基于差分隐私的连续位置隐私保护机制

李洪涛[1]，任晓宇[1]，王洁[1]，马建峰[2]
（1. 山西师范大学数学与计算机科学学院，山西 临汾 041099；2. 西安电子科技大学计算机科学与技术学院，陕西 西安 710071）

摘 要：针对连续使用基于位置的服务（LBS）会造成用户位置隐私泄露的问题，首先基于路网拓扑关系，提出了隐私级别划分算法——RPL 算法，对敏感路段进行隐私级别划分。然后，提出差分隐私位置保护机制 DPLPM，通过为敏感路段分配隐私预算并添加 Laplace 噪声，实现对位置数据的隐私保护。实验结果表明，所提机制能有效保护位置隐私，具有较高的数据可用性。

关键词：基于位置的服务；差分隐私；位置隐私保护；树结构
中图分类号：TN393
文献标识码：A
DOI: 10.11959/j.issn.1000−436x.2021123

Continuous location privacy protection mechanism based on differential privacy

LI Hongtao[1], REN Xiaoyu[1], WANG Jie[1], MA Jianfeng[2]
1. College of Mathematics & Computer Science, Shanxi Normal University, Linfen 041099, China
2. School of Computer Science and Technology, Xidian University, Xi'an 710071, China

Abstract: Aiming at the problem of users' location privacy leakage caused by continuously using LBS, a road privacy level (RPL) algorithm was proposed based on road topological network, which divided the privacy level of the road sections around the sensitive locations. Then, a differential privacy location protection mechanism (DPLPM) was proposed. Privacy budget was allocated for sensitive road sections and Laplace noise was added to realize the privacy protection of location data. The experimental results show that the mechanism has high data availability while protecting the privacy of location information.

Keywords: location-based service, differential privacy, location privacy protection, tree structure

1 引言

移动智能设备和定位技术的快速发展给移动用户带来了各种类型的基于位置的应用，为人们的生活带来诸多便捷。然而，由于移动用户在享受便捷服务的时候需要向基于位置的服务（LBS，location based service）提供他们的位置信息，使大量用户位置信息被不可信的第三方获取，可能使用户遭受严重的位置隐私泄露问题，危害用户的隐私安全。针对位置隐私保护技术，相关学者进行了大量研究。目前，多数位置隐私保护技术以 *k*-匿名或 *l*-多样性为基础，此类技术是将用户的真实位置泛化成一个区域，实现位置信息的隐私保护。其中，文献[1]基于用户的单一敏感属性设计了个性化 *k*-匿名模型与 KAUP（*k*-anonymity algorithm for personalized quasi-identifier attributes），提高了数据发布过程中的隐私保护程度。文献[2]提出一种基于 *l*-多样性大数据隐私保护方法，采用 NER（named

收稿日期：2021−01−04；**修回日期**：2021−04−03
基金项目：国家自然科学基金资助项目（No.61702316）；山西省自然科学基金资助项目（No.201901D111280, No.201801D221177）；山西省软科学基金资助项目（No.2017041016-4）
Foundation Items: The National Natural Science Foundation of China (No.61702316), The Natural Science Foundation of Shanxi Province (No.201901D111280, No.201801D221177)；Soft Science Project of Shanxi Province (No.2017041016-4)

entity recognition）方法将数据表示为结构化形式，进而对数据进行匿名化，实现对隐私数据的保护。抑制和扰乱技术也是近些年使用较多的保护方法。抑制技术的主要思想是不发布用户的敏感位置信息。文献[3]提出了一种基于信息熵的轨迹抑制隐私保护算法，通过函数计算抑制敏感位置点的最低代价，选择合理的抑制方式对原始数据集中包含敏感点的序列进行抑制。扰乱技术是将真实位置通过一定的变换生成假位置，达到保护真实位置的目的。文献[4]基于假位置隐私方法，提出了一种最大最小假位置选择（MMDS, maximum and minimum dummy selection）方案，使攻击者很难结合边信息过滤一些假位置，对位置信息进行隐私保护。以上几种隐私保护技术都有一定的局限性和缺点，攻击者可以通过长期的观察、挖掘和分析等方法获取用户的位置隐私信息[5-7]，因此这些技术无法抵抗相关攻击背景知识攻击。

Dwork 等[8]于 2006 年提出了差分隐私保护模型，其因良好的隐私保护强度成为一种主流的技术，通过对原始查询结果添加随机噪声，使在数据集中添加或删除某一条数据对查询结果不产生影响，从而让攻击者很难通过多次查询反推某条真实数据，实现隐私保护。Chen 等[9]将差分隐私保护机制应用于位置数据保护，通过对位置数据加入 Laplace 噪声，实现对位置数据的隐私保护。霍峥等[10]对自由空间和路网空间分别构造了噪声四分树和噪声 *R*-树，通过添加 Laplace 噪声保护位置数据，但没有考虑 2 个连续时刻位置数据间的相互影响。吴云乘等[11]采用差分隐私位置保护模型，把已知生成位置时真实位置的后验概率与真实位置概率的比值作为满足差分隐私的条件提出了 DPLRM（differential privacy location release mechanism）。Xiao 等[12]将地图转换为带权无向图，给位置区域分配隐私级别，在文献[11]的基础上，用马尔可夫链表示 2 个连续位置的关系，提出基于差分隐私的位置保护方案。然而，现有的差分隐私解决方法多数没有考虑位置间的关联，即使对用户所有位置进行保护，攻击者也会根据地理拓扑、时序关系等方法获取用户隐私信息。

针对以上问题，本文提出了一种差分隐私位置保护机制（DPLPM, differential privacy location protect mechanism），该机制能有效保护位置隐私和最大化数据可用性。本文主要贡献如下。

1) 根据路网的拓扑关系，对敏感位置周围路段进行隐私级别划分，提出道路隐私级别（RPL, road privacy level）划分算法（本文简称为 RPL 算法）。

2) 提出 DPLPM，构建位置树结构并给位置信息分配隐私预算，为敏感路段添加符合差分隐私机制的 Laplace 噪声，实现位置信息的隐私保护。

3) 理论分析和实验结果证明，所提机制能够较好地保护位置隐私和最大化数据可用性。

2　相关定义及其概念

2.1　差分隐私的相关定义

定义 1　邻近数据集。设数据集有相同的属性结构，两者仅相差一条记录，即 $\left|D\Delta D'\right|=1$，则称 D 和 D' 为邻近数据集。

定义 2　差分隐私。给定邻近数据集 D 和 D'，并已知某查询算法 A，若算法 A 在数据集 D 和 D' 的任意输出结果 O 满足不等式(1)，则称算法 A 满足 ε-差分隐私。

$$\Pr\left[A(D)=O\right]\leqslant \mathrm{e}^{\varepsilon}\Pr\left[A(D')=O\right] \tag{1}$$

隐私预算 ε 用来控制算法 A 在 2 个邻近数据集输出相同结果的概率比例，它表明隐私保护的程度，即 ε 越小，隐私保护程度越高，当 ε=0 时，隐私保护程度达到最高。

定义 3　全局敏感度。设函数 $f:D\to R^d$，输入一个数据集，输出 d 维实数向量，对于任意临近数据集 D 和 D'，称 GF_f 为函数 f 的全局敏感度。

$$\mathrm{GF}_f=\max_{D,D'}\left\|f(D)-f(D')\right\|_1 \tag{2}$$

其中，$\left\|f(D)-f(D')\right\|_1$ 是 $f(D)$ 和 $f(D')$ 之间的 1-阶范数距离。

2.2　Laplace 机制

Laplace 机制通过向确切的查询结果中加入服从 Laplace 分布的随机噪声来实现 ε-差分隐私，主要面向数值型查询结果。记初始位置下尺度参数为 b 的 Laplace 分布为 Lap(b)，其概率密度函数为

$$p(x)=\frac{1}{2b}\mathrm{e}^{-\frac{|x|}{b}} \tag{3}$$

定义 4　Laplace 机制。给定数据集 D，设有函数 $f:D\to \mathrm{e}^d$，其敏感度为 Δf，那么随机算法 $A(D)=f(D)=Y$ 提供 ε 差分隐私，其中

$Y \sim \mathrm{Lap}(\Delta f / \varepsilon)$ 为随机噪声，服从尺度参数为 $\Delta f / \varepsilon$ 的 Laplace 分布。

2.3　数据可用性

本文采用文献[12]的定义来测量本文数据可用性。假设 t 时刻发布位置是 O_t，其真实位置是 Z_t，本文采用 O_t 和 Z_t 之间的欧氏距离作为误差评价，即

$$\mathrm{dis}(Z_t - O_t) = \left\| Z_t - O_t \right\|_2 \tag{4}$$

特别地，对于长度为$|W|$的轨迹，本文同样以距离误差为基础衡量数据可用性，如式(5)所示。RMSE 等于位置上处于敏感区域的真实位置与其发布位置之间的均方根误差和。RMSE 越大，表示数据可用性越差。

$$\mathrm{RMSE} = \frac{1}{|W|}\sum_{t=1}^{W}\prod Z_t \mathrm{dis}(Z_t - O_t) \tag{5}$$

其中，$\prod Z_t$ 为指示函数，当 Z_t 为敏感位置时，指示函数 $\prod Z_t$ 的值等于 1，否则该值等于 0。

3　系统结构和威胁模型

3.1　系统结构

本文的系统结构如图 1 所示，主要包括三部分：客户端、隐私保护处理器和位置服务处理器。客户端主要通过 GPS 定位模块获取用户位置，并将位置存储至数据库中；隐私保护处理器分为数据划分模块、连续位置数据保护模块和数据库，数据划分模块将连续位置划分隐私级别，并将经过隐私级别划分的数据存储至数据库，连续位置数据保护模块对数据库中的位置提供差分隐私保护；位置服务处理器根据连续位置保护模块提出的查询请求，获得位置信息查询反馈，并将查询结果存储至数据库中。

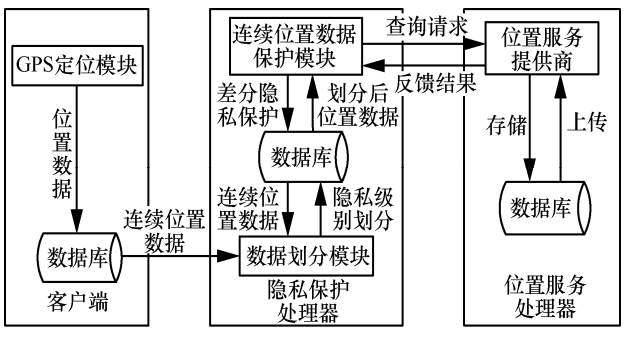


图 1　系统结构

针对用户位置隐私泄露问题，本文提出一种基于差分隐私的连续位置保护机制。在客户端，GPS 定位模块获取用户位置数据，并将其上传至数据库，数据划分模块用 RPL 算法对位置数据划分隐私级别；假设隐私保护处理器是可信任的第三方，连续位置数据保护模块从数据库中获取经过隐私级别划分的数据，通过 DPLPM 添加基于差分隐私的 Laplace 噪声，并生成位置集合；位置服务处理器向连续位置数据保护模块提出查询请求，连续位置数据保护模块将查询结果反馈至位置服务提供商；位置服务提供商在提供服务之后，将数据存储至数据库。

3.2　威胁模型

本文假设攻击者会不定时攻击用户的位置数据。很多位置服务提供商都有不同程度的安全保障，但当其服务器或数据库受到攻击时，用户的位置数据等隐私信息就可能被泄露。基于这个假设，本文提出的隐私威胁模型如图 2 所示。智能手机、便携电脑和近场通信（NFC, near field communication）等便携设备获取用户位置数据，并将连续位置数据传送至服务器端进行处理，进而从位置服务提供商处获取服务。攻击者可以通过攻击服务器端或直接攻击位置服务提供商，获取用户隐私信息。

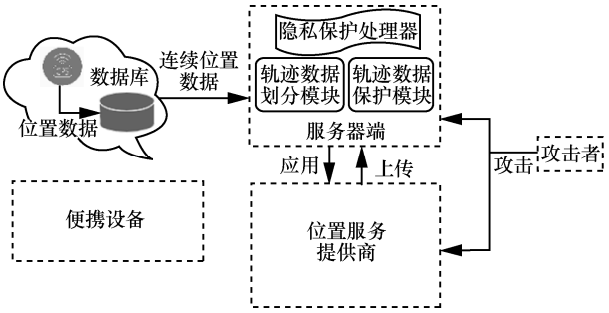


图 2　隐私威胁模型

4　本文机制描述

本文常用符号如表 1 所示。

4.1　RPL 算法

在数据划分模块，改进道路隐私级别划分算法，结合岔路口位置提出 RPL 算法。图 3(a)为真实区域的路网示意，○表示路口 i，★表示真实位置，☆表示用户当前所处位置，2 个路口间的虚线表示路口可直达。假设用户会选择最短路径到达目的位置。图 3(b)统计了到达不同的目的位置最少需要经过路口的个数。其中，用户自定义初始

敏感位置集合为 $SL^{initial}=\{sl_1, sl_2, \cdots, sl_{sl}\}$，对应的隐私级别集合为 $PL^{initial}=\{pl_1, pl_2, \cdots, pl_{pl}\}$。$PL^{initial}$ 中元素的取值范围为[0,1]，值越大表示敏感位置隐私级别越高。

表 1　　**常用符号**

符号表示	符号含义
G	路网
$SL^{initial}$	初始敏感位置集合
$PL^{initial}$	隐私级别集合
$NSL^{initial}$	初始非敏感位置集合
V_{max}	最大运行速度
ε_t	t 位置隐私预算
Z_t	真实位置
O_t	发布位置
W	真实轨迹
W'	发布轨迹
k	$SL^{initial}$ 的大小
η	距离阈值

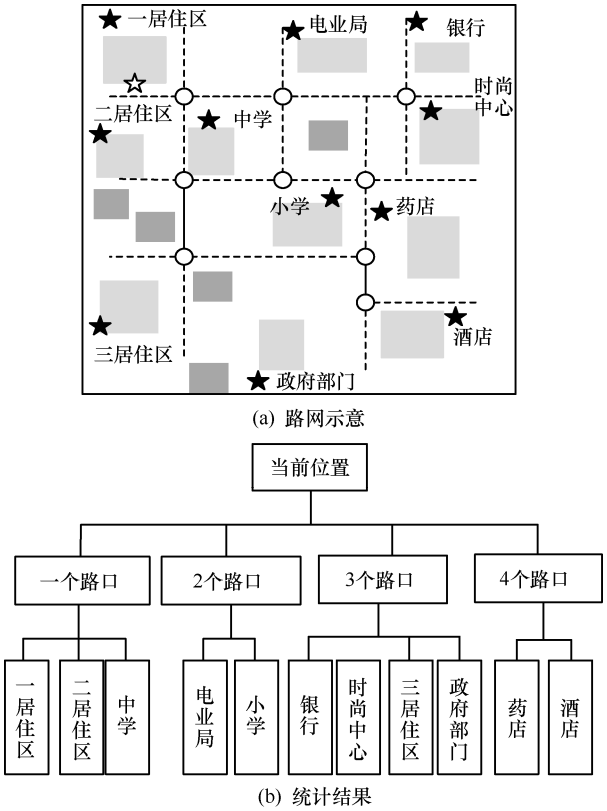


图 3　路网示意及到达不同的目的位置需要经过路口的个数

假设 v 为初始敏感位置，其隐私级别为 v.pl，v 的邻接位置集合为 neighborSet，g 是 neighborSet 中的一个位置，v 与 g 的距离为 g.dis，则 g 的隐私级别为

$$\varepsilon = g.\mathrm{pl} = \frac{1}{g.\mathrm{dis}} v.\mathrm{pl} \tag{6}$$

从式(6)可知，g.dis 越大，g.pl 就越小，即距离敏感位置越远的点隐私级别越小。若 g.dis=0，即 g 与初始敏感位置 v 重合，则取其本身的隐私级别和分配的隐私级别中较大者作为新的隐私级别；若 g.dis≠0，即 g 为初始非敏感位置，则其隐私级别为 g.pl。

已知用户初始敏感位置，根据图 3(a)计算该位置的隐私级别，步骤如下。确定初始集合 $SL^{initial}$ 中的位置 v 至路口 i 的路段上是否存在用户的第二个敏感位置，若不存在，则计算位置 v 与 i 之间的距离 dis(v, i)=R，当 $R<\eta$ 时，输出位置 v 与 i 的路段 $v \rightarrow i$；当 $R \geqslant \eta$ 时，输出距离为 η 的路段。若位置 v 至路口 i 的路段上存在用户的第二个敏感位置 g，则先根据式(6)计算位置 g 的隐私级别，并与 g 的初始隐私级别进行比较，取其中较大者作为位置 g 的最终隐私级别，此时，输出的敏感路段为两段，即 v 至 g 的路段 $v \rightarrow g$，以及 g 至路口 i 的路段 $g \rightarrow i$。用户道路隐私级别划分算法如算法 1 所示。

算法 1　RPL 算法

输入　路网 $G=<S, i, R>$，路网的区域划分 $M=\{SL^{initial}, NSL^{initial}\}$，初始敏感位置集合 $SL^{initial}$，每个敏感位置对应的隐私级别集合 $PL^{initial}$，距离阈值 η

输出　敏感路段及其隐私级别

1) SL = $SL^{initial}$
2) v = SL.head()
//位置 v 为集合中的第一个位置
3) while v! = NULL
//位置 v 不为空
4) neighborSet = find neighborSet(v)
//搜索位置 v 的邻居节点
5) if　g = NULL　//邻居节点为空
6)　dis(v,i) = R
//初始敏感节点到路口的距离为 R
7)　if　$R < \eta$
8)　output: $v \rightarrow i$
//输出初始敏感位置到路口的路段

9) else output $R = \eta$

//若距离 $R=\eta$，输出距离为 η 的路段

10) else

11) for all $g \in$ neighborSet

12) newpl = PL(g)

//计算邻居节点的隐私级别

13) if newpl < 1 then 执行算法

//隐私级别小于 1 可执行

14) if $g \in \mathrm{SL}^{\mathrm{initial}}$ then

//如果邻居节点属于初始敏感位置集合

15) g.pl = max(g.pl , newpl)

//选取初始隐私级别和计算隐私级别中较大者作为位置 g 的最终隐私级别

16) else

17) g.pl = newpl

//否则，g 的最终隐私级别为计算隐私级别

18) end for

19) output: $v \to g, g \to i$

//输出初始敏感位置到邻居节点，再由邻居节点到路口的路段

20) $v = v$.next()

//对初始敏感位置集合中的下一位置进行计算

21) end while

22) return

如果在初始敏感位置 v 到路口的路段上出现用户自定义的另一个一级初始隐私位置，则直接输出位置 v 到路口的全部路段。

4.2 差分隐私位置保护机制

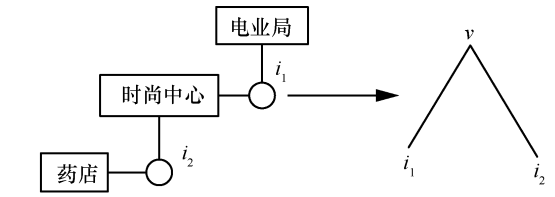
在位置数据保护模块，采用位置树结构，基于差分隐私保护模型提出 DPLPM。本文构建位置树反映路网实际情况。v 表示初始敏感位置，i 表示路口。路网结构转换为位置树如图 4 所示。

1) 如图 4(a)所示，敏感位置周围有 2 条路通往路口 i，转换为树结构后，包含一个根节点、2 个叶子节点。

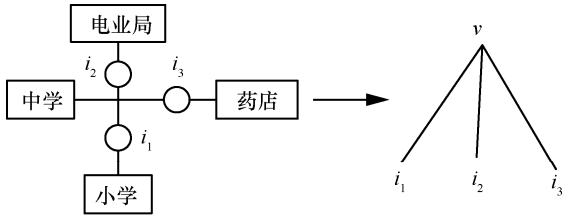
2) 如图 4(b)所示，敏感位置周围有 3 条路通往路口 i，转换为树结构后，包含一个根节点、3 个叶子节点。

3) 如图 4(c)所示，在敏感位置周围的路段上，有一个敏感位置 g，转换为树结构后，深度为 2，包含一个根节点、一个子节点、2 个叶子节点。

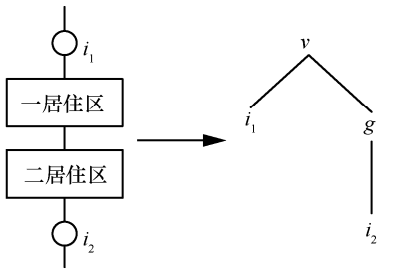
依次类推，将路网转换为位置树。



(a) 敏感位置周围有2条路通往路口i



(b) 敏感位置周围有3条路通往路口i



(c) 敏感位置周围的路段上存在敏感位置g

图 4 路网结构转换为位置树

根据差分隐私定义可知，隐私预算越小，对数据的隐私保护程度越大，以位置树的高度为基础，对隐私预算进行分配。

$$q = \frac{g.\mathrm{pl}}{v.\mathrm{pl}} \tag{7}$$

其中，q 为 2 个敏感位置间隐私级别比值，即其应分配隐私预算的比值。位置 v 为隐私级别最高的敏感位置，相应地，v.pl 最大，即 g.pl<v.pl，因此 q<1。由位置树结构可知，隐私级别较大者在位置树中所在高度为 1，所以需要分配更小的隐私预算，即 $\varepsilon = q\varepsilon_t$。

下面，根据路网示意构建位置树结构，根据树的高度分配隐私预算，添加 Laplace 噪声，实现对用户位置数据的保护。首先，以位置 v 为根节点构建位置树，若构建树的高度为 1，则将隐私预算 ε 平均分配至用户的 t 个敏感位置；若树的高度大于 1，则按照式(7)分配隐私预算，并根据隐私预算为每个位置加入符合差分隐私机制的 Laplace 噪声 Laplace Noisy(ε_t)。差分隐私位置保护机制如算法 2 所示。

算法 2 DPLPM

输入 路网 $G=<S, i, R>$，隐私预算 $\varepsilon=\{\varepsilon_1, \varepsilon_2, \cdots,$

$\varepsilon_h\}$，噪声树的高度 h，噪声值 $N=\{N_1, N_2, \cdots, N_h\}$，初始敏感位置 v，真实位置 Z_t，敏感路段上敏感位置数 t

输出　位置集合 W

1) 以 v 为根节点创建噪声树

2) if　$h=1$

3)　$\varepsilon_t=\varepsilon_h/t$

//隐私预算平均分配给 t 个位置

4)　$O_t=Z_t+\text{Laplace Noisy}(\varepsilon_t)$

//为每个位置添加 Laplace 噪声

5) else if　$h>1$

6)　$\varepsilon=N_1\varepsilon_1+N_2\varepsilon_2+\cdots+N_h\varepsilon_h$

//按照式(6)分配隐私预算

//$\varepsilon_1<\varepsilon_2<\cdots<\varepsilon_h$

7)　$O_t=Z_t+\text{Laplace Noisy}(\varepsilon_t)$

//为每个位置添加 Laplace 噪声

8) else $h<1$

9) 输入错误

10) $W=\{O_t\}$

//位置集合 W

11) return　W

5　差分隐私证明和算法分析

5.1　差分隐私证明

给定隐私预算 ε，本节证明 RPL 算法和 DPLPM 均满足 ε-差分隐私。

RPL 算法不能以差分隐私机制中数据集的概念来定义，因为对于位置和位置隐私来说，用户的每个位置都是需要被保护的，本文假设已知某时刻的发布位置 O_t，根据已发布的位置判断真实位置的后验概率为 $\Pr(Z_t|O_t)$，即

$$\frac{\Pr(Z_t\mid O_t)}{\Pr(Z_t)}=\mathrm{e}^{\varepsilon} \tag{8}$$

由差分隐私定义可知，RPL 算法满足 ε-差分隐私。

在 DPLPM 中，加入 Laplace 噪声，即符合 ε-差分隐私，证明过程如下。

证明　已知 Laplace 机制的概率密度函数为 $p(x)=\frac{1}{2b}\mathrm{e}^{-\frac{|x|}{b}}$，设 p_x 表示 $A_l(x,f,\varepsilon)$ 的概率密度函数，p_y 表示 $A_l(y,f,\varepsilon)$ 的概率密度函数，则对于某个输出值 Z，有

$$\frac{p_x(Z)}{p_y(Z)}=\prod_{i=1}^{k}\frac{\mathrm{e}^{-\frac{\varepsilon|f(x)_i-Z_i|}{\Delta f}}}{\mathrm{e}^{-\frac{\varepsilon|f(y)_i-Z_i|}{\Delta f}}}=\prod_{i=1}^{k}\mathrm{e}^{\frac{\varepsilon(|f(y)_i-Z_i|-|f(x)_i-Z_i|)}{\Delta f}}\leqslant$$

$$\prod_{i=1}^{k}\mathrm{e}^{\frac{\varepsilon(|f(x)_i-f(y)_i|)}{\Delta f}}=\mathrm{e}^{\frac{\varepsilon\|f(x)-f(y)\|_1}{\Delta f}}\leqslant\mathrm{e}^{\varepsilon}$$

可知，RPL 算法和 DPLPM 均满足 ε-差分隐私。证毕。

5.2　算法分析

1) 时间复杂度。假设连续位置数据中共有 n 个位置，在 RPL 算法中，最耗时的部分是遍历所有位置，其时间复杂度为 $O(n)=n$；在 DPLPM 中，最耗时的部分是将隐私预算分配给每层树结构，再将每一层的隐私预算分配给敏感路段中的每一个位置，其时间复杂度为 $O(n)=ht$。总体来说，本文所提算法计算消耗较小。

2) 隐私性。根据 DPLPM，t 时刻发布位置 O_t 使当前位置 Z_t 满足 ε-差分隐私，即轨迹 W 中，每一个位置都满足 ε-差分隐私，可知轨迹 W 满足 ε-差分隐私。

3) 数据安全性。本文经过数据划分模块和连续位置数据保护模块的处理后，将数据上传至位置服务提供商，位置服务提供商不能获得用户原始数据，所以攻击者不能通过位置服务提供商对用户进行攻击；本文采用差分隐私保护模型抵御背景知识攻击，攻击者无法根据用户的行为模式和地理拓扑关系推断出用户的真实位置，并根据用户的隐私级别分配不同的隐私预算，实现对用户位置数据的保护。

4) 数据可用性。本文通过式(5)对数据可用性进行分析。由式(5)可知，影响数据可用性主要有以下 3 个因素。①轨迹长度。轨迹长度越长，需要考虑的时刻越多，使发布位置与真实位置之间的距离越大，即数据可用性越差。②敏感位置 Z_t 与发布位置 O_t 的欧氏距离，距离越大，RMSE 越大，数据可用性越差。③指示函数 $\prod Z_t$ 的值，即判断位置 Z_t 是否为敏感位置，若为敏感位置，则 $\prod Z_t=1$，否则，$\prod Z_t=0$，数据可用性最高。本文所提 RPL 算法使敏感轨迹长度降到最低，同时由概率比值可得真实位置与发布位置之间的距离为 e^{ε}，通过控制隐私预算减少两者的距离，由此可知，本文在数据可用性方面是最优的。

6　实验与分析

6.1　实验设置

本节测试本文所提 DPLPM 的性能。实验使用 Python 实现，在 3.60 GHz CPU、8 GB RAM 的

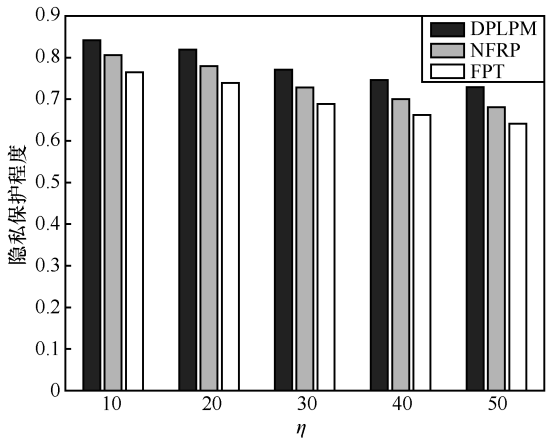
Windows 10 平台上运行，实验数据集为真实位置数据集 Geolife[13]和 Gowalla[14]。将 DPLPM 和 FPT（final private trajectory）算法[15]、基于频繁驻留点的加噪（NFRP, noisy of frequent resident points）算法[16]进行比较，从隐私保护程度、算法运行时间和数据可用性 3 个方面判断本文算法的优劣性。

6.2 实验结果与分析

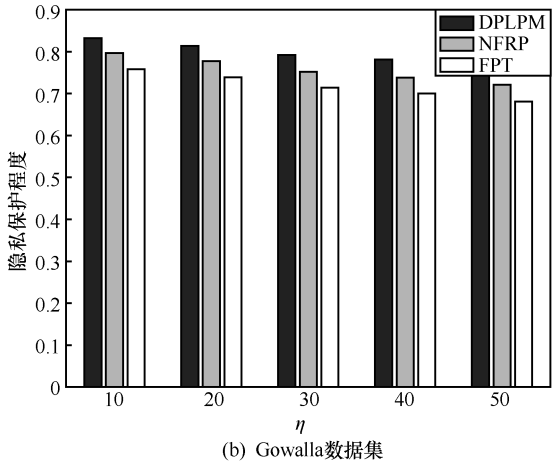
6.2.1 隐私保护程度

本节实验分析了本文所提 DPLPM 的隐私保护程度，通过用户自定义和式(6)计算隐私级别 pl、由仿真实验结果选出距离阈值 η、初始敏感区域 $SL^{initial}$ 大小 k 和隐私预算 ε，比较本文算法和 FPT 算法、NFRP 算法在 Geolife 数据集和 Gowalla 数据集上的隐私保护程度。

k=4，pl=0.25 时，距离阈值 η 对隐私保护程度的影响如图 5 所示。



(a) Geolife数据集

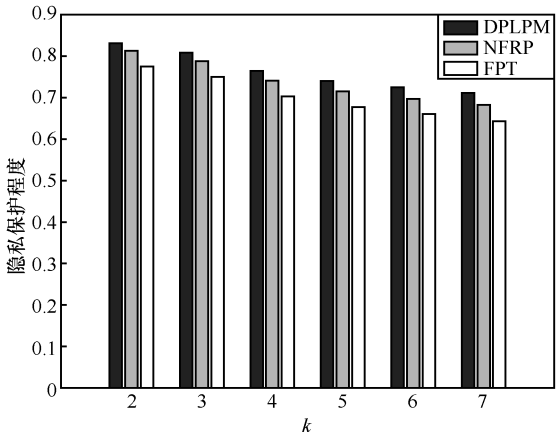


(b) Gowalla数据集

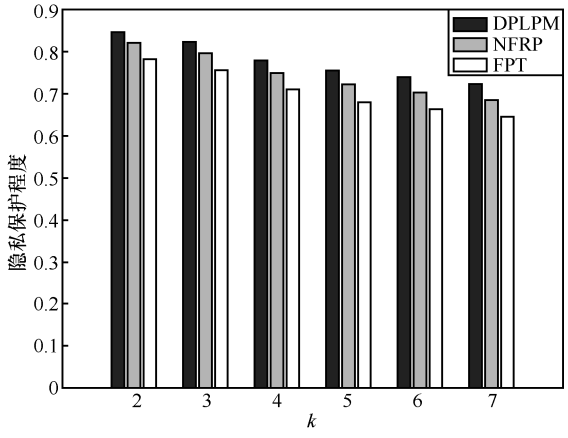
图 5 η 对 3 种算法隐私保护程度的影响

从图 5 可知，3 种算法的隐私保护程度都随着 η 的增大而减小。根据式(5)可知，随着输出路段距离的不断增大，位置的隐私级别不断下降，即隐私保护程度减小。

pl=0.25，η=20 时，初始敏感区域 $SL^{initial}$ 大小 k 对隐私保护程度的影响如图 6 所示。



(a) Geolife数据集



(b) Gowalla数据集

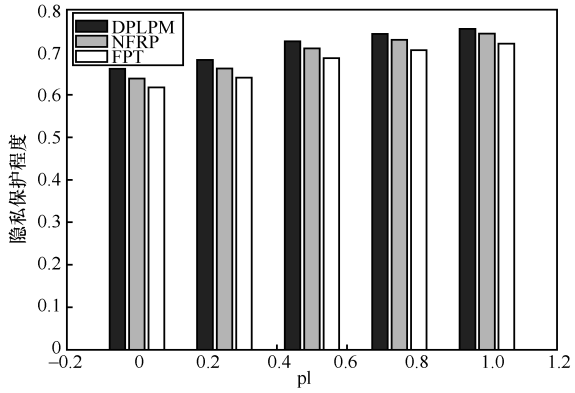
图 6 k 对 3 种算法隐私保护程度的影响

图 6 可知，3 种算法的隐私保护程度都随着 k 的增大而减小。这是因为敏感位置越多，则所需的隐私预算越多，即隐私保护程度越小。

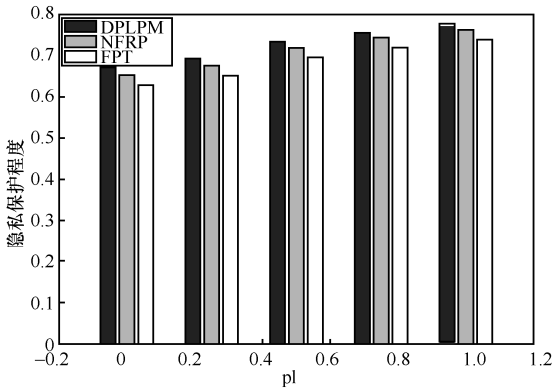
k=4，η=20 时，隐私级别 pl 对隐私保护程度的影响如图 7 所示。从图 7 可知，随着 pl 的增加，3 种算法隐私保护程度都在增加。这是因为隐私级别 pl 越大，为其分配的隐私预算越小，即隐私保护程度要越大。

k=4，η=20，pl=0.25 时，隐私预算 ε 对隐私保护程度的影响如图 8 所示。从图 8 可知，随着 ε 的增加，3 种算法隐私保护程度都在减少，由差分隐私定义可得，隐私预算越大，隐私保护程度越小。

此外，从图 5～图 8 可以看出，在不同参数下，本文所提算法的隐私保护程度都优于其他 2 种算法。

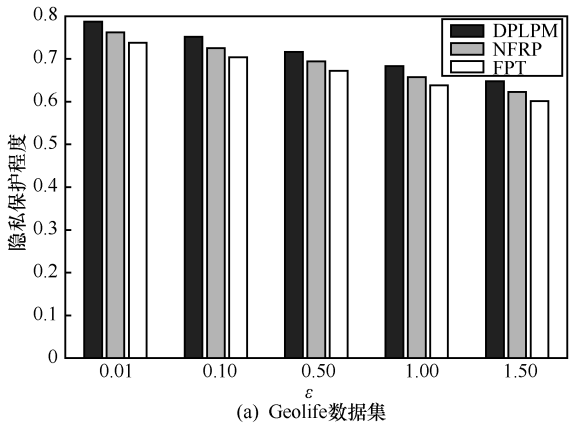


(a) Geolife数据集

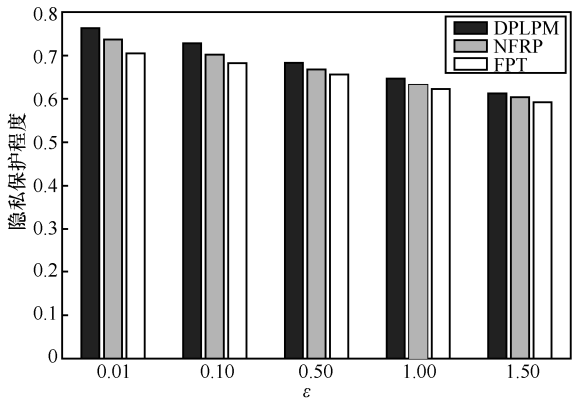


(b) Gowalla数据集

图 7　pl 对 3 种算法隐私保护程度的影响



(a) Geolife数据集



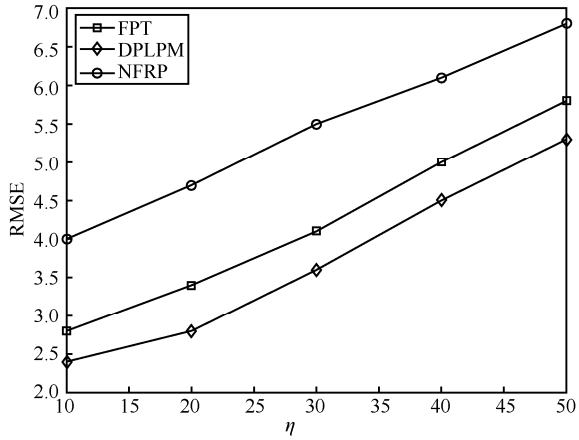
(b) Gowalla数据集

图 8　ε 对 3 种算法隐私保护程度的影响

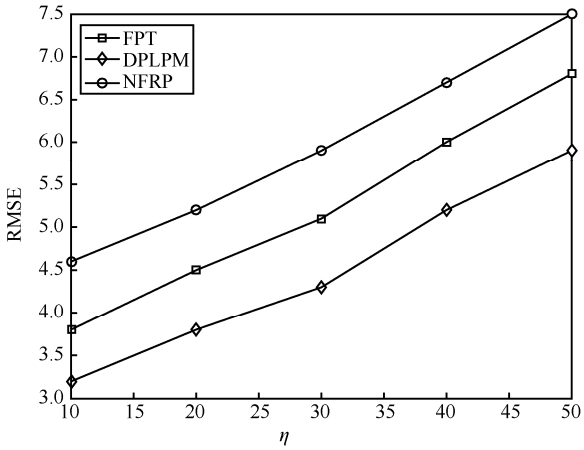
6.2.2　数据可用性

本文用 RMSE 衡量数据可用性。本节实验分别在 Geolife 数据集和 Gowalla 数据集上运行，对比了本文所提 DPLPM、FPT 算法和 NFRP 算法的数据可用性。其中，FPT 算法通过构建噪声前缀树实现对位置数据的保护，NFRP 算法根据统计后的流量图中边的流量值添加噪声。

k=4，pl=0.25 时，距离阈值 η 对 RMSE 的影响如图 9 所示。



(a) Geolife数据集



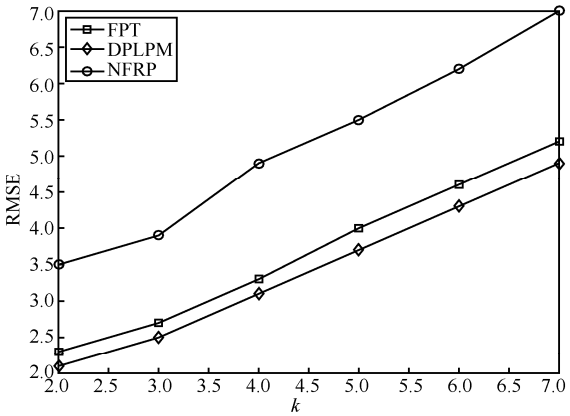
(b) Gowalla数据集

图 9　η 对 3 种算法 RMSE 的影响

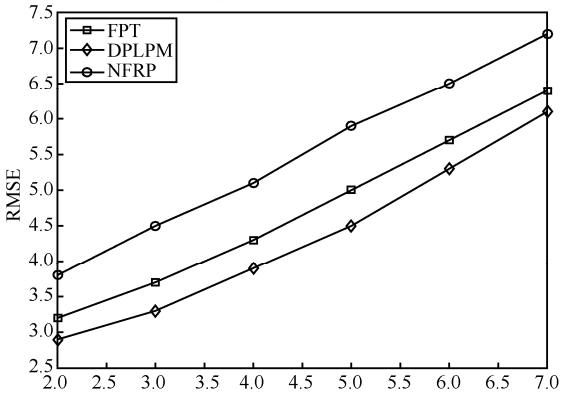
从图 9 可知，3 种算法的 RMSE 都随着 η 的增大而增大。NFRP 算法的可用性最差，因为该算法只为某一位置的经纬度添加噪声；FPT 算法的位置数据可用性相对较好，但 FTP 算法较多地对空节点添加噪声；DPLPM 考虑了位置连续性之间的影响，数据可用性是最好。

pl=0.25，η=20 时，初始敏感区域 $SL^{initial}$ 大小 k 对 DPLPM、FPT 算法和 NFRP 算法 RMSE 的影响如图 10 所示。

从图 10 可知，3 种算法的 RMSE 都随着 k 的增大而增大。这是因为位置个数的增大会导致隐私预算增多，添加的噪声也增加，即数据可用性变差。NFRP 算法的可用性最差，DPLPM 可用性最好，而 FPT 算法介于二者之间。



(a) Geolife数据集



(b) Gowalla数据集

图 10　k 对 3 种算法 RMSE 的影响

k=4，η=20 时，隐私级别 pl 对 DPLPM、FTP 算法和 NFRP 算法 RMSE 的影响如图 11 所示。

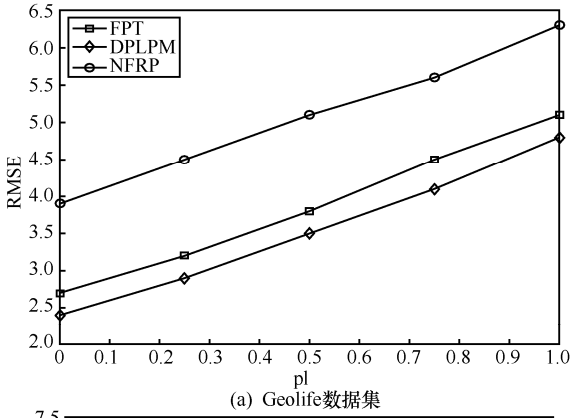
从图 11 可知，3 种算法的 RMSE 都随着 pl 的增大而增大。这是因为随着隐私级别的增大需要添加的噪声增加，数据可用性变差。DPLPM 的可用性最好，FPT 算法次之，NFRP 算法最差。

6.2.3　算法运行时间

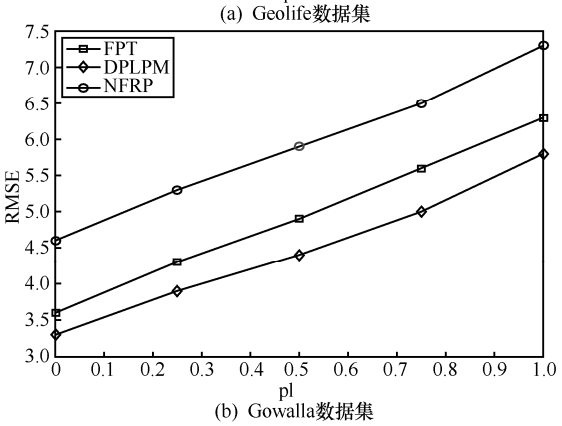
本节实验在Geolife数据集和Gowalla数据集上运行 DPLPM、FPT 算法和 NFRP 算法，对比 3 种算法的运行时间。

k=4，pl=0.25 时，距离阈值 η 对 DPLPM、FPT 算法和 NFRP 算法运行时间的影响如图 12 所示。

从图 12 可知，随着 η 的增加，3 种算法的运行时间都随之增加。因为 DPLPM 只需要提供加噪，所以 DPLPM 运行时间是最少的。

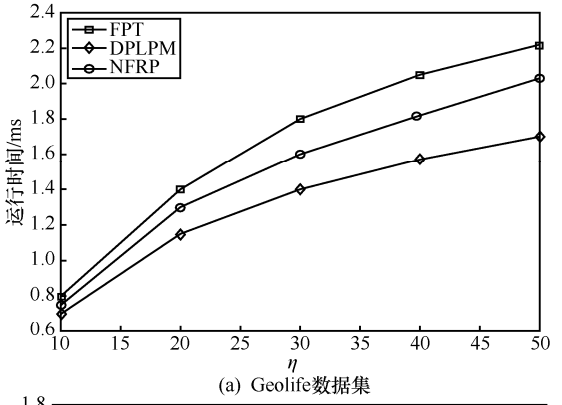


(a) Geolife数据集

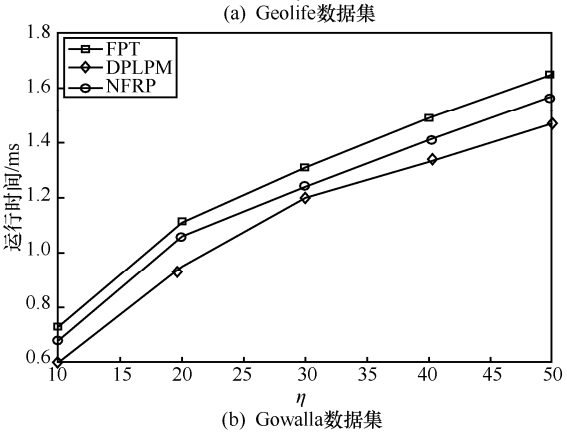


(b) Gowalla数据集

图 11　pl 对 3 种算法 RMSE 的影响



(a) Geolife数据集



(b) Gowalla数据集

图 12　η 对 3 种算法运行时间的影响

pl=0.25，η=20 时，初始敏感区域 $SL^{initial}$ 大小 k

对 DPLPM、FPT 算法和 NFRP 算法运行时间的影响如图 13 所示。

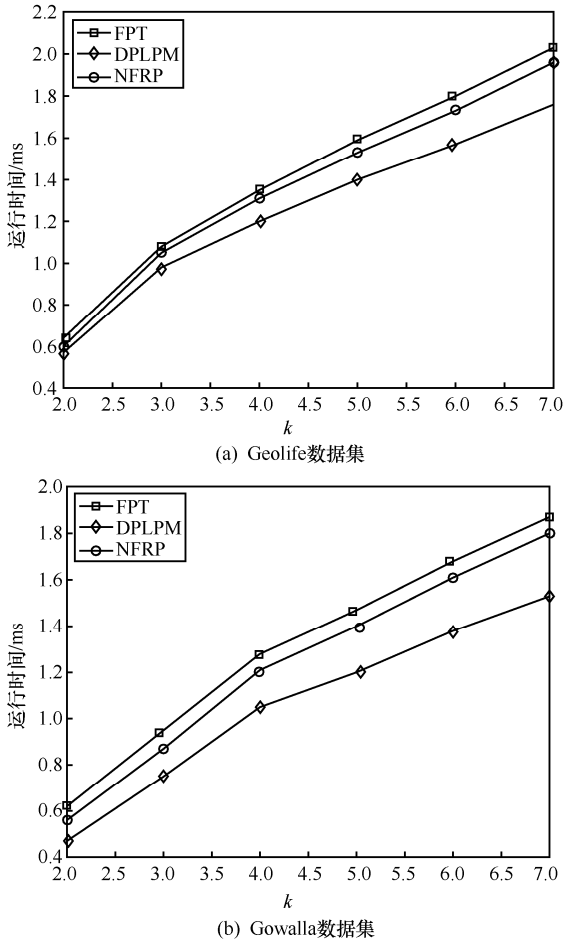


图 13　k 对 3 种算法运行时间的影响

从图 13 可知，随着 k 的增加 3 种算法的运行时间都随之增加。FTP 算法在 2 个数据集上的运行时间都是最长的，而 NFRP 算法次之，DPLPM 的运行时间是最短的。

k=4，η=20 时，隐私级别 pl 对 DPLPM、FTP 算法和 NFRP 算法运行时间的影响如图 14 所示。

从图 14 可知，随着 pl 的增加，3 种算法的运行时间都在增加。FPT 算法在 2 个数据集上运行时间都最长，NFRP 算法次之，DPLPM 的运行时间最短。

6.3　相关工作比较

本文所提 DPLPM 与其他方法的比较如表 2 所示。文献[17]提出一种基于隐私拆分的轨迹隐私保护方法，建立单点位置的发布对查询轨迹的前向和后向隐私风险评估机制，通过拆分查询轨迹，消除轨迹中位置间的相关性，但没有控制隐私预算，同时使算法开销增大。文献[12]提出一种差分隐私保护方法，根据时间相关性，使用马尔可夫链预测前一个位置对后一个位置的影响，考虑了路网中的位置连续性，但没有对隐私预算进行合理分配。文献[18]将不规则树引入差分隐私方法中，减少连续查询时噪声叠加带来的查询精度下降问题，但没有考虑位置间的相关性。文献[19]提出 AC-TFIDF（adaptive clustering based TFIDF）算法，根据重要位置点在不同时刻的分布状况，选择聚类中心代替原始位置，生成发布位置，但没有考虑路网实际情况。

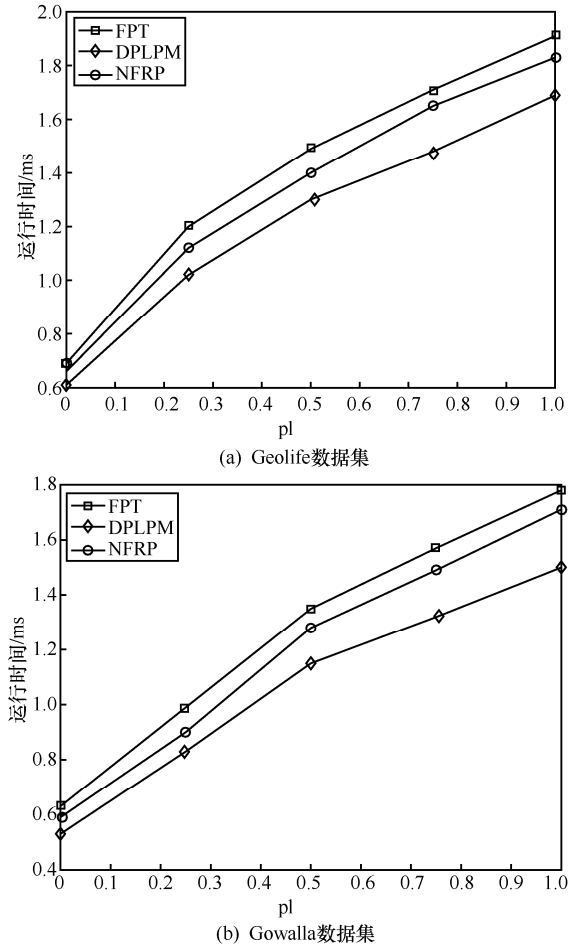


图 14　pl 对 3 种算法运行时间的影响

7　结束语

本文研究了连续位置隐私保护问题，基于差分隐私机制，提出了一种连续位置隐私保护机制。在位置数据划分模块提出了隐私级别划分算法，根据路网关系计算其相邻位置的隐私级别，使攻击者无法推断出用户隐私位置；在位置数据保护模块提出

表 2　　**相关工作比较**

方法	位置连续性	数据可用性	控制隐私预算	计算开销	结合路网实际情况
文献[17]方法	√	√	×	×	√
文献[12]方法	√	√	×	√	√
文献[18]方法	×	√	√	×	√
文献[19]方法	×	√	×	√	×
DPLPM	√	√	√	√	√

差分隐私位置保护机制 DPLPM，根据道路关系映射位置树，根据树的高度为敏感路段分配隐私预算，并添加符合差分隐私机制的 Laplace 噪声，保护了用户的位置信息。理论分析证明，本文算法均满足 ε-差分隐私。仿真实验证明，本文所提 DPLPM 有较好的隐私保护程度和较高的数据可用性。

参考文献：

[1] 何泾沙, 杜晋晖, 朱娜斐. 基于 k 匿名的准标识符属性个性化实现算法研究[J]. 信息网络安全, 2020, 20(10): 19-26.
HE J S, DU J H, ZHU N F. Research on k-anonymity algorithm for personalized quasi-identifier attributes[J]. Netinfo Security, 2020, 20(10): 19-26.

[2] 邹劲松, 李芳. 基于可伸缩 *l*-多样性的大数据发布隐私保护[J]. 计算机应用研究, 2021, 38(2): 564-566,571.
ZOU J S, LI F. Big data publishing privacy protection based on scalable *l*-diversity[J]. Application Research of Computers, 2021, 38(2): 564-566,571.

[3] 汪逸飞, 罗永龙, 俞庆英, 等. 基于信息熵抑制的轨迹隐私保护方法[J]. 计算机应用, 2018, 38(11): 3252-3257.
WANG Y F, LUO Y L, YU Q Y, et al. Trajectory privacy-preserving method based on information entropy suppression[J]. Journal of Computer Applications, 2018, 38(11): 3252-3257.

[4] 王洁, 王春茹, 马建峰, 等. 基于位置语义和查询概率的假位置选择算法[J]. 通信学报, 2020, 41(3): 53-61.
WANG J, WANG C R, MA J F, et al. Dummy location selection algorithm based on location semantics and query probability[J]. Journal on Communications, 2020, 41(3): 53-61.

[5] HU Z W, YANG J, ZHANG J P. Trajectory privacy protection method based on the time interval divided[J]. Computers & Security, 2018, 77: 488-499.

[6] 李婕, 白志宏, 于瑞云, 等. 基于 PSO 优化的移动位置隐私保护算法[J]. 计算机学报, 2018, 41(5): 1037-1051.
LI J, BAI Z H, YU R Y, et al. Mobile location privacy protection algorithm based on PSO optimization[J]. Chinese Journal of Computers, 2018, 41(5): 1037-1051.

[7] HE W. Research on LBS privacy protection technology in mobile social networks[C]//2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference. Piscataway: IEEE Press, 2017: 73-76.

[8] DWORK C, KENTHAPADI K, MCSHERRY F, et al. Our data, ourselves: privacy via distributed noise generation[M]. Berlin: Springer, 2006.

[9] 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. New York: ACM Press, 2012: 213-221.

[10] 霍峥, 孟小峰. 一种满足差分隐私的轨迹数据发布方法[J]. 计算机学报, 2018, 41(2): 400-412.
HUO Z, MENG X F. A trajectory data publication method under differential privacy[J]. Chinese Journal of Computers, 2018, 41(2): 400-412.

[11] 吴云乘, 陈红, 赵素云, 等. 一种基于时空相关性的差分隐私轨迹保护机制[J]. 计算机学报, 2018, 41(2): 309-322.
WU Y C, CHEN H, ZHAO S Y, et al. Differentially private trajectory protection based on spatial and temporal correlation[J]. Chinese Journal of Computers, 2018, 41(2): 309-322.

[12] XIAO Y H, XIONG L. Protecting locations with differential privacy under temporal correlations[C]//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2015: 1298-1309.

[13] ZHENG Y, XIE X, MA W Y. Geolife: a collaborative social networking service among user, location and trajectory[J]. IEEE Data Engineering Bulletin, 2010, 33(2): 32-39.

[14] CHO E, MYERS S A, LESKOVEC J. Friendship and mobility: user movement in location-based social networks[C]//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2011: 1082-1090.

[15] 石秀金, 徐嘉敏, 王锐, 等. 基于噪声前缀树的轨迹数据发布隐私保护算法研究[J]. 智能计算机与应用, 2019, 9(2): 9-15.
SHI X J, XU J M, WANG R, et al. Research on privacy protection algorithm of trajectory data distribution based on noise prefix tree[J]. Intelligent Computer and Applications, 2019, 9(2): 9-15.

[16] 兰微, 林英, 包聆言, 等. 融入兴趣区域的差分隐私轨迹数据保护方法[J]. 计算机科学与探索, 2020, 14(1): 59-72.
LAN W, LIN Y, BAO L Y, et al. Trajectory-differential privacy-protection method with interest region[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(1): 59-72.

[17] 程保容, 叶阿勇, 张强, 等. 一种基于隐私拆分的轨迹隐私保护方法[J]. 福建师范大学学报(自然科学版), 2020, 36(6): 28-35.

CHENG B R, YE A, ZHANG Q, et al. A trajectory privacy preserving algorithm based on privacy splitting[J]. Journal of Fujian Normal University (Natural Science Edition), 2020, 36(6): 28-35.

[18] 胡德敏, 廖正佳. 不规则线段树的差分隐私位置隐私保护方法[J]. 小型微型计算机系统, 2020, 41(2): 333-337.
HU D M, LIAO Z J. Differential privacy of location privacy protection method for irregular segment tree[J]. Journal of Chinese Computer Systems, 2020, 41(2): 333-337.

[19] 张双越, 田丰, 吴振强. 一种基于差分隐私机制的自适应轨迹数据发布算法[J]. 陕西师范大学学报(自然科学版), 2018, 46(5): 9-15,21.
ZHANG S Y, TIAN F, WU Z Q. An adaptive trajectory data publishing algorithm based on differential privacy[J]. Journal of Shaanxi Normal University (Natural Science Edition), 2018, 46(5): 9-15,21.

[作者简介]

李洪涛（1984− ），男，山东临沂人，博士，山西师范大学副教授、硕士生导师，主要研究方向为网络信息安全、大数据安全及隐私保护、物联网安全等。

任晓宇（1996− ），女，山西大同人，山西师范大学硕士生，主要研究方向为大数据安全及隐私保护、物联网安全等。

王洁（1977− ），女，山西霍州人，博士，山西师范大学副教授、硕士生导师，主要研究方向为网络信息安全、数据隐私保护。

马建峰（1963− ），男，陕西西安人，博士，西安电子科技大学教授、博士生导师，主要研究方向为信道编码、密码学、无线和移动安全、系统可生存性等。