

$$\min_{\mathbf{R}} \sum_{i=1}^{N_{i1}} \sum_{j=1}^{N_{i2}} p_i^{(\omega-)} \cdot r_{ij}^i \cdot dis(q_i^{(\omega)}, h_j^{(\omega)}) \quad (7)$$

约束条件为式(6)及每个元素的值不小于0且每行的和为1. 其中 $dis(\cdot)$ 函数是两个区域中心的欧几里得距离. 这样我们就将 t 时刻的差分隐私位置发布机制转换成了上式的基于欧几里得距离(与第3.4节的可用性对应)的优化问题. 上述目标函数的解 $\mathbf{R}^{(\omega)*}$ 即不仅使得 t 时刻的真实位置发布满足其在所在区域的差分隐私强度, 还能使得之前的 $\omega-1$ 个时刻的真实位置满足各自区域的差分隐私强度.

5.2 差分隐私位置发布机制 DPLRM

为了求解上述目标函数, 首先应当求解当 t 时刻发布位置 o_t 已知时, t 时刻真实位置 z_t 的后验概率分布. 根据贝叶斯公式有

$$\begin{aligned} \Pr(z_t = q_i^{(\omega)} | o_t = h_j^{(\omega)}) &= \frac{\Pr(o_t = h_j^{(\omega)} | z_t = q_i^{(\omega)}) \Pr(z_t = q_i^{(\omega)})}{\sum_{a=1}^{N_{i1}} \Pr(o_t = h_j^{(\omega)} | z_t = q_a^{(\omega)}) \Pr(z_t = q_a^{(\omega)})} \\ &= \frac{r_{ij} p_i^{(\omega-)}}{\sum_{a=1}^{N_{i1}} r_{aj} p_a^{(\omega-)}}, \end{aligned}$$

用向量形式表示 z_t , 当给定发布位置 o_t 时, t 时刻真实位置的后验概率分布向量为

$$\begin{aligned} \mathbf{p}^{(\omega)+}(o_t = h_j^{(\omega)}) &= \Pr(z_t \in \mathbf{q}^{(\omega)} | o_t = h_j^{(\omega)}) \\ &= \frac{\mathbf{R}_j^T \cdot \mathbf{p}^{(\omega-)}}{\mathbf{R}_j^T \times (\mathbf{p}^{(\omega-)})^T} \end{aligned} \quad (8)$$

其中 \mathbf{R}_j^T 是原始矩阵第 j 列的转置, 分子表示元素乘, 分母是向量内积. 对于所有的 $o_t \in \mathbf{h}^{(\omega)}$, t 时刻的后验概率分布可以写成矩阵形式

$$\mathbf{P}^{(\omega)+} = (\mathbf{p}^{(\omega)+}(o_t = h_1^{(\omega)}), \dots, \mathbf{p}^{(\omega)+}(o_t = h_{N_{i2}}^{(\omega)}))^T \quad (9)$$

现在求当 t 时刻发布位置 o_t 已知时, $t-1$ 时刻真实位置的后验概率分布. 设 $t-1$ 时刻真实位置区域的集合为 $\mathbf{q}^{(t-1)}$, 则根据全概率公式有

$$\begin{aligned} \Pr(z_{t-1} = q_i^{(t-1)} | o_t) &= \sum_{a=1}^{N_{i1}} \Pr(z_{t-1} = q_i^{(t-1)} | z_t = q_a^{(\omega)}) \Pr(z_t = q_a^{(\omega)} | o_t), \end{aligned}$$

其中 $\Pr(z_{t-1} | z_t)$ 表示当已知 t 时刻真实位置 z_t 时, $t-1$ 时刻真实位置为 z_{t-1} 的概率. 因为真实位置的转移概率服从马尔可夫链模型, 可由贝叶斯公式得该后验概率分布.

$$\begin{aligned} \Pr(z_{t-1} = q_i^{(t-1)} | z_t = q_a^{(\omega)}) &= \frac{\Pr(z_t = q_a^{(\omega)} | z_{t-1} = q_i^{(t-1)}) \Pr(z_{t-1} = q_i^{(t-1)} | o_{t-1})}{\sum_{b=1}^{N_{(t-1)1}} \Pr(z_t = q_a^{(\omega)} | z_{t-1} = q_b^{(t-1)}) \Pr(z_{t-1} = q_b^{(t-1)} | o_{t-1})} \end{aligned}$$

值得注意的是, 这里用的是 $t-1$ 时刻的后验概率分布 $\Pr(z_{t-1} | o_{t-1})$ 作为当前的先验概率分布. 特别地, 可以得到从 t 时刻真实位置到 $t-1$ 时刻真实位置的后验概率转移矩阵 $\mathbf{X}^{t(t-1)}$, 矩阵的每列代表 t 时刻真实位置为 $z_t = q_a^{(\omega)}$, 相应 $t-1$ 时刻真实位置的转移向量. 因此我们可以将已知 o_t 时, $t-1$ 时刻的后验概率分布写成矩阵相乘的形式:

$$\mathbf{P}^{(t-1)+} = \mathbf{P}^{(t)+} \mathbf{X}^{t(t-1)} \quad (10)$$

依次类推, 可得 $t-2, \dots, t-\omega+1$ 时刻的后验概率分布的矩阵形式如下所示:

$$\begin{cases} \mathbf{P}^{(t-2)+} = \mathbf{P}^{(t)+} \mathbf{X}^{t(t-1)} \mathbf{X}^{(t-1)(t-2)} \\ \dots \\ \mathbf{P}^{(t-\omega+1)+} = \mathbf{P}^{(t)+} \mathbf{X}^{t(t-1)} \dots \mathbf{X}^{(t-\omega+2)(t-\omega+1)} \end{cases} \quad (11)$$

根据上述推导, 在 $t-1$ 时刻, 计算得出相应的差分隐私位置发布机制 $\mathbf{R}^{(t-1)}$ 后, 可以得到一个发布位置 o_{t-1} . 根据 o_{t-1} 及 $\mathbf{R}^{(t-1)}$ 可以得到 $t-\omega$ 到 $t-1$ 时刻的真实位置后验概率分布 $\mathbf{p}^{(t-1)+}, \dots, \mathbf{p}^{(t-\omega)+}$ 和 t 时刻真实位置的先验概率分布 $\mathbf{p}^{(t)-}$. 值得注意的是 $t-1$ 时求得真实位置间的概率转移矩阵 $\mathbf{X}^{(t-1)(t-2)}$ 与 t 时刻所求的不同. 在 t 时刻, 我们希望找到一个差分隐私位置发布机制 $\mathbf{R}^{(t)}$, 使得根据该机制得到的任意发布位置 o_t 都满足目标函数的条件, 且损失最小.

综上所述, 我们可以计算出目标函数式(7)的约束条件(式(6))具体值. 特别地, 因为 t 时刻之前的 $\omega-1$ 个时刻的约束条件都可以通过式(10)和式(11)转换成与 t 时刻约束条件类似的不等式, 所以目标函数约束条件的数量可以从 ωN^2 减少为 N^2 , 其中 $N \geq \max\{N_{i1}, N_{i2}\}, \forall i \in [t-\omega+1, t]$.

在获得各线性不等式约束条件后, 采用优化算法对其进行迭代求解^[35-36]. 下面给出 t 时刻时序相关的差分隐私位置发布机制 DPLRM 的主要流程(伪代码见算法2).

(1) 初始化矩阵 $\mathbf{R}^{(t)}$;

(2) 由式(9)计算 o_t 已知时 z_t 的后验概率分布 $\mathbf{P}^{(t)+}$;

(3) 由式(10)和式(11)计算时刻间的概率转移矩阵 \mathbf{X} 和 o_t 已知时 $z_{t-1}, \dots, z_{t-\omega+1}$ 的后验概率分布 $\mathbf{P}^{(t-1)+}, \dots, \mathbf{P}^{(t-\omega+1)+}$;

(4) 计算各个约束条件的值, 并根据各个时刻的后验概率矩阵 $\mathbf{P}^{(t)+}, \mathbf{P}^{(t-1)+}, \dots, \mathbf{P}^{(t-\omega+1)+}$ 更新矩阵 $\mathbf{R}^{(t)}$ 的参数;

(5) 重复(2)~(4)步, 直到达到迭代结束条件;

得到差分隐私位置发布机制 $\mathbf{R}^{(t)}$;

(6) 根据当前所在的真实位置和 $\mathbf{R}^{(t)}$ 相应的行, 依概率随机在 $\mathbf{h}^{(t)}$ 中选择一个位置发布.

算法 2. 差分隐私位置发布 (DPLRM).

输入: 真实位置的区域转移概率矩阵 \mathbf{M} , 时间窗口长度 w , 各个时刻的隐私保护预算向量 $\boldsymbol{\varepsilon} = \{\varepsilon_1, \dots, \varepsilon_t\}$, 各个时刻真实位置的可能集合 $\mathbf{Q} = \{\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(t-1)}\}$, t 时刻的先验概率分布向量 $\mathbf{p}^{(t)}$, 前 $t-1$ 时刻的后验概率分布矩阵集合 $\hat{\mathbf{P}} = \{\mathbf{P}^{(1)+}, \dots, \mathbf{P}^{(t-1)+}\}$, t 时刻的真实位置 z_t , 终止条件 $cond$

输出: t 时刻的发布位置.

1. $q_vec, h_vec = chooseLoc(\mathbf{p}^{(t)})$;
2. $N1 = q_vec.length(), N2 = h_vec.length()$;
3. $R = initialize()$;
4. IF $t = 1$ THEN:
5. $Q_used = \emptyset; P_used = \emptyset$;
6. ELSE IF $t > 1$ and $t < w$ THEN:
7. $Q_used = \mathbf{Q}; P_used = \hat{\mathbf{P}}$
8. ELSE:
9. $Q_used = selectQ(\mathbf{Q}); P_used = selectP(\hat{\mathbf{P}})$;
10. WHILE $cond \neq TRUE$:
11. $P_plus = calPosterior(\mathbf{R}, \mathbf{p}^{(t)}, q_vec, h_vec)$;
12. $X_used = \emptyset$;
- // 时刻间的真实位置概率转移矩阵集合
13. FOR all $j \in Q_used.length()$:
14. $X_used.append(compute(Q_used, P_used, P_plus))$;
15. $P_used[t-j] = calPosterior2(P_used, Q_used, P_plus)$;
- // 根据式(10)、(11)计算前面时刻的后验概率矩阵
16. END FOR
17. $Conditions = calConds(P_used, P_plus, \boldsymbol{\varepsilon}, X_used)$;
18. $updateR(\mathbf{R}, Conditions, N1, N2)$;
19. END WHILE
20. $o_t = RandomChoose(\mathbf{R}, z_t)$;
21. RETURN o_t ;

5.3 算法分析

关于时间复杂度, 在 DPLRM 算法的每次迭代中, 最耗时的部分在于对 w 个时刻计算其后验概率矩阵 $\mathbf{P}^{(t)+}, \dots, \mathbf{P}^{(t-w+1)+}$ (根据式(9)~(11)), 因此每次迭代的时间复杂度为 $O(wN^3)$, 其中 N 为 $N_{i1}, N_{i2} (\forall i \in [t-w+1, t])$ 的最大值. 值得注意的是, 尽管 DPLRM 算法的计算复杂度相对较高, 但 N 的值不会很大, 因为下一时刻的位置不会偏离上一时刻的位置太远. 另外, 也可以将算法中 while 循环这部分计算量 (即差分隐私位置发布矩阵 \mathbf{R} 的计算) 分配到服务器或进行并行计算, 而在移动端仅仅通过当前的真实位置计算最终的发布位置 (即使服务器

知道每个时刻的发布位置和 $\varepsilon_i (i \in [1, t])$, 它也很难将 ε_i 与真实位置对应起来, 因为这种对应关系只有移动端自身知道).

关于隐私性, 根据 t 时刻差分隐私位置发布机制 DPLRM, 可知 t 时刻的发布位置 o_t 既使得当前时刻的真实位置 z_t 满足 ε_t 差分隐私, 还使得之前 $w-1$ 个时刻的所有真实位置 $z_{t-w+1}, \dots, z_{t-1}$ 分别满足 $\varepsilon_{t-w+1}, \dots, \varepsilon_{t-1}$ 差分隐私. 因此, 由位置 γ -隐私模型可知, 在长度为 w 轨迹上的每个点都满足位置 γ -隐私, 即满足轨迹 γ -隐私 (见定义 3).

关于可用性, 由于 DPLRM 算法主要是通过对式(7)进行求解来生成扰乱的发布位置, 而式(7)的目标就是最小化所有情况下真实位置和发布位置间的距离总和, 因此在保证之前 $w-1$ 时刻真实位置满足 γ -隐私的前提下, DPLRM 算法发布位置的整体可用性是最优的. 当 w 增大时, 需要纳入考虑的時刻增多, 目标函数(式(7))中的约束条件可能会变得更加严格, 从而使发布位置偏离真实位置的程度和概率变大, 一定程度上降低可用性.

6 实验与分析

6.1 实验设置

CPL 算法和 DPLRM 机制均采用 Python 实现, 在 3.60GHz CPU、8.00GB RAM 的 Windows 7 平台上运行. 本文采用的数据集是 Geolife^① 和 Gowalla^② 真实数据集. 数据集 Geolife 采集了 182 个用户从 2007 年 4 月到 2012 年 8 月在北京活动的真实数据, 数据集共包含 17 621 条轨迹. Geolife 数据集中包括用户编号、时间戳、经度、纬度、海拔等属性, 我们抽取五环以内轨迹的前 4 个属性作为新的数据集. 数据集 Gowalla 采集了 2009 年 2 月到 2010 年 10 月共 15 116 个用户在移动社交网站上 (加州范围内) 的签到数据. 同 Geolife 一样, 我们抽取洛杉矶范围内的用户编号、时间戳、经度、纬度作为新的数据集.

对于 CPL 算法, 我们将地图转换成无向图表示, 并将每个用户停留时间最长或访问次数最多的 $k (k \in [5, 10, 15, 20, 25])$ 个区域作为初始敏感位置集合, 且设其隐私级别等于 1, 我们衡量阈值参数 δ ($\delta \in [0.1, 0.2, 0.3, 0.4, 0.5]$) 对算法运行时间的影

① <http://research.microsoft.com/en-us/downloads/>

② <http://snap.stanford.edu/data/loc-gowalla.html>

响. 对于 DPLRM 机制, 与文献[30]类似, 我们将北京地图划分为大小为 $0.34 \text{ km} \times 0.34 \text{ km}$ 的网格, 将洛杉矶地图划分为 $0.89 \text{ km} \times 0.89 \text{ km}$ 的网格, 并在此基础上获得相对应的马尔可夫概率转移矩阵. 我们衡量 500 个时间戳内 γ ($\gamma \in [0.2, 0.4, 0.6, 0.8, 1.0]$) 对位置可用性的影响、CPL 算法阈值参数 δ ($\delta \in [0.1, 0.2, 0.3, 0.4, 0.5]$) 和初始敏感集合大小 k ($k \in [5, 10, 15, 20, 25]$) 对位置可用性的影响及时间窗口 w ($w \in [1, 3, 5, 7, 9]$) 对算法运行时间和位置可用性的影响. 其中, 位置可用性的定义为时间戳内敏感区域上真实位置和发布位置之间的距离总误差的均值 $RMSE$ (见第 3.4 节式(2)), $RMSE$ 越小, 位置可用性越高.

6.2 实验结果和分析

6.2.1 CPL 算法表现

我们首先分析阈值参数 δ 对 CPL 算法运行时间的影响, 结果如图 4 所示. 在这一系列实验中, 初始敏感位置集合 SL^{initial} 大小 k 的默认值是 10. 由图 4 可以看出, 当 δ 逐渐增加时, CPL 算法在两个数据集上的运行时间逐渐降低. 这是因为, 当 δ 变大时, CPL 算法的剪枝操作很有效 (见算法 1 的第 8 行), 因此算法的运行效率会有很大的提高. 同时, 我们可以看出 CPL 算法在 Geolife 数据集上的运行时间略大于在 Gowalla 数据集上的运行时间, 这主要是因为北京在地理上的语义集合略大于洛杉矶的, 同样的遍历, Geolife 数据集所需要的时间稍多.

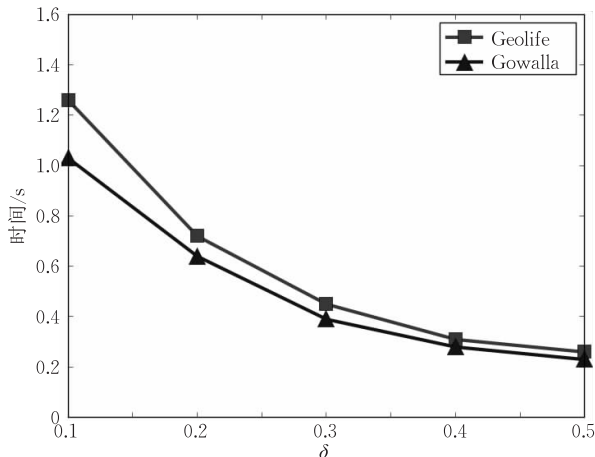


图 4 阈值 δ 对 CPL 算法运行时间的影响

然后, 我们分析初始敏感位置集合 SL^{initial} 大小 k 对 CPL 算法运行时间的影响, 结果如图 5 所示. 在这一系列实验中, δ 的默认值设为 0.2. 由图 5 可以看出, 当 k 不断增加时, CPL 算法在两个数据集上的运行时间也相应增加. 这是因为 k 很大时, 算法所遍历的地理空间增加, 进而所需时间也增加.

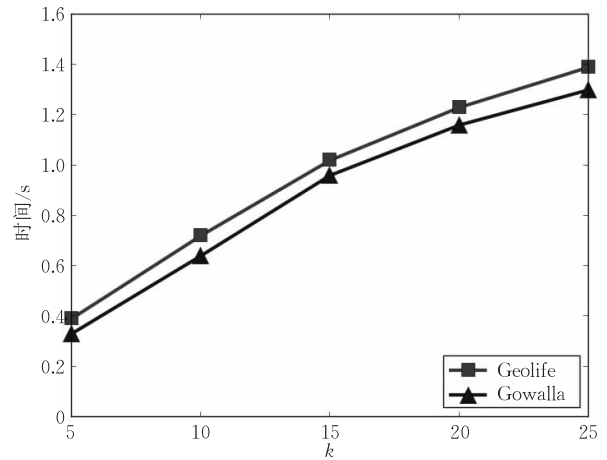


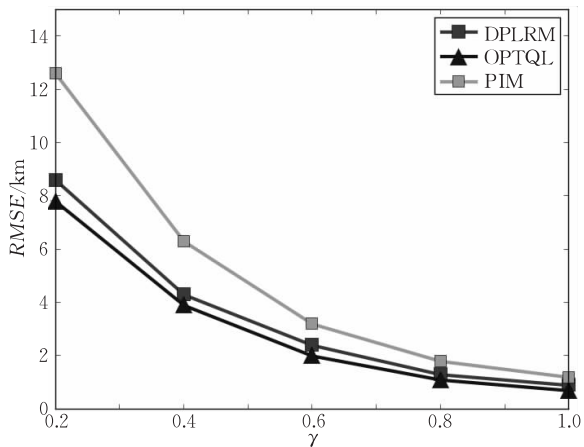
图 5 k 对 CPL 算法运行时间的影响

6.2.2 DPLRM 算法表现

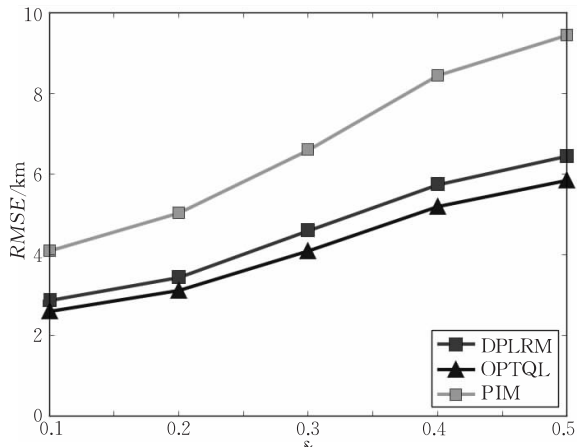
对于 DPLRM 算法, 我们主要关注隐私模型参数 γ 、CPL 算法阈值 δ 和时间窗口长度 w 的影响. 在衡量 γ 和 δ 对位置可用性的影响时, 我们将 DPLRM 与 OPTQL 算法^[18] 和 PIM 算法^[30] 对比. OPTQL 考虑单个时刻的最优差分隐私位置发布, PIM 是一种基于敏感度包的差分隐私位置发布机制.

我们首先分析 γ 对位置可用性的影响, 结果如图 6 所示. 在这一系列实验中, 为方便衡量 γ 的影响, 我们假设时间窗口长度 $w=3$, 且阈值 $\delta=1.0$, 也就是说只有初始设定的敏感集合需要采用三种算法. 此时的 γ 与传统差分隐私中的隐私保护预算 ϵ 等价. 由图 6(a) 可以看出, PIM 算法的位置可用性相对最差, 这是因为 PIM 算法仅提出了一种满足差分隐私的位置发布机制, 在执行该发布机制时并未将位置的可用性考虑在内; OPTQL 算法的位置可用性最好, 这是因为该算法的目标就是在满足差分隐私的同时最小化发布位置的误差; DPLRM 算法介于两个算法之间, 并接近于 OPTQL 算法的表现. 这是因为 DPLRM 不仅考虑了当前发布位置对当前时刻的隐私影响, 还考虑了当前发布位置对之前发布位置的隐私影响, 所以位置可用性略弱于 OPTQL 算法. 图 6(b) 在 Gowalla 数据集上显示了类似的表现. 另外, 三种算法在 Geolife 数据集上的表现优于在 Gowalla 数据集上的表现, 原因是在 Geolife 上划分的网格面积小于在 Gowalla 上的划分, 所以网格更密集、结果更精确.

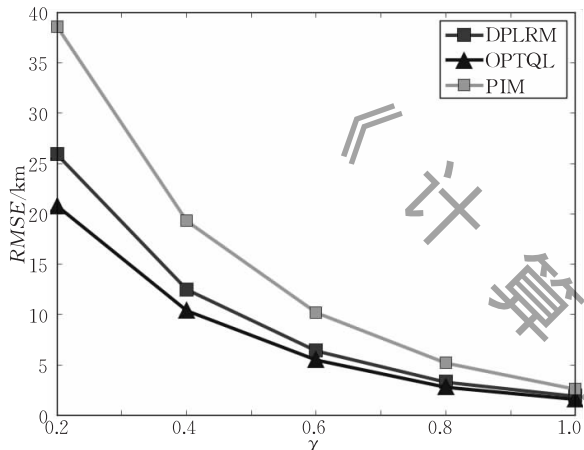
然后, 我们分析阈值 δ 对三种算法位置可用性的影响, 结果如图 7 所示. 在这一系列实验中, 我们设定 $\gamma=0.2, w=3$; 并且三种算法采用相同的 CPL 算法结果, 即在敏感区域集合上的差分隐私预算根



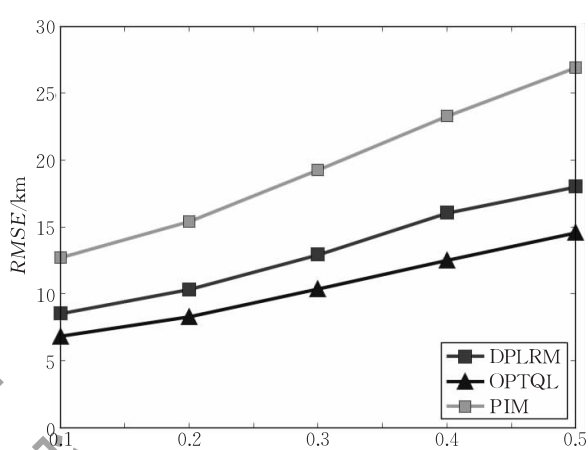
(a) Geolife数据集



(a) Geolife数据集



(b) Gowalla数据集



(b) Gowalla数据集

图 6 γ 对 RMSE 的影响

图 7 阈值 δ 对 RMSE 的影响

据定义 2 获得. 由图 7(a)可以看出, δ 越小时, 三种算法的 RMSE 越小. 这是因为当 δ 小时, CPL 算法所得结果中会包含较多的隐私级别 pl 较低的区域; 又由定义 2 可知, 各区域的差分隐私预算 ϵ 与其隐私级别 pl 成反比, 因而在这种情况下, 会有较多的区域拥有较大的差分隐私预算 ϵ , 最终使得平均的位置可用性提高(即 RMSE 降低). 图 7(b)在 Gowalla 数据集上显示了相似的结果.

时刻发布位置对之前真实位置的后验概率, 所以所需时间也越长. 值得注意的是, w 的大小对最优化问题的求解影响相对较小, 如第 5.2 节所述, 最优化问题的条件最终可以约减到 N^2 , 因此, 这部分的运行时间与 w 的值关系较小.

最后, 我们分析时间窗口长度 w 对位置可用性的影响, 结果如图 8 所示. 在这一系列实验中, 我们假设 $\gamma=0.2$ 、 $\delta=0.2$. 由图 8 可知, 当 w 逐渐增大时, RMSE 也相应增加, 这是因为当 w 很大时, DPLRM 算法求得的解可能要更偏离真实位置才能满足当前发布位置在 w 个时刻上都满足差分隐私, 因而位置误差相对较大. 图 9 显示了时间窗口长度 w 对 DPLRM 算法运行时间的影响. 如图所示, 当 w 增大时, 算法所需的运行时间也越长, 因为 DPLRM 算法要逆向在 w 时间窗口内的每个时刻求解当前

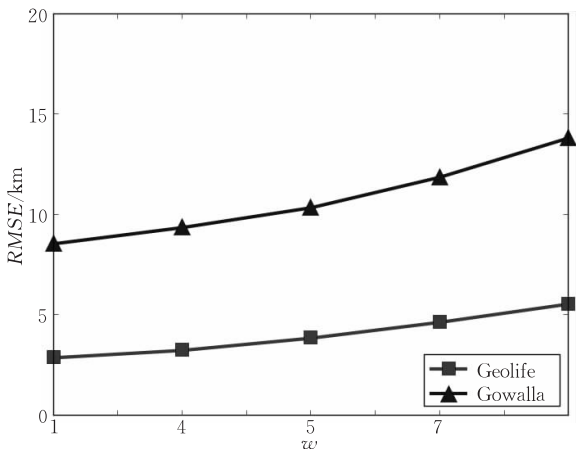


图 8 时间窗口长度 w 对 RMSE 的影响

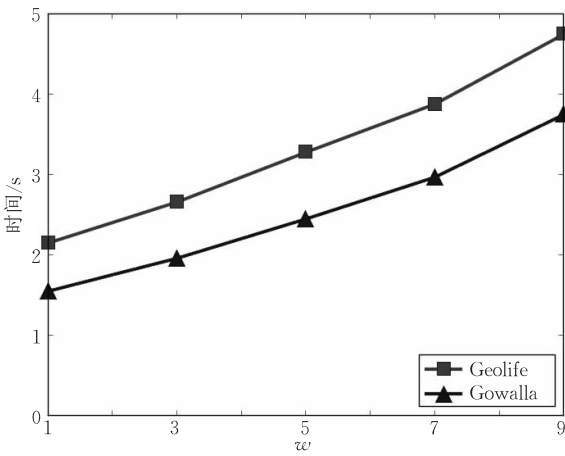


图9 时间窗口长度 ω 对 DPLRM 运行时间的影响

7 总 结

本文针对基于位置的服务中的轨迹隐私保护问题,提出了一种基于地理空间拓扑关系的区域隐私级别计算方法 CPL 以及结合区域隐私级别与差分隐私预算的 γ -隐私模型. CPL 算法利用无向图表示地理间的拓扑关系,并根据节点的度和节点间的距离将预先设定的敏感区域的隐私级别分配给相邻节点,以使得攻击者无法根据地理限制推测用户的隐私.然后,分析了当前发布位置对轨迹上真实位置的隐私影响,提出了一个基于可用性的差分隐私位置发布机制的最优化问题,以及相应的算法 DPLRM.实验结果表明,DPLRM 在达到轨迹差分隐私效果的同时具有较好的可用性.今后的研究将考虑如下两个方面:(1)在根据地理拓扑关系计算初始敏感位置附近位置的隐私级别时,将用户的移动模式(如交通方式)以及时间因素(如用户可以自定义初始敏感位置的隐私级别随时间变化的曲线)考虑在内,以提供更细致的隐私级别计算;(2)研究如何对 DPLRM 算法进行继续优化,以提高扰乱位置发布的可用性并减少算法的运行时间,更好地扩展到实时的位置服务中.

参 考 文 献

- [1] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking//Proceedings of the 1st International Conference on Mobile Systems, Applications, and Services (MobiSys2003). San Francisco, USA, 2003: 31-42
- [2] Mokbel M F, Chow C Y, Aref W G. The new casper: Query processing for location services without compromising privacy //Proceedings of the 32nd Conference of Very Large Data Bases (PVLDB2006). Seoul, South Korea, 2006: 763-77
- [3] Huo Zheng, Meng Xiao-Feng. A survey of trajectory privacy preserving techniques. Chinese Journal of Computers, 2011, 34(10): 1820-1830(in Chinese)
(霍峥, 孟小峰. 轨迹隐私保护技术研究. 计算机学报, 2011, 34(10): 1820-1830)
- [4] Chow C Y, Mokbel M F. Enabling private continuous queries for revealed user locations//Proceedings of the 10th International Symposium on Spatial and Temporal Databases (SSTD 2007). Boston, USA, 2007: 258-275
- [5] Pan X, Meng X, Xu J. Distortion-based anonymity for continuous queries in location-based mobile services//Proceedings of the 17th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems (ACM-GIS 2009). Seattle, USA, 2009: 256-265
- [6] Bamba B, Liu L, Pesti P, Wang T. Supporting anonymous location queries in mobile environments with PrivacyGrid//Proceedings of the 17th International Conference on World Wide Web (WWW 2008). Beijing, China, 2008: 237-246
- [7] Ghinita G, Kalnis P, Skiadopoulos S. PRIVE: Anonymous location-based queries in distributed mobile systems//Proceedings of the 16th International Conference on World Wide Web (WWW 2007). Banff, Canada, 2007: 371-380
- [8] Huo Zheng, Meng Xiao-Feng, Huang Yi. PrivateCheckIn: Trajectory privacy-preserving for check-in services in MSNS. Chinese Journal of Computers, 2013, 36(4): 716-726(in Chinese)
(霍峥, 孟小峰, 黄毅. PrivateCheckIn: 一种移动社交网络中的轨迹隐私保护方法. 计算机学报, 2013, 36(4): 716-726)
- [9] Freudiger J, Raya M, Felegyhazi M, et al. Mix-zones for location privacy in vehicular networks//Proceedings of the 1st International Workshop on Wireless Networking for Intelligent Transportation Systems (WiN-ITS 2007). Vancouver, Canada, 2007: 1-7
- [10] Freudiger J, Shokri R, Hubaux J P. On the optimal placement of mix zones//Proceedings of the 9th International Symposium on Privacy Enhancing Technologies (PETS 2009). Seattle, USA, 2009: 216-234
- [11] Palanisamy B, Liu L. MobiMix: Protecting location privacy with mix zones over road networks//Proceedings of the 27th International Conference on Data Engineering (ICDE 2011). Hannover, Germany, 2011:494-505
- [12] Chen R, Fung B C M, Mohammed N, et al. Privacy-preserving trajectory data publishing by local suppression. Information Science, 2013, 231: 83-97
- [13] Gruteser M, Liu X. Protecting privacy in continuous location-tracking applications. IEEE Security and Privacy, 2004, 2(2): 28-34

- [14] Zhao Jing, Zhang Yuan, Li Xing-Hua, Ma Jian-Feng. A trajectory privacy protection approach via trajectory frequency suppression. *Chinese Journal of Computers*, 2014, 37(10): 2096-2106(in Chinese)
(赵婧, 张渊, 李兴华, 马建峰. 基于轨迹频率抑制的轨迹隐私保护方法. *计算机学报*, 2014, 37(10): 2096-2106)
- [15] Lee K C K, Lee W C, Leong H V, Zheng B. Navigational path privacy protection//*Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*. Hong Kong, China, 2009: 691-700
- [16] Shokri R, Theodorakopoulos G, Troncoso C, et al. Protecting location privacy: Optimal strategy against localization attacks//*Proceedings of the ACM Conference on Computer and Communications Security (CCS2012)*. Raleigh, USA, 2012: 617-627
- [17] Andrés M E, Bordenabe N E, Chatzikokolakis K, Palamidessi C. Geo-indistinguishability: Differential privacy for location-based systems//*Proceedings of the ACM Conference on Computer and Communications Security (CCS 2013)*. Berlin, Germany, 2013: 901-914
- [18] Bordenabe N E, Chatzikokolakis K, Palamidessi C. Optimal geo-indistinguishable mechanisms for location privacy//*Proceedings of the ACM Conference on Computer and Communications Security (CCS 2014)*. Scottsdale, USA, 2014: 251-262
- [19] Theodorakopoulos G, Shokri R, Troncoso C, et al. Prolonging the hide-and-seek game: Optimal trajectory privacy for location-based services//*Proceedings of the 13th Workshop on Privacy in the Electronic Society (WPES2014)*. Scottsdale, USA, 2014: 73-82
- [20] Huo Z, Meng X, Hu H, Huang Y. You can walk alone: Trajectory privacy-preserving through significant stays protection//*Proceedings of the 17th International Conference on Database Systems for Advanced Applications (DASFAA 2012)*. Busan, South Korea, 2012: 351-366
- [21] Dwork C, Kenthapadi K, McSherry F, et al. Our data, ourselves: Privacy via distributed noise generation//*Proceedings of the 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT 2006)*. Petersburg, Russia, 2006: 486-503
- [22] Dwork C, McSherry F, Nissim K, Smith A D. Calibrating noise to sensitivity in private data analysis//*Proceedings of the 3rd Theory of Cryptography Conference (TCC 2006)*. New York, USA, 2006: 265-284
- [23] Dwork C, Roth A. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 2014, 9(3-4): 211-407
- [24] Chen R, Fung B C M, Desai B C, Sossou N M. Differentially private transit data publication: A case study on the Montreal transportation system//*Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2012)*. Beijing, China, 2012: 213-221
- [25] Chen R, Acs G, Castelluccia C. Differentially private sequential data publication via variable-length n -grams//*Proceedings of the ACM Conference on Computer and Communications Security (CCS 2012)*. Raleigh, USA, 2012: 638-649
- [26] Shao D, Jiang K, Kister T, et al. Publishing trajectory with differential privacy: a priori vs. a posteriori sampling mechanisms //*Proceedings of the 24th International Conference on Database and Expert Systems Applications (DEXA 2013)*. Prague, Czech Republic, 2013: 357-365
- [27] Jiang K, Shao D, Bressan S, et al. Publishing trajectories with differential privacy guarantees//*Proceedings of the Conference on Scientific and Statistical Database Management (SSDBM2013)*. Baltimore, USA, 2013: 12: 1-12
- [28] He X, Cormode G, Machanavajjhala A, et al. DPT: Differentially private trajectory synthesis using hierarchical reference systems//*Proceedings of the 41st International Conference on Very Large Data Bases (VLDB 2015)*. Hawaii, USA, 2015: 1154-1165
- [29] Hua J, Gao Y, Zhong S. Differentially private publication of general time-serial trajectory data//*Proceedings of the IEEE Conference on Computer Communications (INFOCOM2015)*. Hong Kong, China, 2015: 549-557
- [30] Xiao Y, Xiong L. Protecting locations with differential privacy under temporal correlations//*Proceedings of the 22nd ACM Conference on Computer and Communications Security (CCS2015)*. Denver, USA, 2015: 1298-1309
- [31] You T H, Peng W C, Lee W C. Protecting moving trajectories with dummies//*Proceedings of the 8th International Conference on Mobile Data Management (MDM2007)*. Mannheim, Germany, 2007: 278-282
- [32] Kellaris G, Papadopoulos S, Xiao X, Papadias D. Differentially private event sequences over infinite streams//*Proceedings of the 40th International Conference on Very Large Data Bases (VLDB 2014)*. Hangzhou, China, 2014: 1155-1166
- [33] Gotz M, Nath S, Gehrke J. Maskit: Privately releasing user context streams for personalized mobile applications//*Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 2012)*. Scottsdale, USA, 2012: 289-300
- [34] Shokri R, Theodorakopoulos G, Boudec J Y L, Hubaux J P. Quantifying location privacy//*Proceedings of the 32nd IEEE Symposium on Security and Privacy (S&P 2011)*. Berkeley, USA, 2011: 247-262
- [35] Li Hang. *Statistical Learning Methods*. Beijing: Tsinghua University Press, 2012(in Chinese)
(李航. *统计学习方法*. 北京: 清华大学出版社, 2012)
- [36] Platt J C. Sequential minimal optimization: A fast algorithm for training support vector machines. Seattle, USA: Microsoft Research, Technical Report: MSR-TR-98-14, 1998



WU Yun-Cheng, born in 1989, Ph. D. candidate. His research interest includes private data analysis and data management in IoTs.

CHEN Hong, born in 1965, Ph. D. , professor, Ph. D. supervisor. Her research interests include database, data warehouse, Internet of Things and privacy preservation.

ZHAO Su-Yun, born in 1979, Ph. D. , associate professor. Her research interests include fuzzy sets, rough sets and

uncertainty data processing.

LIANG Wen-Juan, born in 1980, Ph. D. candidate. Her research interests include data privacy preservation and database.

WU Yao, born in 1990, Ph. D. candidate. His research interests include crowd sensing and big data management.

LI Cui-Ping, born in 1971, Ph. D. , professor, Ph. D. supervisor. Her research interests include social network analysis, data mining and recommender systems.

ZHANG Xiao-Ying, born in 1987, Ph. D. , lecturer. Her research interests include wireless sensor network and privacy preservation.

Background

Protecting location privacy and trajectory privacy in location-based services (LBS) has become a hot spot of research. There are two types of LBS, namely, snapshot and continuous LBS. For a snapshot LBS, a mobile user only needs to report its current location to a service provider once to get its desired information. For continuous LBS, a mobile user has to report its location in a periodic manner to obtain the desired information. Privacy preserving techniques for LBS can be classified into four categories. (1) Location Obfuscation. The basic idea is to send a cloak region instead of the real location to the service provider. (2) Mix Zones. This technique is to change the pseudonyms when several users enter a mix-zone, ensuring unlinkability between the incoming users and outgoing users. (3) Suppression. The basic idea is not to report the current location if user is in a sensitive area. (4) Perturbation. This technique sends a false location that is close to the real location to the service provider. At present, the existing studies mainly focus on privacy preserving snapshot queries. Differential privacy, a popular paradigm for providing privacy with strong theoretical guarantees, has recently gained significant attention in snapshot LBS. Several differentially private approaches that generates a false location according to the real location have been proposed, however, these snapshot approaches cannot be directly applied to the trajectory privacy protection scenario (continuous LBS), since they seldom consider the geo-spatial

and temporal correlation of the locations between several timestamps.

Aiming at solving the trajectory privacy problem, this paper firstly propose CPL algorithm to calculate the privacy level of each region in the map according to geo-spatial correlation, and define a privacy model that integrates privacy level and differential privacy. In addition, we analyze the effect of released location on the true locations based on Markov chain, and present a differentially private location release mechanism DPLRM, to protect the privacy of true locations and trajectory. Experimental results on real datasets demonstrate the performance of proposed mechanism. Although this paper provides a possible way to protect trajectory privacy using the notion of differential privacy, there are two improvements that could be explored to enhance the performance of this approach. Firstly, the concrete mobility pattern of user, e. g. , the way of transportation on a specific timestamp can be integrated into the calculation of privacy level, which might improve the accuracy of privacy level. Secondly, how to reduce the computational cost for DPLRM is also an important direction.

This work is supported by the National Natural Science Foundation of China (Grant Nos.61532021, 61702522, 61772537, 61772536), the National High Technology Research and Development Program of China (863 Program) (No.2014AA015204).