$c^{\mathcal{A}}_{(\cdot)}$的中心, 并且这些单元格的下标与式(11)中相应定义的下标一致.

我们在本段落揭示EDE/GEO的LPPM在π_0上存在的PU异常现象. 使用不同粒度构建的EDE/GEO LPPM被转换成$\omega_{1\times2}$以及连续位置域, 与最佳LPPM实现的隐私和质量损失进行比较. P(或Q_{loss})表示转换后的LPPM实现的隐私(或质量损失), P^{opt}(或$Q^{\mathrm{opt}}_{\mathrm{loss}}$)表示最佳LPPM实现的隐私(或质量损失), 转换后的LPPM与最佳LPPM之间的隐私(或质量损失)差异可以计算为隐私距离$|P-P^{\mathrm{opt}}|$(或质量损失距离$\left|Q_{\mathrm{loss}}-Q^{\mathrm{opt}}_{\mathrm{loss}}\right|$).

结果如图3所示. 在图3(a)中, 总体趋势验证了最佳EDE LPPM在细网格上有更好的隐私表现. 然而, 粒度$\omega_{4\times8}$处存在异常. 图3(b)则说明了最佳GEO LPPM在效用上同样存在异常. EDE的效用性能趋势与GEO的隐私性能趋势相似, 粒度$\omega_{4\times8}$上都存在异常. 因此, 对于在细粒度网格上构建的最佳LPPM, 无论是EDE还是GEO, 都无法确保其在应用于粗粒度网格时仍然保持最佳, 也无法确保其仍然满足相关约束条件.

在第6节中, 我们对目前技术进行了调查. 所有构建最佳LPPM的技术都部分或完全基于EDE和/或GEO程序, 因此它们都受到PU异常的影响.



(a) 由EDE LPPM实现的隐私 (b) 由GEO LPPM实现的质量损失 (c) 计算$d(r_i, r_j)$的差异值

Fig. 3 Performance of optimal EDE/GEO LPPMs constructed with different granularities

图3 在不同粒度上构建的最佳EDE/GEO LPPM的性能表现

3 粒度独立作为异常解决方法

本节将介绍一种PU异常的解决办法.

3.1 粒度独立

对于一个独立于粒度的真正最佳LPPM, 如果使用下级粒度生成的原始概率密度函数是最佳的, 那么根据式(12)计算的上级粒度下的转换概率密度函数也应该是最佳的. 这引出了我们对粒度独立性的定义:

定义 1. 粒度独立. 一个最佳LPPM $f^{\mathrm{opt}}_{\omega'_i,\omega_i}$满足粒度独立性, 当且仅当对于任何一对粒度$(\omega'_j,\omega_j)$, 其中$\omega'_j$是相对于$\omega'_i$的上级粒度, 而$\omega_j$是相对于$\omega_i$的上级粒度, 式(13)(14)都成立.

$$Q_{\mathrm{loss}}\left(f^{\mathrm{tr}}_{\omega'_i,\omega_i}\left(r'_{\omega'_j},r_{\omega_j}\right),\pi\left(r_{\omega_j}\right)\right)=Q_{\mathrm{loss}}\left(f^{\mathrm{opt}}_{\omega'_j,\omega_j}\left(r'_{\omega'_j},r_{\omega_j}\right),\pi\left(r_{\omega_j}\right)\right), \tag{13}$$

$$P\left(f^{\mathrm{tr}}_{\omega'_i,\omega_i}\left(r'_{\omega'_j},r_{\omega_j}\right),\pi\left(r_{\omega_j}\right)\right)=P\left(f^{\mathrm{opt}}_{\omega'_j,\omega_j}\left(r'_{\omega'_j},r_{\omega_j}\right),\pi\left(r_{\omega_j}\right)\right). \tag{14}$$

显然, 一个满足粒度独立性(granularity independence, GI)的最佳LPPM不会受到PU异常的影响. 此外, 一个最佳LPPM满足粒度独立性的一个充分条件是: $\mathcal{R}$和$\mathcal{R}'$都不受粒度的影响. 为了满足这一条件, 一个合理的设置是:

1) $\mathcal{R}$是一个有限的兴趣点集合(POIs);

2) $\mathcal{R}'$是一个感兴趣的位置区域(一个无限不可数的连续位置集合).

3.2 实现粒度独立的挑战

考虑到可计算性, 将上述设置应用于实际程序并

不是一件容易的事情. 如果将$\mathcal{R}$设置为有限的兴趣点(POIs)集合，将$\mathcal{R}'$设置为感兴趣的区域$\mathcal{A}$，我们可以修改约束条件式(1)、质量损失和隐私为：

$$\int_{\mathcal{R}'} f(r',r)\mathrm{d}r' = 1,\ \forall r, \tag{15}$$

$$Q_{\mathrm{loss}}(\pi,f,d_Q) = \sum_r \pi(r)\int_{\mathcal{R}'} f(r',r)d_Q(r',r)\mathrm{d}r', \tag{16}$$

$$P(\pi,f,d_P) = \int_{\mathcal{R}'} \min_{\hat{r}} \sum_r \pi(r) f(r',r) d_P(\hat{r},r)\mathrm{d}r'. \tag{17}$$

给定一个名为π的位置分布和一个名为$Q_{\mathrm{loss}}^{\max}$的上限，理论上可以通过解决式(3)中定义的优化问题来构建最佳的 EDE+GI LPPM，并同时满足式(2)(4)(15)(16)(17)的条件. 类似地，原则上也可以构建 GEO+GI LPPM.

然而，出于可计算性原因，这样的程序在实际应用中并不可行. 该程序需要为每对$r\in\mathcal{R}$和$r'\in\mathcal{R}'$都引入一个变量，而当$\mathcal{R}'$是一个无限集时，就会产生无限数量的变量. 因此，最优化问题无法在多项式时间内求解.

3.3 EDE+GI 程序作为衡量标准

我们解决如下 EDE+GI 程序的计算性挑战. 式(16)中的质量损失和式(17)中的隐私都具有一个有趣且重要的特性，对于$\mathcal{R}'$的任何给定划分$\mathcal{P}$，可以通过累积计算$\mathcal{P}$中每个元素的质量损失和隐私来计算质量损失和 EDE 隐私. 假设$\mathcal{P}=\{\mathcal{A}_1,\mathcal{A}_2,\cdots,\mathcal{A}_K\}$是$\mathcal{R}'$的一个划分，我们总有

$$Q_{\mathrm{loss}}(\pi,f,d_Q) = \sum_i Q_{\mathrm{loss},i}(\pi,f,d_Q), 1\leqslant i\leqslant k, \tag{18}$$

$$P(\pi,f,d_P) = \sum_i P_i(\pi,f,d_P), 1\leqslant i\leqslant k, \tag{19}$$

其中$Q_{\mathrm{loss},i}$和P_i被定义为使用$\mathcal{A}_i$分别获得的质量损失和隐私

$$Q_{\mathrm{loss},i}(\pi,f,d_Q) \triangleq \int_{\mathcal{A}_i}\sum_r \pi(r) f(r',r) d_Q(r',r)\mathrm{d}r', \tag{20}$$

$$P_i(\pi,f,d_P) \triangleq \int_{\mathcal{A}_i}\min_{\hat{r}}\sum_r \pi(r) f(r',r) d_P(\hat{r},r)\mathrm{d}r'. \tag{21}$$

这一特性意味着，尽管对于 EDE+GI 来说，无法解决最优化问题以获得报告任何$r'\in\mathcal{R}'$的最佳概率密度函数，但我们仍然可以获得报告$\mathcal{R}'$任何一部分的最佳概率密度函数. 这个概率密度函数可以用来生成一个位置区域$\mathcal{A}_i\in\mathcal{P}$，而不是一个位置点$r'\in\mathcal{R}'$. 在用户必须报告一个点位置的情况下，我们可以使用$\mathcal{A}_i$的中心来代表$\mathcal{A}_i$. 同时，由于划分$\mathcal{P}$并不唯一，$\mathcal{P}$中的位置区域$\mathcal{A}_i$的面积也可以被限制在某个阈值内以保证使用位置区域的中心来代表位置点所导致的误差在用户的可接受范围内. 我们提出 EDE+GI 程序作为一个衡量标准：给定用户位置分布π、划分$\mathcal{P}$和上限$Q_{\mathrm{loss}}^{\max}$，一个最佳 LPPM $f_{\mathcal{P}}^{\mathrm{opt}}$可以通过解决下述最优化问题构建

$$f_{\mathcal{P}}^{\mathrm{opt}} \triangleq \arg\max_{f_{\mathcal{P}}} \sum_{\mathcal{A}_i}\min_{\hat{r}}\sum_r \pi(r) f_{\mathcal{P}}(\mathcal{A}_i,r) d_P(\hat{r},r), \tag{22}$$

$$\text{s.t.}\sum_{\mathcal{A}_i,r}\pi(r) f_{\mathcal{P}}(\mathcal{A}_i,r) d_Q(\mathcal{A}_i,r) \leqslant Q_{\mathrm{loss}}^{\max}, \tag{23}$$

$$d_Q(\mathcal{A}_i,r) \triangleq \int_{\mathcal{A}_i} d_Q(r',r)\mathrm{d}r',\ \forall r,\mathcal{A}_i, \tag{24}$$

$$f_{\mathcal{P}}(\mathcal{A}_i,r)\geqslant 0,\ \forall r,\mathcal{A}_i, \tag{25}$$

$$\sum_{\mathcal{A}_i} f_{\mathcal{P}}(\mathcal{A}_i,r) = 1,\ \forall r. \tag{26}$$

该程序实现了最佳的权衡，通过该程序构建的任何 LPPM $f_{\mathcal{P}}^{\mathrm{opt}}$都是最佳的.

定理 1. 对于任何划分$\mathcal{P}$，由 EDE+GI 程序构建的位置隐私保护机制$f_{\mathcal{P}}^{\mathrm{opt}}$都是最佳的，即始终存在一个最佳 LPPM f^{opt}，它是 EDE+GI 程序的解，满足

$$f_{\mathcal{P}}^{\mathrm{opt}}(\mathcal{A}_i,r) = \int_{\mathcal{A}_i} f^{\mathrm{opt}}(r',r)\mathrm{d}r',\ \forall r,\mathcal{A}_i. \tag{27}$$

证明. 假设$f_{\mathcal{P}}^{\mathrm{opt}}$是实际 EDE+GI 程序的任意解，我们可以找到$f^{\mathrm{opt}}$. 考虑一个新的程序，它使用一个新的约束条件式(27)作为 EDE+GI 程序的约束. 令f^{opt}是这个新程序的解，我们现在证明f^{opt}必定也是 EDE+GI 程序的解. 首先，若f^{opt}不是 EDE+GI 程序的解，这意味着存在一个 LPPM f'满足条件：

$$P(\pi,f',d_P) > P(\pi,f^{\mathrm{opt}},d_P). \tag{28}$$

我们可以基于f'计算 LPPM $f'_{\mathcal{P}}$：

$$f'_{\mathcal{P}}(\mathcal{A}_i,r) = \int_{\mathcal{A}_i} f'(r',r)\mathrm{d}r',\ \forall r,\mathcal{A}_i. \tag{29}$$

很明显，$f'_{\mathcal{P}}$满足式(25)(26)中的约束条件. 此外，根据式(4)(18)(20)，$f'_{\mathcal{P}}$也满足式(23)(24)中的约束. 根据式(19)(21)(27)，$f'_{\mathcal{P}}$实现的隐私大于$f_{\mathcal{P}}^{\mathrm{opt}}$，这与式(22)中的定义相矛盾. 因此，我们得出结论，f^{opt}也是 EDE+GI 程序的解. 证毕.

4 即时的最佳隐私保护机制

在本节中，我们提出一种名为 Divide-and-Coin 的即时最佳隐私保护机制，用于报告位置的即时生成.

4.1 基本概念

Divide-and-Coin 不是一个构建 LPPM 的程序. 它本身就是一个 LPPM，并以迭代的方式基于真实位置随机生成一个报告位置区域，如图 4 所示. 假设用户能够指定一个感兴趣的区域$\mathcal{R}'$和一个报告区域大小

的上限$S^{\max}$. Divide-and-Coin 保证生成一个大小不会超过$S^{\max}$报告区域. 假设用户位于r_0. 首先, 考虑$\mathcal{R}'$的一个划分$\mathcal{P}$, 再以$f_{\mathcal{P}}^{\text{opt}}(\mathcal{A}_i, r_0)$的概率随机选择一个元素$\mathcal{A}_i \in \mathcal{P}$. 然后检查这个元素的大小($\mathcal{A}_i$同样是一个区域). 如果$\mathcal{A}_i$的大小不超过$S^{\max}$, 那么就报告$\mathcal{A}_i$; 否则, 用户考虑区域$\mathcal{A}_i$的新划分, 并选择这个新划分中的一个元素$\mathcal{A}_i'$. 上述步骤以迭代方式执行, 生成一个最终报告区域.

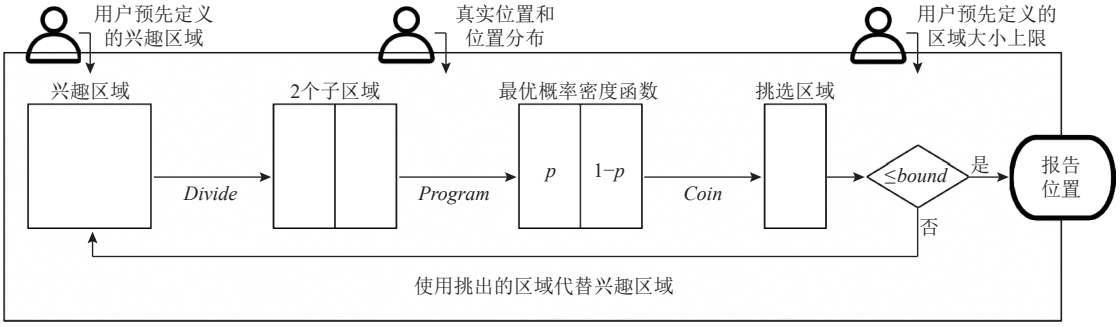


Fig. 4 Overview of Divide-and-Coin

图 4 Divide-and-Coin 概述图

4.2 最佳机制: Divide-and-Coin

Divide-and-Coin 主要由 3 个迭代执行的操作组成: *Divide*()操作将给定区域划分为子区域, *Program*() 操作构建报告子区域的最佳概率密度函数, *Coin*() 操作随机选择一个子区域.

1) *Divide* . 其以区域$\mathcal{R}' \subset R^2$作为输入, 输出区域$\mathcal{R}'$的一个划分$\mathcal{P}$. 为了保证正确性, *Divide* 操作确保对于相同的输入始终生成相同的输出. 为了降低时间复杂度, 它总是将 1 个区域划分为 2 个大小相同的子区域.

2) *Program* . 其以 *Divide* 操作生成的划分$\mathcal{P}$、位置分布π、真实位置r_0、当前迭代的质量损失上限$Q_{\text{loss}}^{\max}$以及上一轮迭代的$f_{\mathcal{P}}$作为输入(记为$f^{\text{temp}}(r)$), 输出为本轮迭代划分$\mathcal{P}$上的最佳分布$f_{\mathcal{P}}$. 其通过解决式(22)中定义的最优化问题得到, 同时满足式(23)(24)(25)以及

$$\sum_{\mathcal{A}_i} f_{\mathcal{P}}(\mathcal{A}_i, r) = f^{\text{temp}}(r),\ \forall r. \tag{30}$$

3) *Coin* . 它的输入是由 *Divide* 和 *Program* 操作分别生成的$\mathcal{P}$和$f_{\mathcal{P}}$. *Coin* 基于$f_{\mathcal{P}}$随机选择$\mathcal{P}$中的一个元素(区域$\mathcal{A}_i \in \mathcal{P}$)进行输出, 该区域$\mathcal{A}_i$成为下一次迭代中 *Divide* 的输入区域.

我们使用$f^{\text{D\&C}}(r', r)$来表示 Divide-and-Coin 的隐式概率密度函数. 其输入包括感兴趣区域$\mathcal{R}' \subset R^2$、用户位置分布π、真实位置$r_0 \in \mathcal{R}$, 输出区域大小上限$S^{\max}$、估计的质量损失Q_0的上限. 其输出是一个区域$\mathcal{A}' \subseteq \mathcal{R}'$.

我们的最终机制 Divide-and-Coin 包括 7 个步骤.

1)初始化.在第 1 次迭代中, $Q_{\text{loss}}^{\max}$设置为Q_0, $f^{\max}(r)$对所有$r \in \mathcal{R}$都设置为 1.

$$\mathcal{A}' = A = \mathcal{R}', \tag{31}$$

$$f^{\max}(r) = 1,\ \forall r \in R, \tag{32}$$

$$Q_{\text{loss}}^{\max} = Q_0, \tag{33}$$

$$f^{\text{D\&C}}(\mathcal{A}, r_0) = 1. \tag{34}$$

2)退出.当区域$\mathcal{A}'$的大小不再大于$S^{\max}$, 退出并报告区域$\mathcal{A}'$.

3)更新区域$\mathcal{A}$.构建步骤 2 中不满足要求的区域$\mathcal{A}'$的新划分$\mathcal{P}$, 计算划分$\mathcal{P}$上最佳分布$f_{\mathcal{P}}$, 并基于该分布随机选择该划分中的某个元素.

$$\mathcal{P} = Divide(\mathcal{A}'), \tag{35}$$

$$f_{\mathcal{P}}(\mathcal{A}_i, r) = Program\left(\pi, r_0, \mathcal{P}, Q_{\text{loss}}^{\max}, f^{\text{temp}}\right),\ \forall r \in \mathcal{R}, \mathcal{A}_i \in \mathcal{P}, \tag{36}$$

$$\mathcal{A} = Pick(\mathcal{P}, f_{\mathcal{P}}). \tag{37}$$

4)更新f^{temp}.暂存本轮迭代的最佳分布.

$$f^{\text{temp}}(r) = f_{\mathcal{P}}(\mathcal{A}, r),\ \forall r \in \mathcal{R}. \tag{38}$$

5)更新$Q_{\text{loss}}^{\max}$.$Q_{\text{loss}}^{\max}$取决于步骤 3 所选择的元素.

$$Q_{\text{loss}}^{\max} = \sum_r \pi(r) f_{\mathcal{P}}(\mathcal{A}, r) d_Q(\mathcal{A}, r). \tag{39}$$

6)计算报告区域$\mathcal{A}$的累积概率分布$f^{\text{D\&C}}$并更新$\mathcal{A}'$.

$$f^{\text{D\&C}}(\mathcal{A}, r_0) = f^{\text{D\&C}}(\mathcal{A}', r_0) f_{\mathcal{P}}(\mathcal{A}, r_0), \tag{40}$$

$$\mathcal{A}' = \mathcal{A}. \tag{41}$$

7)返回步骤 2.

Divide-and-Coin 本身是一个 LPPM, 可以形式化为一个隐式概率密度函数$f^{\text{D\&C}}(r', r)$. Divide-and-Coin 基于用户的真实位置r_0生成一个区域$\mathcal{A}'$作为报告位置, 并可以保证$\mathcal{A}'$的大小不会超过参数$S^{\max}$. Divide-

and-Coin 不需要计算完整的分布 $f^{\mathrm{D\&C}}$. 它根据真实位置 r_0 以及概率 $f^{\mathrm{D\&C}}(\mathcal{A}', r_0)$ 生成报告位置 $\mathcal{A}'$，概率 $f^{\mathrm{D\&C}}(\mathcal{A}', r_0)$ 可以通过计算每次迭代中选择一个元素的概率乘积来获得，如式(40)中所示. 概率 $f^{\mathrm{D\&C}}(\mathcal{A}', r_0)$ 描述了报告 $\mathcal{A}'$ 中每个点位置的累积概率：

$$f^{\mathrm{D\&C}}(\mathcal{A}', r_0) = \int_{\mathcal{A}'} f^{\mathrm{D\&C}}(r', r_0)\,\mathrm{d}r'. \tag{42}$$

4.3 Divide-and-Coin 的优势

Divide-and-Coin 具有 3 个优点：

1)实现最佳的权衡. Divide-and-Coin 是最佳的，因为它等于最佳概率密度函数在所有 $r' \in \mathcal{A}'$ 上的累积概率，根据定理 2 给出.

定理 2. Divide-and-Coin 是最佳的，即对于任何可能的报告区域 $\mathcal{A}'$(可能的程序输出)和任何真实位置 r_0，始终存在一个最佳 LPPM f^{opt}，它是 EDE+GI 程序的解，满足：

$$f^{\mathrm{D\&C}}(\mathcal{A}', r_0) = \int_{\mathcal{A}'} f^{\mathrm{opt}}(r', r_0)\,\mathrm{d}r',\ \forall r_0, \mathcal{A}'. \tag{43}$$

证明. 由于 *Divide* 操作只能为相同的区域生成一个划分，我们可以假设 Divide-and-Coin 机制迭代 n 次才最终输出 $\mathcal{A}'$. 进一步假设在第 k 次迭代中，由 *Coin* 操作生成的选定区域是 $\mathcal{A}'_k$. 根据定理 1，$f^{\mathrm{D\&C}}\left(\mathcal{A}'_1, r_0\right)$ 是最佳的. 对于 $1 \leqslant k < n$，如果 $f^{\mathrm{D\&C}}\left(\mathcal{A}'_k, r_0\right)$ 是最佳的，根据式(36)(38)(39)和命题 1 证明 $f^{\mathrm{D\&C}}(\mathcal{A}'_{k+1}, r_0)$ 也是最佳的，从而 $f^{\mathrm{D\&C}}(\mathcal{A}', r_0)$ 是最佳的. 证毕.

2)从不出现 PU 异常. Divide-and-Coin 满足粒度独立性，(根据定理 2 出发)因此从不出现 PU 异常.

3)较低的计算开销. 我们将 Divide-and-Coin 的时间复杂度与传统的 EDE 程序和 EDE+GI 程序进行比较. 我们假设可能的报告位置数为 n，真实位置数为 $m << n$. 我们进一步假设 $S^{\max}$ 不会超过 EDE 程序中的单元格大小. 根据文献 [28]，EDE 程序和 EDE+GI 程序的时间复杂度不小于 $O\left(m^{2\frac{1}{18}} n^{2\frac{1}{18}}\right) \approx O\left(m^{2.055} n^{2.055}\right) = O\left(n^{2.055}\right)$. 对于 Divide-and-Coin，它迭代 $\log n$ 次来生成最后的输出，每次迭代的时间复杂度只有 $O\left(m^{2.055}\right)$. 因此，整个程序的时间复杂度为 $O\left(m^{2.055} \log n\right) = O\left(\log n\right)$.

4.4 实现

我们将 Divide-and-Coin 实现为 C++中的一个位置隐私工具包. 基于此工具包，我们开发了一个由位置 API 和 Divide-and-Coin 机制组成的 LPPM 服务器，如图 5 所示. 该 LPPM 服务器需要用户授权来访问电动汽车 API 平台，因此其实际上是一个可信的位置服务提供商(LBS)，可以从 API 平台获取实时的真实位置信息. 当用户访问其他 LBS 时，该 LBS 无需借助电动汽车 API 平台授权来访问用户的真实位置信息，而是实时获取由 LPPM 服务器提供的报告位置.

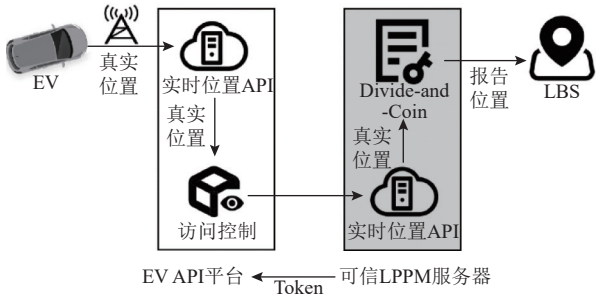


Fig. 5 Framework of using Divide-and-Coin in EV scenarios

图 5 电动汽车场景中使用 Divide-and-Coin 框架

5 评 估

在本节中，我们评估 PU 异常的影响来验证 Divide-and-Coin 的最佳性并评估其运行性能. 设置 $Q_{\mathrm{loss}}^{\max}$ 为单元格边长的0.869倍[16] 来构建最佳的 EDE LPPMs，并将 $P^{\min}$ 设置为 1[9] 来构建最佳的 GEO LPPMs.

5.1 实验配置

所有实验运行在操作系统为 Ubuntu 16.04 的台式机上，该台式机配置了 7.7 GB 内存和一颗 i7-6700CPU @ 3.40 GHz. 比较程序基于 C++实现. 实验共用到 4 种位置分布数据集来评估 PU 异常对 EDE 程序[10] 和 GEO 程序[9] 的影响以及本文算法的最佳性：

1)3 种假定的位置分布，如图 6 所示；

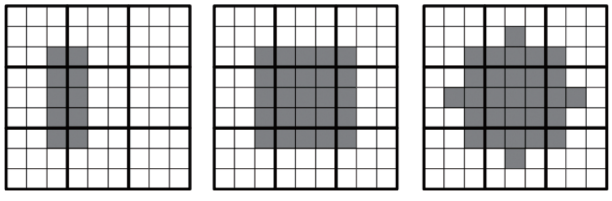
2)6 个随机生成的位置分布；

3)100 个随机生成的位置分布，每个大小为 16×16 个单元格，包含 64 个随机真实位置；

4)GeoLife GPS 轨迹数据集[29]，包含 182 个用户的 17 621 条轨迹.

5.2 PU 异常的影响

我们通过比较最佳 LPPMs 和转换后的 LPPMs 的隐私和效用性能来评估 PU 异常对 EDE 程序[10] 和



(a) π_1 (b) π_2 (c) π_3

Fig. 6 Three assumed profiles π_1, π_2 and π_3 distribute uniformly over the dark cells

图 6 3 个假定的位置分布 π_1，π_2，π_3 均匀分布在黑色单元格上

GEO 程序[9] 的影响.

对于图 6 中 3 种假定的位置分布进行的评估, 我们考虑 2 种不同的粒度, 即 $\omega_{9\times9}$ 和 $\omega_{3\times3}$. 我们根据式(12)将基于粒度 $\omega_{9\times9}$ 构建的最佳 LPPMs 转换为粒度 $\omega_{3\times3}$ 的 LPPMs, 以便与基于粒度 $\omega_{3\times3}$ 构建的最佳 LPPMs 进行比较. 隐私比和质量损失比如图 7 所示. 如果一个转换后的 LPPM 是最佳的, 那么隐私和质量损失的比率都应该是 1. 我们可以发现, 最佳的 EDE LPPMs 和最佳的 GEO LPPMs 都受到 PU 异常的影响, 而转换后的 LPPMs 始终在隐私方面表现更好但在效用方面表现更差, 这是因为这些位置分布中的真实位置彼此相近.

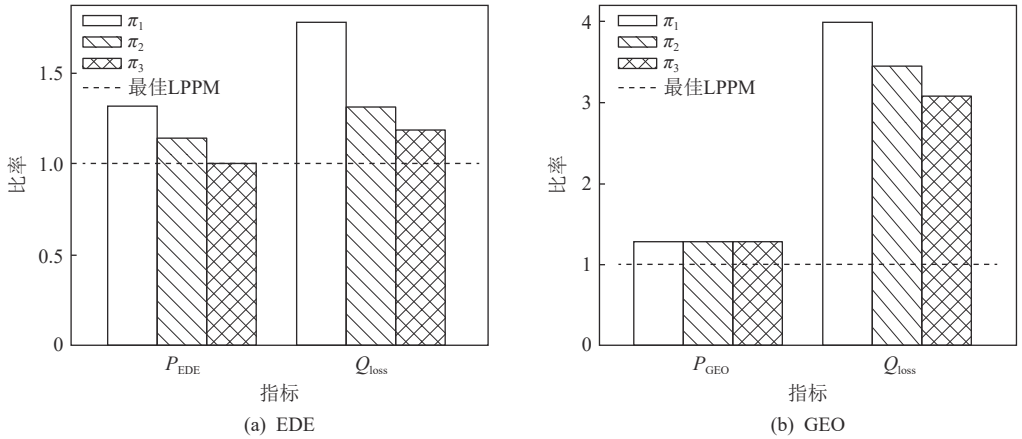


Fig. 7 Ratios of privacy and quality loss achieved by the transformed LPPMs from $\omega_{9\times9}$ and the optimal LPPMs with $\omega_{3\times3}$

图 7 由 $\omega_{9\times9}$ 转换的 LPPM 与由 $\omega_{3\times3}$ 构建的最佳 LPPM 在隐私比以及质量损失比

在使用 100 个随机位置分布的评估中, 我们旨在评估 PU 异常对最佳 EDE/GEO LPPMs 的平均影响. 我们考虑 4 种粒度, 即 $\omega_{16\times16}$, $\omega_{8\times8}$, $\omega_{4\times4}$, $\omega_{2\times2}$. 不同粒度的最佳 LPPMs 都被转换为 $\omega_{2\times2}$ 以进行比较. 结果如图 8 所示, 使用更精细的网格构建的最佳 LPPM 不一定在较粗的网格上仍然保持最佳, 并且其隐私保护和质量损失都不能保证.

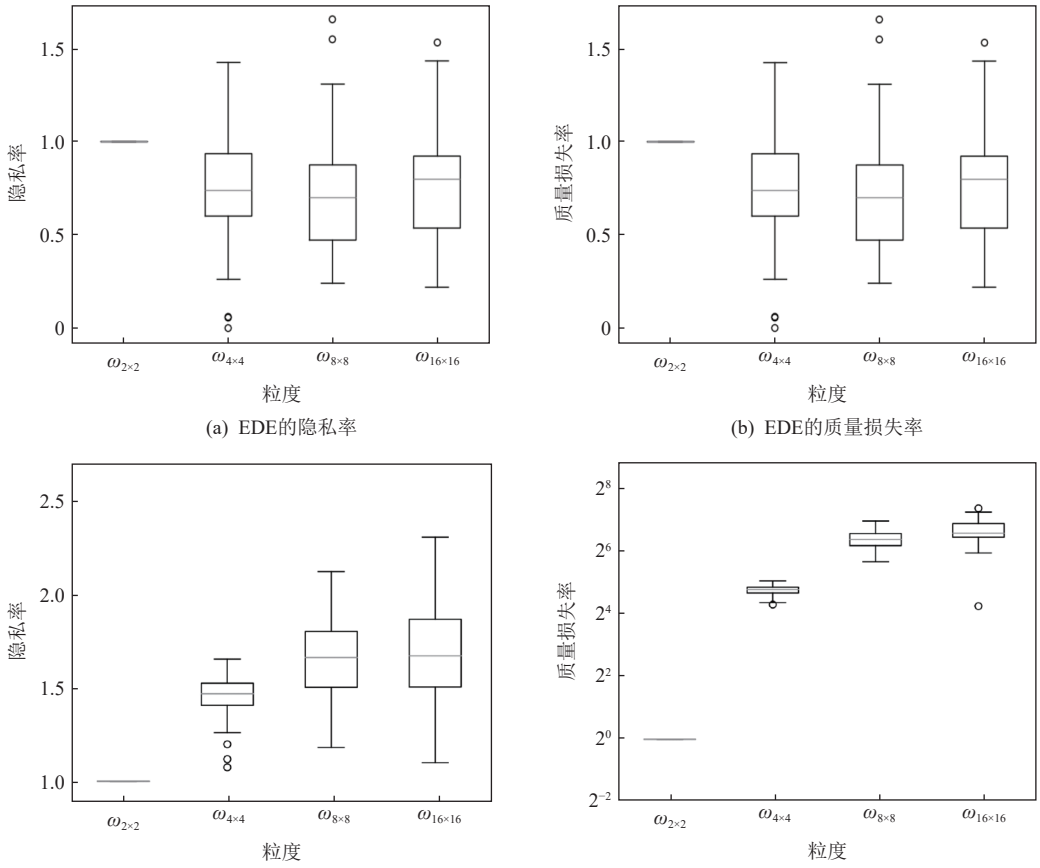
我们还使用真实用户位置分布来评估 PU 异常的影响. 同样, 我们考虑 4 种粒度, 即 $\omega_{16\times16}$, $\omega_{8\times8}$, $\omega_{4\times4}$, $\omega_{2\times2}$. 使用 $\omega_{16\times16}$ 粒度的单元格的边长约为 496 m, 结果如图 9 所示. 我们使用 P_{EDE} 和 P_{GEO} 来区分 EDE 隐私和 GEO 隐私. 可以发现, 对于最佳 EDE LPPMs 来说, Q_{loss} 的值应等于 Q_{loss}^{max}, 对于最佳 GEO LPPMs 来说, 实现的隐私应等于 P^{min}. 所有转换 LPPMs 都不是最佳的, 因此 EDE 和 GEO 都受到了 PU 异常的影响.

5.3 Divide-and-Coin 的最佳性

我们设计了实验来验证 Divide-and-Coin 的隐式概率密度函数是否是最佳概率密度函数. 实验方法为判断在给定真实位置的情况下生成报告位置的概率与由 EDE+GI 程序构建的最佳 LPPM 是否一致. 我们生成了 6 个随机位置分布, 其中 1 个如图 10 所示(其余类似, 不再详细描述), 每个真实位置的先验概率标注在单元格上. 我们使用 Divide-and-Coin 为每个真实位置生成一个概率密度函数. 在 Divide-and-Coin 执行的每轮迭代中, 我们获得 2 个子区域的概率. 然后, 对于每个子区域, 我们继续执行 Divide-and-Coin 以构建 2 个更小的子区域, 并获得相应的概率. 最终报告一个子区域的概率是选择其父区域的概率和在选择其父区域的条件下选择它的条件概率的乘积. 最后, 我们可以计算出报告任何位置的概率, 实现了最佳隐私保护的最终概率密度函数, 如图 11 所示.

5.4 Divide-and-Coin 的运行时间

EDE 程序和 EDE+GI 程序的时间复杂度都是 $O\left(|\mathcal{R}|^{2.055}|\mathcal{R}'|^{2.055}\right)$, 而 Divide-and-Coin 的时间复杂度仅为 $O\left(|\mathcal{R}|^{2.055}\log|\mathcal{R}'|\right)$. 我们考虑使用 GeoLife 数据集作为用户位置分布, 并进行实验以评估 Divide-and-Coin 的运行时间. 这些真实位置是用户的 POIs, 因此 $|\mathcal{R}|$ 不应该很大. 在这里, 我们提供一个有关 $\mathcal{R}$ 合理规模的计算. 假设可以为用户位置分布设置一个阈值 π^{min}, 来代表位置分布中的最小概率. 例如, π^{min}=0.001 表示如果用户位于某个位置的概率小于 0.001, 那么我们便不将该位置视为真实位置. 然后, 我们假定“80-20 原则”, 即 80% 的真实位置占据了位置分布中 20% 的概率. 基于这些假设, 我们得出结论, 如果 π^{min} = 0.001, 则 $|\mathcal{R}| < \left(20\%/\pi^{min}\right)/80\% = 250$. 如果 π^{min}= 0.005, 则 $|\mathcal{R}| < 50$. 我们设置 $50 \leqslant |\mathcal{R}| \leqslant 250$. 我们将 Divide-and-



(a) EDE的隐私率 (b) EDE的质量损失率

(c) GEO的隐私率 (d) GEO的质量损失率

Fig. 8 Boxplot of the privacy ratio and quality loss ratio using 100 random profiles

图 8 使用 100 个随机位置分布绘制的隐私率和质量损失率的箱线图

Coin 的平均运行时间与通过 EDE+GI 程序构建的最佳 LPPM 的运行时间进行比较，结果如图 12 和图 13 所示. Divide-and-Coin 的运行时间要小得多，而且不太受$|\mathcal{R}'|$值的影响，这与我们前文对时间复杂度的分析一致.

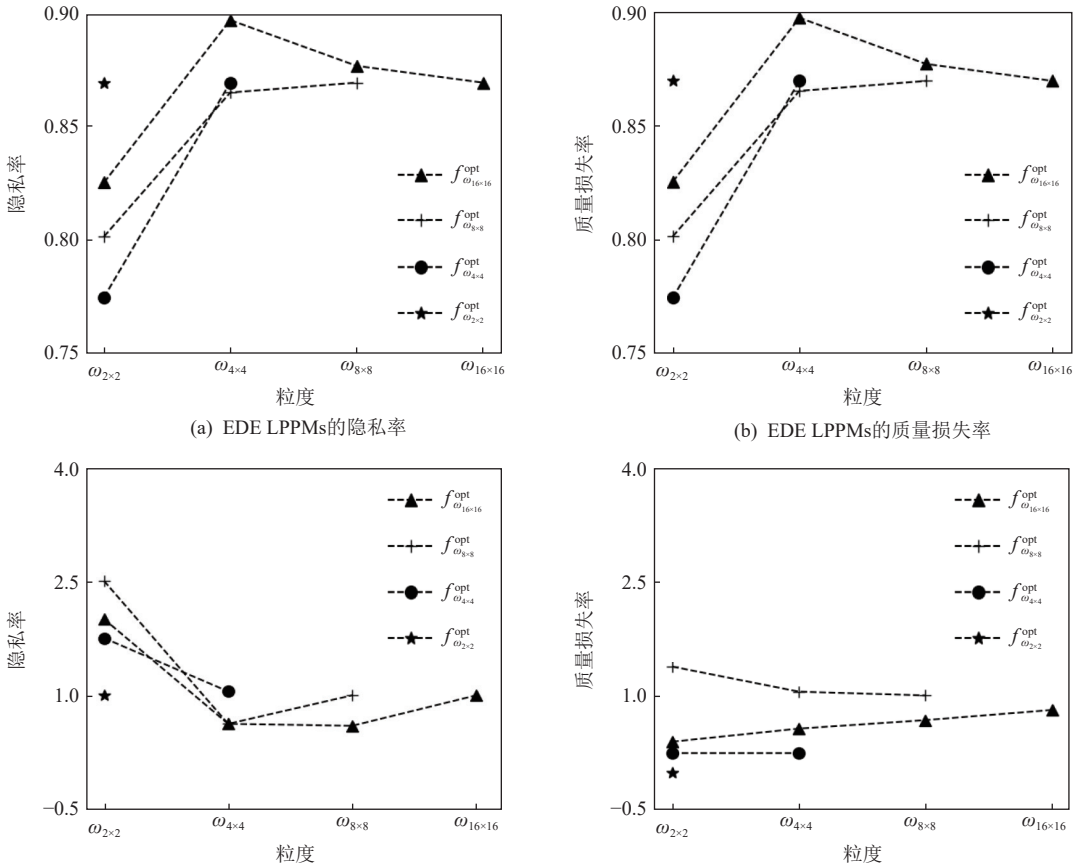
此外，我们进一步评估 Divide-and-Coin 是否能够实时执行. 图 14 展示了在不同$|\mathcal{R}|$值的情况下 Divide-and-Coin 的运行时间与迭代次数的关系. 我们考虑一个极端情况，即报告位置的数量$|\mathcal{R}'| = 2^{17} = 131\,072$（对于现有技术几乎无法计算），以及真实位置的数量$|\mathcal{R}| = 256$（这表明即使某个真实位置的概率小于 0.001，也会被考虑在内）. 在这种情况下，Divide-and-Coin 能够从一个2^{27} m^2（约 134.22 km^2）的城市级区域中生成一个2^{10} m^2（约 0.001 km^2）的建筑级位置. 实验结果表明，即使在这种极端条件下，Divide-and-Coin 仍能在 1 s 内完成.

6 相关工作

如何保护用户的位置隐私在很大程度上取决于 LBS 的访问方式. 如果一个 LBS 被频繁且匿名地访问，位置隐私就是避免被跟踪的能力[30-34]；如果一个 LBS 被偶尔且不匿名地访问，位置隐私就是避免暴露用户真实位置的能力[9,11-13,16-17,20-27]. 我们专注于后一种情况.

6.1 隐私概念和 LPPMs

考虑到用户的位置分布可以获得的情况，一个传统的位置隐私概念是位置熵[20-22]，其描述了攻击者的不确定性. Shokri 等人[16]引入了攻击者不正确性作为不确定性和不准确性的结合，并提出了估计距离误差（EDE）作为隐私的概念，这已经成为该领域的标准. 在另一方面的研究中，考虑用户个人资料是否未知（或不准确[35]），Andrés 等人[17]基于 2 维空间下差分隐私（DP）[36]的一般化[37]引入了地理不可区分性（GEO），它定义了位置隐私为攻击者在任何 2 个位置上的条件后验概率之间的最大差异. 尽管 GEO 不完美，因为它使用参数 ϵ 量化隐私，而这可能会对隐私级别[27,38]产生误导，但它已经成为位置隐私的另一个正式概念，并被许多研究[23-25]采用. EDE 和 GEO 的结合将同时保护位置免受贝叶斯推断攻击和侧信道推断攻击



(a) EDE LPPMs的隐私率　(b) EDE LPPMs的质量损失率

(c) GEO LPPMs的隐私率　(d) GEO LPPMs的质量损失率

Fig. 9　Privacy ratio and quality loss ratio achieved by transformed LPPMs and optimal LPPMs using real user profile

图 9　使用真实用户位置分布由转换后 LPPM 和最佳 LPPM 实现的隐私率和质量损失率

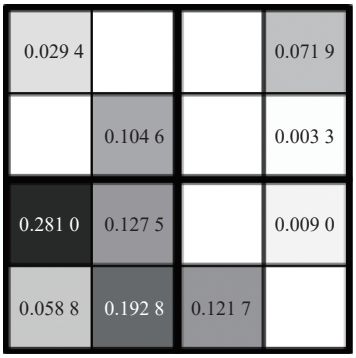


Fig. 10　A random generated user profile

图 10　一个随机生成的用户位置分布

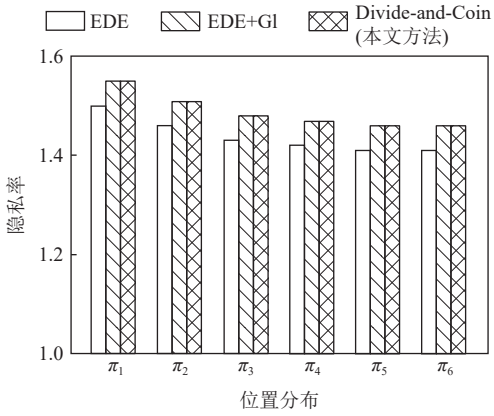


Fig. 11　Evaluation on the optimality of Divide-and-Coin

图 11　Divide-and-Coin 的最佳性评估

[26].

6.2　最佳 LPPMs

Shokri 等人[16]通过考虑攻击者的贝叶斯建模，提出了一个线性程序来构建最佳的 EDE LPPM，并随后使用 Stackelberg 贝叶斯博弈框架提出了一个改进的程序[10]. Chatzikokolakis 等人[9,11]提出了一个用于构建最佳 GEO LPPM 的程序. 文献 [12] 构建的 LPPM 在满足 GEO 约束的同时实现了最佳的 EDE，而文献 [13] 构建的 LPPM 则在保持 EDE 最佳的同时实现了最大化条件熵(CE). 根据不可区分性、DP 和互信息之间的关系[39]，基于熵的隐私概念还可以与 GEO 或 DP 相结合. 一些研究也考虑报告位置在连续平面上的情况[11,13,17]. 然而，它们要么无法构建最佳的 LPPM[11,17]，要么仍然需要将输出限制为离散的字母表[11,13].

6.3　电动汽车用户的位置隐私

Campolo 等人[40]提出了一个综合的基于智能手

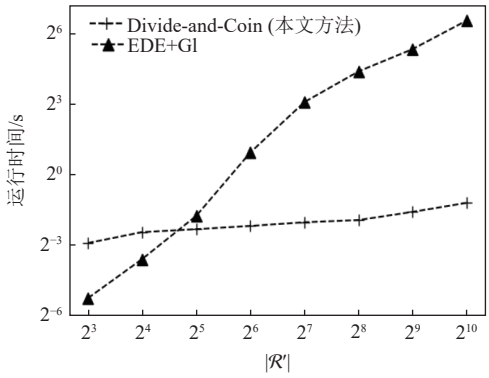


Fig. 12 Running time on $|\mathcal{R}'|$ with $|\mathcal{R}|$=256

图 12 在$|\mathcal{R}|$=256 时不同$|\mathcal{R}'|$的运行时间

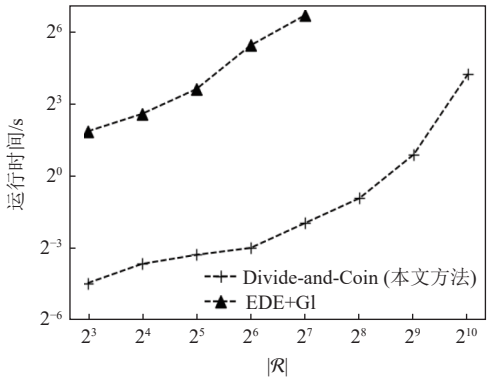


Fig. 13 Running time on $|\mathcal{R}|$ with $|\mathcal{R}'|$= 1 024

图 13 在$|\mathcal{R}'|$=1 024 时不同$|\mathcal{R}|$的运行时间

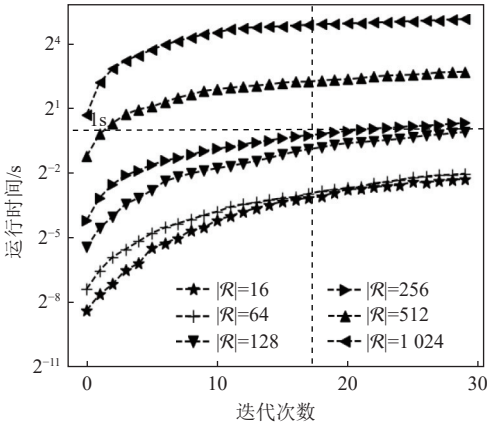


Fig. 14 Running time of Divide-and-Coin on $|\mathcal{R}|$

图 14 Divide-and-Coin 在不同$|\mathcal{R}|$的运行时间

机的平台，该平台检索车辆数据并提供实时 API. 通过使用基于身份的访问控制机制，该平台保护了包括位置在内的所有敏感数据. Gao 等人[41]提出了一种使用区块链来保护位置隐私的安全支付方案. Pazos-Revilla 等人[42]提出了一种使用物理层辅助来保护位置隐私的安全充电方案. Zhang 等人[8]提出了一个保护位置隐私的认证方案.

7 总 结

在这项工作中，我们的目标是为电动汽车场景中的实时位置 API 构建最佳的位置隐私保护机制(LPPM). 我们揭示了现有最佳 LPPM 的 PU 异常现象，并将其称为 PU 异常. 为了解决这种异常现象，我们将粒度独立性的概念正式转化为一种最佳性的属性. 最后，我们提出了一种可以实时构建的最佳机制 Divide-and-Coin.

作者贡献声明：董恺提出了算法思路和实验方案并负责撰写论文；王立夫负责完成实验并参与撰写论文；凌振参与算法思路设计与修改论文.

参 考 文 献

[1] Monta. The charging platform built to EV bettry [EB/OL]. [2022-07-20]. https://monta.com

[2] EvDotEnergy. Plug into bettey [EB/OL]. [2022-07-20]. https://ev.energy

[3] Hyundai Motor Company. BlueLink live services [EB/OL]. [2022-07-20].https://www.hyundai.com/eu/driving-hyundai/owning-a-hyundai/bluelink-connectivity/live-services.html

[4] EasyPark. Making cities more livable [EB/OL]. [2022-07-20]. https://easyparkgroup.com

[5] Freetrailer. Rent a trailer for free with freetrailer [EB/OL]. [2022-07-20]. https://freetrailer.com

[6] SmartCar. Easiest way to integrate apps cars [EB/OL]. [2022-07-20]. https://smartcar.com

[7] Jiang Hongbo, Li Jie, Zhao Ping, et al. Location privacy-preserving mechanisms in location-based services: A comprehensive survey[J]. ACM Computing Surveys, 2021, 54(1): 1−36

[8] Zhang Yinghui, Zou Jian, Guo Rui. Efficient privacy-preserving authentication for V2G networks[J]. Peer-to-Peer Networking and Applications, 2021, 14(3): 1−13

[9] Bordenabe N, Chatzikokolakis K, Palamidessi C. Optimal geo-indistinguishable mechanisms for location privacy[C]//Proc of the 2014 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2014: 251–262

[10] Shokri R, Theodorakopoulos G, Troncoso C, et al. Protecting location privacy: Optimal strategy against localization attacks[C]//Proc of the 2012 ACM Conf on Computer and Communications Security. New York: ACM, 2012: 617–627

[11] Chatzikokolakis K, ElSalamouny E, Palamidessi C. Efficient utility improvement for location privacy[J]. Privacy Enhancing Technologies, 2017, 2017(4): 308−328

[12] Shokri R. Privacy games: Optimal user-centric data obfuscation[J].

arXiv preprint, arXiv:1402.3426, 2014

[13] Oya S, Troncoso C, Pérez-González F. Back to the drawing board: Revisiting the design of optimal location privacy-preserving mechanisms[C]//Proc of the 2017 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2017: 1959−1972

[14] Cui Jie, Chen Xuefeng, Zhang Jing, et al. Bus cache-based location privacy protection scheme in the Internet of vehicles[J]. Journal on Communications, 2021, 42(7): 150−161 (in Chinese)
(崔杰, 陈学峰, 张静, 等. 基于公交车缓存的车联网位置隐私保护方案[J]. 通信学报, 2021, 42(7): 150−161)

[15] Li Hongtao, Ren Xiaoyu, Wang Jie, et al. Continuous location privacy protection mechanism based on differential privacy[J]. Journal on Communications, 2021, 42(8): 164−175 (in Chinese)
(李洪涛, 任晓宇, 王洁, 等. 基于差分隐私的连续位置隐私保护机制[J]. 通信学报, 2021, 42(8): 164−175)

[16] Shokri R, Theodorakopoulos G, Le Boudec J Y, et al. Quantifying location privacy[C]//Proc of the 2011 IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2011: 247−262

[17] Andrés M E, Bordenabe N E, Chatzikokolakis K, et al. Geo-indistinguishability: Differential privacy for location-based systems[C]//Proc of the 2013 ACM SIGSAC Conf on Computer & Communications Security. New York: ACM, 2013: 901−914

[18] Wang Bin, Zhang Lei, Zhang Guoyin. A gradual sensitive indistinguishable based location privacy protection scheme[J]. Journal of Computer Research and Development, 2020, 57(3): 616−630 (in Chinese)
(王斌, 张磊, 张国印. 敏感渐进不可区分的位置隐私保护[J]. 计算机研究与发展, 2020, 57(3): 616−630)

[19] Kim J W, Edemacu K, Jang B. Privacy-preserving mechanisms for location privacy in mobile crowdsensing: A survey[J]. Journal of Network and Computer Applications, 2022, 200: 103315

[20] Zhao Dapeng, Jin Yuanyuan, Zhang Kai, et al. EPLA: Efficient personal location anonymity[J]. GeoInformatica, 2018, 22(1): 29−47

[21] Niu Ben, Li Qinghua, Zhu Xiaoyan, et al. Achieving k-anonymity in privacy-aware location-based services[C]//Proc of the 2014 IEEE Conf on Computer Communications. Piscataway, NJ: IEEE , 2014: 754−762

[22] Sweeney L. K-anonymity: A model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557−570

[23] Wang Leye, Yang Dingqi, Han Xiao, et al. Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation[C]//Proc of the 26th Int Conf on World Wide Web. New York: ACM, 2017: 627–636

[24] To H, Shahabi C, Li Xiong. Privacy-preserving online task assignment in spatial crowdsourcing with untrusted server[C]//Proc of IEEE 34th Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2018: 833−844

[25] Hua Jingyu, Tong Wei, Xu Fengyuan, et al. A geo-indistinguishable location perturbation mechanism for location-based services supporting frequent queries[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(5): 1155−1168

[26] Yu Lei, Liu Ling, Pu C. Dynamic differential location privacy with personalized error bounds[C/OL]//Proc of NDSS. 2017 [2022-07-20].https://www.eecis.udel.edu/~ruizhang/CISC859/S17/Paper/p28.pdf

[27] Dong Kai, Guo Taolin, Ye Haibo, et al. On the limitations of existing notions of location privacy[J]. Future Generation Computer Systems, 2018, 86: 1513−1522

[28] Jiang Shunhua, Song Zhao, Weinstein O, et al. Faster dynamic matrix inverse for faster LPS[J]. arXiv preprint, arXiv:2004.07470, 2020

[29] Zheng Yu, Xie Xing, Ma Weiying. Geolife: A collaborative social networking service among user, location and trajectory[J]. IEEE Data Engineering Bulletin, 2010, 33(2): 32−39

[30] Mendes R, Cunha M, Vilela J P. Impact of frequency of location reports on the privacy level of geo-indistinguishability[J]. Proceedings on Privacy Enhancing Technologies, 2020, 2020(2): 379−396

[31] Zhang Wenjing, Li Ming, Tandon R, et al. Online location trace privacy: An information theoretic approach[J]. IEEE Transactions on Information Forensics and Security, 2018, 14(1): 235−250

[32] He Xiaofan, Jin Richeng, Dai Huaiyu. Leveraging spatial diversity for privacy-aware location-based services in mobile networks[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(6): 1524−1534

[33] Gursoy M E, Liu Ling, Truex S, et al. Utility-aware synthesis of differentially private and attack-resilient location traces[C]//Proc of the 2018 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2018: 196−211

[34] Terrovitis M, Poulis G, Mamoulis N, et al. Local suppression and splitting techniques for privacy preserving publication of trajectories[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(7): 1466−1479

[35] Oya S, Troncoso C, Pérez-González F. Rethinking location privacy for unknown mobility behaviors[C]//Proc of the 2019 IEEE European Symp on Security and Privacy (EuroS&P). Piscataway, NJ: IEEE, 2019: 416−431

[36] Dwork C. Differential privacy[C/OL]//Proc of the Int Colloquium on Automata, Languages and Programming. Berlin: Springer, 2006 [2022-07-20].https://link.springer.com/chapter/10.1007/11787006_1

[37] Chatzikokolakis K, Andrés M E, Bordenabe N E, et al. Broadening the scope of differential privacy using metrics[C]//Proc of the 13th Privacy Enhancing Technologies. Berlin: Springer, 2013: 82–102

[38] Oya S, Troncoso C, Pérez-González F. Is geo-indistinguishability what you are looking for[C]//Proc of the 2017 Workshop on Privacy in the Electronic Society. New York: ACM, 2017: 137−140

[39] Wang Weina, Ying Lei, Zhang Junshan. On the relation between identifiability, differential privacy, and mutual-information privacy[J]. IEEE Transactions on Information Theory, 2016, 62(9): 5018−5029

[40] Campolo C, Iera A, Molinaro A, et al. Smartcar: An integrated smartphone-based platform to support traffic management applications[C/OL]//Proc of the 1st Int Workshop on Vehicular Traffic Management for Smart Cities. Piscataway, NJ: IEEE, 2012 [2022-07-20]. https://ieeexplore.ieee.org/abstract/document/6398700

[41] Gao Feng, Zhu Leihuang, Shen Meng, et al. A blockchainbased

privacy-preserving payment mechanism for vehicle-to-grid networks[J]. IEEE Network, 2018, 32(6): 184−192

[42] Pazos-Revilla M, Alsharif A, Gunukula S, et al. Secure and privacy-preserving physical-layer-assisted scheme for ev dynamic charging system[J]. IEEE Transactions on Vehicular Technology, 2018, 67(4): 3304−3318



Dong Kai, born in 1985. PhD, associate professor. His main research interests include privacy protection, AI security, and IoT.

董　恺, 1985 年生. 博士, 副教授. 主要研究方向为隐私保护、AI 安全、物联网.



Wang Lifu, born in 2002. Bachelor. His main research interests include IoT and trigger action programming.

王立夫, 2002 年生. 学士. 主要研究方向为物联网、触发响应编程.



Ling Zhen, born in 1982. PhD, professor. His main research interests include IoT, system security, privacy protection, and trusted computing.

凌　振, 1982 年生. 博士, 教授. 主要研究方向为物联网、系统安全、隐私保护、可信计算.