基于本地化差分隐私的时序位置发布方案研究

康海燕,冀源蕊
(北京信息科技大学信息管理学院信息安全系,北京 100192)

摘 要: 为了解决基于位置的服务(Location Based Service, LBS)在收集用户位置数据时造成的隐私泄露,提出一种本地化差分隐私位置发布模型. 首先,该模型采用了灵活的位置隐私保护方案(个性化隐私设置),即由用户选择已设定的多种隐私策略或定制隐私策略,在此基础上设计了定制隐私策略位置扰动算法(Customized Privacy policy Location Perturbation algorithm, CPLP);其次,提出并设计一种基于隐马尔可夫模型的时序关联位置隐私发布算法(Temporal Relational Location Privacy publishing algorithm,TRLP),解决发布时序位置时产生的隐私泄露;最后,在GeoLife数据集和Gowalla数据集上通过对比实验验证了该模型的有效性.

关键词: 差分隐私; 位置服务; 时序数据; 隐私发布; 隐马尔可夫模型
中图分类号: TP309.2 **文献标识码:** A **文章编号:** 0372-2112(2022)09-2222-11
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20210338

Research on Time-Serial Location Data Publication Based on Local Differential Privacy

KANG Hai-yan, JI Yuan-rui
(*School of Information Management, Beijing Information Science and Technology University, Beijing* 100192, *China*)

Abstract: In order to solve the privacy leakage problem in location based service(LBS) when collecting user's location data, we proposed a time-serial location data publication model based on local differential privacy. Firstly, the model adapts a flexible location privacy preservation method, allows users to choose or customize their privacy policy(personalized privacy Settings), based on customized privacy policy, we designed a customized privacy policy location perturbation algorithm(CPLP); Secondly, we proposed and designed temporal relational location privacy publishing algorithm(TRLP) based on hidden Markov model(HMM), which can reduce the privacy leakage when releasing the time-serial location data. Finally, we verified the usability of the algorithm on data sets Geolife and Gowalla.

Key words: differential privacy; location based service; time-series data; privacy releasing; hidden markov model

1 引言

随着GPS定位技术和可穿戴设备的广泛应用,基于位置的服务(Location Based Service, LBS)可以提供诸如实时位置共享、路线导航和兴趣点查询等服务,为人们的生活提供了便利. 与此同时,这些移动设备不断记录着用户的位置数据,LBS服务通过收集并发布这些位置数据可以为数据分析提供基础数据,结合数据挖掘和机器学习等技术对位置轨迹大数据的分析,企业可以用挖掘有价值的商业信息,政府部门可以通过分析交通数据进行道路规划[1],在COVID-19疫情防控背景下,政府部门可以通过对用户位置数据的收集和分析实现接触者追踪和疫情传播监控[2]. 然而位置数据中包含大量个体的隐私信息,如果不加保护直接发布会造成大量用户的隐私泄露,用户对隐私问题的顾虑限制了其分享个人位置数据的意愿,阻碍了位置大数据的收集和分析工作,因此针对位置数据发布时的隐私保护研究具有必要性.

最早用于解决位置数据隐私发布的方案是k-匿名技术,如Gedik等[3]提出的匿名位置发布方法,通过对轨迹Ti在任意时刻采样,使得在该时刻内,至少有$k-1$条轨迹在相应的位置能和Ti泛化在同一区域内然而k-匿名理论面临的最大问题是,一旦攻击者能力超过了预先的假设,就能够进一步区分等价类内的不同记录,实

收稿日期:2021-03-10;修回日期:2021-07-30;责任编辑:梅志强
基金项目:国家社科基金(No.21BTQ079);教育部人文社科项目(No.20YJAZH046);国家自然科学基金(No.61370139)

现去匿名化. 由 Dwork 等[4]提出的差分隐私技术因其严格的数学定义备受青睐,是目前最受欢迎的隐私保护技术,在隐私保护的数据分析与数据发布领域应用广泛[5]. 而在位置轨迹隐私发布背景下,基于差分隐私模型的研究同样取得了许多成果[6~17]. 然而现有的差分隐私模型下位置轨迹发布的方案存在以下不足:(1)现有方案基于第三方 LBS 服务可信的前提假设,对收集到的位置数据进行集中加噪处理后再发布,没有考虑第三方服务器不可信的情况.(2)现有方案在隐私保护选择上不够灵活,只允许用户根据单一的隐私预算决定隐私保护程度.(3)现有方案将轨迹发布看作一系列连续位置的发布,没有考虑连续提交多个位置数据的情况下时序关联对位置隐私发布的影响.

为了解决以上问题,本文进行了深入研究,主要贡献如下:(1) 提出一种定制隐私策略位置扰动算法(Customized Privacy policy Location Perturbation algorithm, CPLP),允许用户在本地对位置数据加噪后上传给 LBS 服务器,通过定制隐私策略实现多隐私因子支持.(2) 结合隐马尔可夫模型提出一种时序关联位置隐私发布算法(Temporal Relational Location Privacy publishing algorithm, TRLP),解决时序关联对位置隐私发布的影响.(3) 在 GeoLife 数据集和 Gowalla 数据集上通过大量的对比实验验证了该模型的可用性.

2　研究现状

2.1　中心化差分隐私位置数据发布

中心化差分隐私模型下的位置数据发布形式包括空间直方图、地理位置熵和轨迹数据集. 空间直方图是位置轨迹发布的经典形式,差分隐私模型通过对空间直方图添加随机噪声,发布加噪后的直方图来抵御攻击者对于数据集中是否包含某用户的推测攻击. 然而直接对构成直方图的每个网格单元添加噪声导致查询结果的误差太大,为了提高任意范围内查询结果的精度,通常采用四叉树[9]对划分后的网格提供索引服务,树中的每个节点代表本区域范围内的位置轨迹数目,由于该数目涉及用户的位置隐私,通过对节点计数查询添加噪声实现差分隐私,在响应用户的计数查询时,采用类似索引树的查询模式自上而下在四叉树中搜索与该查询节点匹配的节点集合,根据加噪后的节点计数结果进行回答. 为了解决位置数据分布不均衡导致四叉树中添加噪声过大的问题,Hay 等[10]提出一种基于 k-叉平均树的位置数据构建方法,使得划分后直方图各个区间内位置轨迹数目比较均衡,同时通过最优线性无偏估计对其进行一致性修正,降低中间节点造成的查询噪声误差. Zhang 等[11]提出一种基于不完全四叉树的位置发布方法 PrivTree,通过对高维空间数据进行合理的划分摆脱了加入噪声时树高的影响,并采用局部敏感度和近似误差的方法降低噪声误差. 为了解决轨迹中多个单元的子轨迹被重复计数导致大范围内轨迹聚集查询误差较大的问题,Xie 等[12]提出了层次化模型(Euler Histograms Tree, EHT),通过降低子轨迹重复带来的查询误差以支持矩形空间聚集查询. 除了直接发布位置轨迹的计数信息,位置相关统计信息还包括位置熵(Location Entropy, LE),位置熵以熵的形式衡量地理位置受用户欢迎的程度,熵越大,说明该地理位置越受用户欢迎. 为了防止攻击者根据发布的位置熵直方图推测出某个用户的位置隐私,To 等[13]采用拉普拉斯机制对位置熵进行噪声处理,注入噪声的规模由全局敏感度决定,为了解决全局敏感度导致噪声过大的问题,采用本地敏感度或平滑敏感度的方法代替全局敏感度.

空间直方图和位置熵的缺陷在于丢弃了用户轨迹的时序信息,无法满足数据使用者对序列进行深入分析的需求,而直接发布轨迹数据可以在最大程度上保留轨迹的时序特征. 目前有基于树重构和轨迹聚类等多种差分隐私轨迹发布方法. 如 Chen 等[14]提出一种基于前缀树的轨迹差分隐私保护方法,通过构造加噪前缀树并为加噪前缀树中每个节点计数添加噪声实现差分隐私. 霍峥等[15]在噪音树的基础上分别针对自由空间和路网空间提出了两种差分隐私轨迹数据发布方法. 基于聚类的轨迹发布方法[16]在位置精度高,候选位置集合规模大的情况下更适用. 这种方法采用分阶段处理的思想,将长度为的轨迹集处理分为个阶段,在每个阶段对所有位置进行聚类分组,使用聚类中心点代替该聚类中的真实位置点,通过在聚类中心点的随机化和轨迹计数的随机化过程中引入随机化噪声实现差分隐私,最后发布扰动后的数据集. 如 Zhao 等[17]提出的差分隐私轨迹聚类算法 TLDP (Trajectory Location Data Protection),通过将 Laplace 噪声添加到轨迹计数中来抵抗连续查询攻击.

2.2　本地化差分隐私研究现状

与传统的差分隐私[4]技术相比,本地化差分隐私技术已经成为一种更为健壮的隐私保护模型,与传统的中心化差分隐私不同,该技术的核心思路是在本地给用户数据添加满足本地化差分隐私的扰动,将扰动后数据传输给第三方数据收集者,再通过一系列查询操作得到有效的结果. 本地化差分隐私的目标在于解决服务器不可信场景下数据的安全采集与分析问题. 本地化差分隐私中常用的扰动机制是随机响应,如杨高明等提出一种满足本地化差分隐私约束的关联属性不变后随机响应扰动方法[18]. 除此以外还有压缩机制 Compression[19]和扭曲机制 Distortion[20]. 这些扰动机制

在频数统计、均值估计和机器学习等学术领域有大量应用[21]. 除了学术研究以外,本地化差分隐私在工业界也有所应用,如苹果公司将该技术应用在操作系统 ios 10 上以隐私保护的方式收集用户的统计数据[22],谷歌公司同样使用该技术从 Chrome 浏览器上采集用户的行为统计数据[23].

3 背景知识

3.1 位置数据建模

本文使用两种坐标系来表示用户的位置点. 一种是状态坐标系,将原始地图划分成多个网格,使得每个网格单元表示用户的一个位置状态. 另一种是地图坐标,通过二维经纬度坐标点来表示用户的位置.

这两种坐标系之间可以互相转化,状态坐标中每个网格单元的索引可以由经纬度表示,从而对应到地图坐标上. 如图 1 所示,横坐标表示向东方向数据,纵坐标表示向北方向数据,网格表示位置状态.

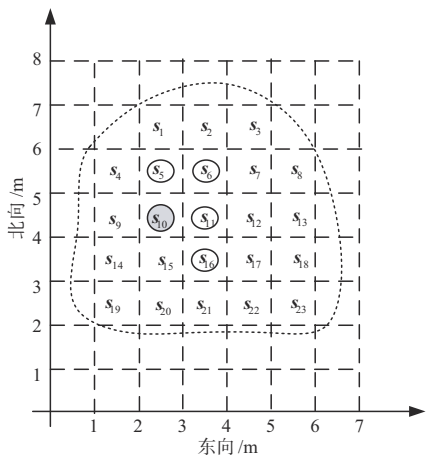


图 1 状态坐标系和地图坐标转化示意图

对于位置域 $S=\{\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_N\}$,记 $\boldsymbol{s}_i(1\le i\le N)$ 代表图上的第 i 个网格单元,该网格单元表示为一个单位向量,其中第 i 个元素为 1,其余 $N-1$ 个元素为 0,用二维向量 $\boldsymbol{x}_t$ 表示 t 时刻用户在地图坐标上的位置,$\boldsymbol{x}_t[0]$ 表示该位置点的经度,$\boldsymbol{x}_t[1]$ 表示该位置点的维度,与 t 时刻用户在状态坐标上的真实位置 $\boldsymbol{s}_t^*$ 互相对应,位置查询 $f(\boldsymbol{s}):\boldsymbol{s}\to\mathbb{R}^2$ 表示网络坐标中位置点到地图坐标的映射,以图 1 为例,设用户位置域 $S=\{\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_{56}\}$,若用户在 t 时刻真实位置状态为 $\boldsymbol{s}_{10}$,则 $\boldsymbol{s}_{10}=[0,0,0,0,0,0,0,0,0,0,1,\cdots,0]$,对应地图坐标 $\boldsymbol{x}_t=[3,5]$,即 $\boldsymbol{s}_{10}$ 的维度坐标为 3,经度坐标为 5,存在位置查询 $f(\boldsymbol{s}_{10})=[3,5]$.

3.2 位置隐私攻击

在位置隐私发布的研究背景下,针对位置的隐私攻击可视为根据发布的扰动位置推测出某时刻用户真实位置的过程,本文采用隐马尔可夫模型对该过程进行建模. 每一时刻用户的真实位置是不可观测的,对应隐马尔可夫模型中的隐藏状态,经过隐私保护处理(如 4.2 中的 CPLP 算法)后发布的位置数据可由攻击者直接观察得到,对应隐马尔可夫模型中的观测状态,记矩阵 $\boldsymbol{M}\in[0,1]^{N\times N}$ 为用户的位置状态转移矩阵,矩阵中的元素 m_{ij} 表示由位置状态 $\boldsymbol{s}_i$ 转移到位置状态 $\boldsymbol{s}_j$ 的概率大小,在后续隐私保护位置扰动算法设计中,假设矩阵 $\boldsymbol{M}$ 可根据用户的历史位置数据训练得到. 在位置攻击模型中,t 时刻用户的位置状态可以通过概率分布 $\boldsymbol{p}_t\in[0,1]^{1\times N}$ 表示,其中 $\boldsymbol{p}_t[i]=\Pr(\boldsymbol{s}_t^*=\boldsymbol{s}_i)=\Pr(\boldsymbol{x}_t^*)$ 代表 t 时刻用户真实位置位于 $\boldsymbol{s}_i$ 的可能性大小,假设 t 时刻用户以相同的概率分布于位置集合 $S=\{\boldsymbol{s}_1,\boldsymbol{s}_3,\boldsymbol{s}_4,\boldsymbol{s}_6\}$ 中,则此时用户的位置概率分布表示为 $\boldsymbol{p}_t=[1/4,0,1/4,1/4,0,1/4,0,0,\ldots,0]$. 再使用 $\boldsymbol{p}_t^-$ 和 $\boldsymbol{p}_t^+$ 分别表示攻击者观察扰动输出 $\boldsymbol{z}_t$ 前后该时刻位置状态的先验概率和后验概率. t 时刻的先验概率可以通过前一时刻 $t-1$ 的后验概率结合状态转移矩阵 $\boldsymbol{M}$ 计算得到,即 $\boldsymbol{p}_t^-=\boldsymbol{p}_{t-1}^+\boldsymbol{M}$,后验概率 $\boldsymbol{p}_t^+$ 可根据式(1)中的贝叶斯公式计算,其中 $\Pr(\boldsymbol{z}_t|\boldsymbol{s}_t^*=\boldsymbol{s}_i)$ 表示隐马尔可夫模型的发射概率,即在给定真实位置概率分布的情况下输出扰动位置为 $\boldsymbol{z}_t$ 的概率.

$$\boldsymbol{p}_t^+[i]=\Pr(\boldsymbol{s}_t^*=\boldsymbol{s}_i|\boldsymbol{z}_t)=\frac{\Pr(\boldsymbol{z}_t|\boldsymbol{s}_t^*=\boldsymbol{s}_i)\boldsymbol{p}_t^-[i]}{\sum_j\Pr(\boldsymbol{z}_t|\boldsymbol{s}_t^*=\boldsymbol{s}_j)\boldsymbol{p}_t^-[j]} \tag{1}$$

假设攻击者掌握的背景知识包括隐马尔可夫模型的状态转移矩阵和初始概率分布,则攻击者可以推测出 t 时刻用户可能出现的位置,表现为该时刻先验概率大于 0 的位置,将这个区域定义为时序关联域 C_t,如定义 1 所示.

定义 1 时序关联域. 时序关联域 C_t 代表 t 时刻用户所有可能出现的位置集合,即 $C_t=\{\boldsymbol{s}_i|\boldsymbol{p}_t^-=\Pr(\boldsymbol{s}_t^*=\boldsymbol{s}_i)>0,\boldsymbol{s}_i\in S\}$.

3.3 本地化差分隐私

本地化差分隐私模型基于严格的数学背景,形式化定义如下所示.

定义 2 ε-本地化差分隐私[21]. 给定 n 个用户,每个用户对应一条记录,对于随机化算法 A,其定义域为 $\mathrm{Dom}(A)$,值域为 $\mathrm{Ran}(A)$,若算法 A 在任意两条记录 t 和 $t'(t,t'\in\mathrm{Dom}(A))$ 上得到相同输出结果 $(o(o\subseteq\mathrm{Ran}(A)))$ 的概率满足 $\Pr(A(t)=o)\le\mathrm{e}^{\varepsilon}\Pr(A(t')=o)$,则称算法 A 满足 ε-本地化差分隐私.

本地化差分隐私技术通过控制任意两条记录输出结果的相似性来确保算法的隐私性,即根据随机化算法 A 的某个输出结果无法推测出输入数据为哪一条记

录,根据3.2节中位置隐私攻击的描述,位置隐私发布的目标就是确保攻击者不能根据已发布的扰动位置推测出某个时刻用户的真实位置,也就是保证时序关联域中任意两个位置不能被攻击者区分出来,基于此本文提出ε-不可区分性的定义,表示时序关联域内的差分隐私.

定义3　ε-不可区分性. 对于时序关联域中相邻的两个位置$\mathbf{s}_i$和$\mathbf{s}_j$,在随机化算法A的作用下,若对任意输出$o\subseteq \mathrm{Ran}(A)$,存在$\Pr(A(\boldsymbol{s}_i)=o)\le \mathrm{e}^{\varepsilon}\Pr(A(\boldsymbol{s}_j)=o)$成立,则称随机化算法$A$满足不可区分性.

定义3中参数ε非负,表示隐私保护的程度,该参数越小隐私保护的程度越高. 由于定义2只是理论模型,而要实现具体的位置差分隐私则需要噪声机制的介入,Laplace机制是实现差分隐私最常用的方法,该机制建立在l_1-敏感度(l_1-norm Sensitivity)的基础上,相关定义如下:

定义4　l_1-敏感度[4]. 对于查询$f(\boldsymbol{s}):\boldsymbol{s}\to\mathbb{R}^2$,$l_1$-敏感度指$f(x_1)-f(x_2)$的最大$l_1$范式值,如式(2),其中$x_1$和$x_2$是相邻数据集中的两个元素.

$$\Delta f=\max\left\| f(x_1)-f(x_2)\right\|_1 \tag{2}$$

定义5　Laplace机制[4]. 对于查询$f(\boldsymbol{s}):\boldsymbol{s}\to\mathbb{R}^2$,查询函数的敏感度为$\Delta f$,如果查询算法$A$满足式(3),则算法$A$具有$\varepsilon$-不可区分性,所添加的噪声符合位置参数为0,尺度参数为$\Delta f/\varepsilon$的拉普拉斯分布. 其中,敏感度Δf表示两个相邻位置查询结果的最大l_1范式值.

$$A=A(\boldsymbol{s})+\mathrm{Lap}\left(\frac{\Delta f}{\varepsilon}\right) \tag{3}$$

位置隐私保护中另一种高效的扰动机制是Xiao等[24]提出的平面各向同性扰动机制(Planar Isotropic Mechanism, PIM),该机制基于计算几何学[25]中凸包(Convex Hull)和各向同性位置(Isotropic Position)的定义构造凸包敏感度,并基于K-机制[25]生成噪声.

定义6　凸包[25](Convex Hull). 对于给定集合$X=\{x_1,x_1,\cdots,x_n\}$,包含X中所有点的凸集称作X的凸包,记作Conv(X),凸包可以用X中所有点的线性组合来构造.

定义7　各向同性位置[25](Isotropic Position). 若凸集$K\subseteq\mathbb{R}^d$满足式(4),则称K位于各向同性位置上,式中L_K表示每个单位向量的各向同性常数.

$$\frac{1}{\mathrm{Vol}(K)}\int_K\left|\langle z,v\rangle\right|^2\mathrm{d}z=L_K^2\mathrm{Vol}(K)^{2/d} \tag{4}$$

定义8　凸包敏感度[24](Sensitivity Hull). 对于位置$\mathbf{s}$和查询$f(\boldsymbol{s}):\boldsymbol{s}\to\mathbb{R}^2$,凸包敏感度$K$是$\Delta f$的凸包,如式(5)所示,$\Delta f$表示时序关联域中两个位置点$x_1$和$x_2$的查询差值.

$$\Delta f=\bigcup_{x_1,x_2\in C}(f(x_1)-f(x_2)),K=\mathrm{Conv}(\Delta f) \tag{5}$$

定义9　K-机制[25](K-norm Mechanism). 对于给定的查询函数$f(\boldsymbol{s}):\boldsymbol{s}\to R^d$以及凸包敏感度$K$,若任意扰动输出$\boldsymbol{z}$的概率分布满足式(6),则称其满足$K$-机制,式中$K=FB_1^n$表示凸包敏感度,$\Gamma()$表示伽马函数,$\|\cdot\|_K$表示凸包敏感度的闵可夫斯基范数.

$$\Pr(\boldsymbol{z})=\frac{1}{\Gamma(d+1)\mathrm{Vol}(K/\varepsilon)}\exp(-\varepsilon\left\|\boldsymbol{z}-f(\boldsymbol{s})\right\|_K) \tag{6}$$

4　本地化差分隐私时序位置发布模型

为了解决位置数据发布时存在的隐私泄露问题,本文设计了一种基于本地化差分隐私的时序位置发布模型,如图2所示,模型主要思想为允许用户在本地进行隐私策略的定制,根据定制的隐私策略对时序关联的位置数据添加噪声后发布,实现位置数据发布时的隐私保护.

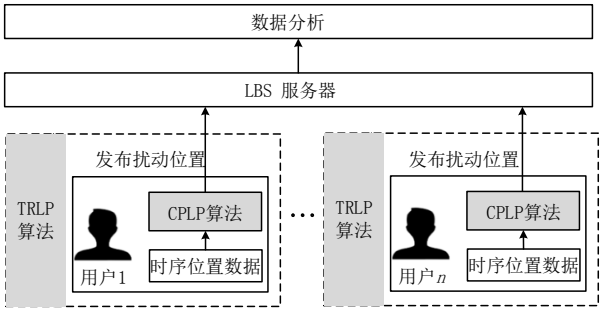


图2　本地化差分隐私时序位置发布模型

在用户端,模型由两个主要算法构成,分别是基于定制隐私策略的位置扰动算法CPLP和基于隐马尔可夫模型的时序关联位置隐私发布算法TRLP. 将经过隐私保护处理后的时序位置进行发布,并上传给LBS服务器,用于后续的位置大数据分析工作.

4.1　定制隐私策略

本文参考Blowfish Privacy[26]来设计位置隐私发布时的定制隐私策略. Blowfish Privacy是一种针对统计数据集的定制隐私保护方案,使用无向图的节点表示需要保护的数据集,边表示对两个数据集提供不可区分性,用户可以在本地通过定制无向图来决定隐私保护程度,然而Blowfish Privacy并不能直接应用在位置数据中,因此结合3.1中定义的位置网格坐标,将定制隐私引入位置数据的隐私保护发布中,提出隐私策略的定义,如定义9所示.

定义10　隐私策略. 隐私策略表示为一个无向图$G=(S,\xi)$,其中S是无向图的节点,代表网格坐标中需要保护的位置状态点,ξ是无向图的边,代表为两个节点

提供ε-不可区分性.

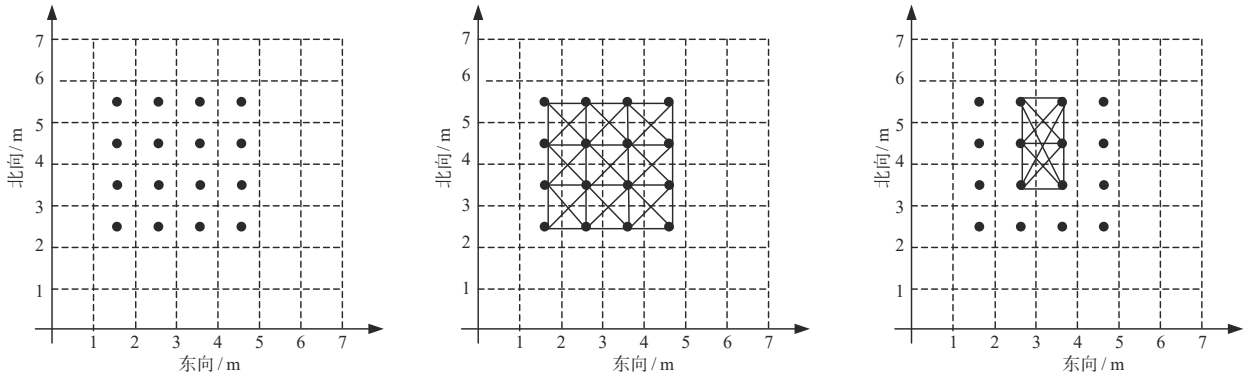
图3展示了几种不同的隐私策略,如图3(a)表示一种宽松的隐私策略,图中所有节点都没有连线,表示可以直接发布用户真实位置,不提供位置隐私保护(仍然需要提供匿名隐私保护),图3(b)的隐私策略为区域内部分位置点之间提供不可区分性,但不要求对图中所有节点提供不可区分性. 与图3(b)相比,图3(c)的隐私要求更为严格,需要保护所选区域内所有位置点之间的隐私性,表现为一个全连接图,这种隐私策略适用于对隐私需求很高的用户.

除了选择图3所示的隐私保护级别外,若用户需要更严格的隐私策略,还可进一步通过定制隐私策略的粒度调整隐私保护级别,粒度代表所保护最小位置范围. 模型为用户提供如图4所示三种粒度的隐私策略,分别是PG_{k9},PG_{k16},PG_{k25},下标中的数字表示隐私策略的粒度,图4中黑色边框表示提供隐私保护的最小位置范围,以PG_{k9}为例,该隐私策略表示网格坐标中每9个网格单元(3×3)内所有位置点彼此完全连接,即在该区域内所有点之间都有连接路径,是一个3×3的全连接图,需要保证该区域内所有位置点不可区分.

为了将定制隐私策略应用到位置差分隐私中,本文结合定义1中的时序关联域C_t,给出时序关联域隐私策略的定义.

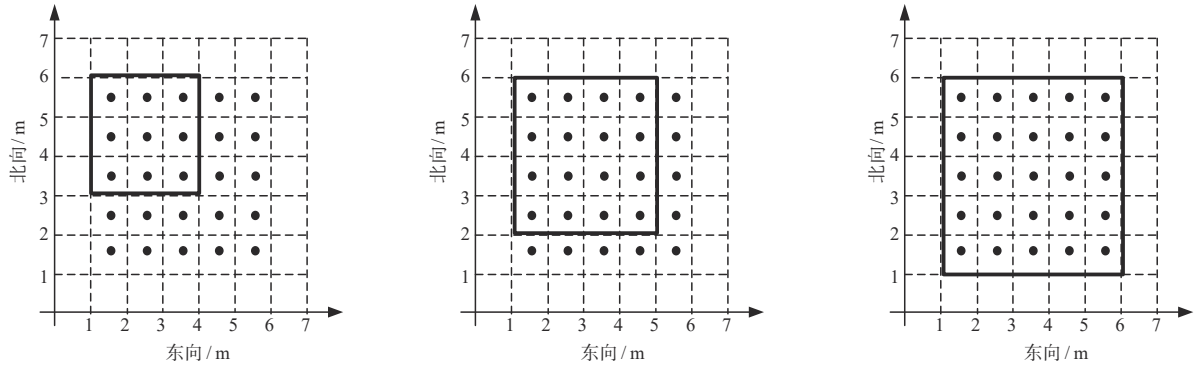
定义11 时序关联域隐私策略G_t^C. t时刻的时序关联域隐私策略G_t^C是隐私策略G在时序关联域中C_t的子图,G_t^C只包含属于时序关联域C_t中的边,即$G^C=(C,\xi^C)$,其中$C\subseteq S$且$\xi^C\subseteq\xi$.

在传统差分隐私定义中,相邻数据集(neighboring databases)被定义为只相差一条记录的两个数据集,在定制隐私策略的背景下,引入相邻节点的概念.



(a) 宽松隐私策略 (b) 普通隐私策略 (c) 严格隐私策略

图3 定制隐私策略示意图



(a) PG_{k9} (b) PG_{k16} (c) PG_{k25}

图4 三种隐私策略粒度示意图

定义12 相邻节点集合$N(\boldsymbol{s})$. 位置$\boldsymbol{s}$的相邻节点是指和$\mathbf{s}$有公共边连接的一系列节点集合,记作$N(\boldsymbol{s})$,则有$N(\boldsymbol{s})=\{\boldsymbol{s}'|d_G(\boldsymbol{s},\boldsymbol{s}')=1,\boldsymbol{s}'\in S\}$,用$d_G(\boldsymbol{s}_i,\boldsymbol{s}_j)$表示隐私策略上点$\boldsymbol{s}_i$和$\boldsymbol{s}_j$之间的距离,该距离可通过两点间最短路径数计算.

结合时序关联域隐私策略,本文提出$\{\varepsilon,G\}$-位置差分隐私的定义,通过确保时序关联域隐私策略中每一对相邻节点的ε-不可区分性,使得攻击者无法区分时序

关联域隐私策略中的相邻位置点.

定义 13　$\{\varepsilon, G\}$-位置差分隐私. 给定一个随机化算法 A,对于时序关联域隐私策略 G_t^C 中所有相邻节点 $\boldsymbol{s}$ 和 $\boldsymbol{s}'$,若对于任意的输出 $z \subseteq \mathrm{Ran}(A)$,存在 $\Pr(A(\boldsymbol{s})=z) \leq \mathrm{e}^{\varepsilon} \Pr(A(\boldsymbol{s}')=z)$ 成立,则认为 $\boldsymbol{s}$ 和 $\boldsymbol{s}'$ 满足 $\{\varepsilon, G\}$-位置差分隐私.

引理 1　对于随机化算法 A,当且仅当时序关联域隐私策略中任意两个节点满足 ε-不可区分性时,算法 A 才满足 $\{\varepsilon, G\}$-位置差分隐私.

4.2　定制隐私策略位置扰动算法

在 4.1 节中,本文根据定制隐私策略对位置差分隐私模型进行了拓展,提出了 $\{\varepsilon, G\}$-位置差分隐私,本节设计一种基于定制隐私策略的位置扰动算法 CPLP,对单一时刻真实位置的查询结果添加噪声,生成扰动位置,在 4.4.1 节证明所提算法满足 $\{\varepsilon, G\}$-位置差分隐私. 对位置数据的扰动可以看作以隐私保护的方式响应查询函数,使得攻击者无法根据扰动后的位置推测出用户的真实位置,具体流程如算法 1 所示.

算法 1　定制隐私策略位置扰动算法(CPLP)

输入:隐私预算 ε,时序关联隐私策略 G_t^C,真实位置 $\boldsymbol{s}$

输出:扰动位置 z

```
根据定义 12 计算相邻位置点集合 N^P(s):
Δf^G = []
FOR i IN range(len(N^P(s)))
    FOR j IN range(i, len(N^P(s)))
        Δf^G.append(f(s_i) − f(s_j))
    END FOR
END FOR
K_I(G_t^C) = Conv(Δf^G)
从 K(G_t^C) 中采样得到 (y_1, y_2, ···, y_l)
根据式(8)计算矩阵 T
K_I(G) = TK(G)
从 K_I(G) 中采样得到 z''
从 Γ(3, ε^{-1}) 中采样得到 r
z'' = rT^{-1}z''
z' ← f(s) + z''
z ← find_nearest_location(z')
RETURN z
```

首先是查询函数敏感度的计算. 传统差分隐私中查询函数的敏感度代表有无某条数据记录对查询结果的最大影响值,在本文定制隐私策略的背景下,查询函数的敏感度代表查询时序关联隐私策略域 G_t^C 中相邻位置节点时查询结果的最大变化值,在定制隐私策略背景下,查询函数的的敏感度计算方法进行对应算法 1 中步骤 1~7,对于 t 时刻的位置状态 $\boldsymbol{s}$,根据时序关联域隐私策略 G_t^C,结合定义 12 计算当前时刻真实位置状态在时序关联域隐私策略中相邻位置点的集合 $N^P(\boldsymbol{s})$,用 Δf^G 表示 $N^P(\boldsymbol{s})$ 中每两个位置查询差值结果的集合,计算公式如式(7)所示.

$$\Delta f^G = \bigcup_{\boldsymbol{s}_i, \boldsymbol{s}_j \in N^P(\boldsymbol{s})} (f(\boldsymbol{s}_i) - f(\boldsymbol{s}_j)) \tag{7}$$

其次将凸包敏感度应用到定制隐私策略位置扰动的背景中. 通过计算 Δf^G 的凸包得到 $K(G_t^C)$,凸包可直观理解为由集合 $X=\{x_1, x_1, \cdots, x_n\}$ 最外沿的所有点连接所组成的凸多边形,$K(G_t^C)$ 表示 t 时刻查询函数的敏感度,记作隐私策略凸包敏感度,表现为一组二维坐标对. 将平面各向同性扰动机制应用到定制隐私策略位置扰动的过程对应算法 1 中步骤 8 ~ 11,对于所得的隐私策略凸包敏感度 $K(G_t^C)$,根据定义 7 将 $K(G_t^C)$ 转化为其各向同性位置 $K_I(G_t^C)$:从集合 $K(G_t^C)$ 中均匀采样得到 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_l$,代入式(8)中计算矩阵 $\boldsymbol{T}$,$K_I(G_t^C)$ 可根据矩阵 $\boldsymbol{T}$ 与 $K(G_t^C)$ 相乘得来,即 $K_I(G_t^C)=\boldsymbol{T}K(G_t^C)$,在二维平面上一个凸包的各向同性位置可直观理解为保持凸包原始方向不变,以凸包的各个顶点为坐标中心对凸包进行旋转排列构成的图形.

$$\boldsymbol{T} = \left(\frac{1}{l} \sum_{i=1}^{l} \boldsymbol{y}_i \boldsymbol{y}_i^{\mathrm{T}} \right)^{-\frac{1}{2}} \tag{8}$$

最后是扰动噪声的生成过程. 对应算法 1 中步骤 12~14,先从 $K_I(G_t^C)$ 中均匀采样得到 $\boldsymbol{z}''$,再从伽马分布 $\Gamma(3, \varepsilon^{-1})$ 中随机产生变量 r,此时得到噪声 $\boldsymbol{z}''=r\boldsymbol{z}''$,将得到的结果转换回 $K(G_t^C)$ 中得 $\boldsymbol{z}''=\boldsymbol{T}^{-1}\boldsymbol{z}''$,这里的 $\boldsymbol{z}''$ 就是添加的噪声大小,表现为一个二维向量. 算法 1 中第 15 步表示对 t 时刻真实位置状态的查询结果添加噪声,得 $\boldsymbol{z}'=f(\boldsymbol{s})+\boldsymbol{z}''$. 算法 1 中第 16 步所用到的函数 find_nearest_location($\boldsymbol{z}'$) 表示在地图坐标系中找到距离 $\boldsymbol{z}'$ 最近的真实位置 $\boldsymbol{z}$ 作为扰动输出返回. 记 t 时刻经过算法 1 处理后所发布的扰动位置为 $\boldsymbol{z}_t$,用 $\Pr(\boldsymbol{z}_t|\boldsymbol{s}_t^*=\boldsymbol{s}_i)$ 表示发布扰动位置 $\boldsymbol{z}_t$ 的概率大小,根据定义 8 中的 K-机制,使用式(9)计算.

$$\Pr(\boldsymbol{z}_t|\boldsymbol{s}_i) = \frac{1}{\Gamma(3)\mathrm{Vol}(K_I/\varepsilon)} \mathrm{e}^{(-\varepsilon \left\| \boldsymbol{z}_t' - \boldsymbol{s}_i' \right\|_{K_I})} \tag{9}$$

$$\boldsymbol{z}_t' = \boldsymbol{T}\boldsymbol{z}_t;\ \boldsymbol{s}_i' = \boldsymbol{T}\boldsymbol{s}_i$$

4.3　时序关联位置隐私发布算法

由于 4.2 节中所提的定制隐私策略位置扰动算法只能用于单一时刻的位置扰动,而在发布连续时刻的位置数据时,需要考虑时序关联的影响,即发布历史时刻的扰动位置对攻击者预测下一时刻真实位置的影响,图 5 展示了时序关联位置隐私发布的过程.

根据 2.2 节中描述的位置隐私攻击模型,攻击者掌握的背景知识包括用户的历史位置数据、用户初始位置的概率分布 $\boldsymbol{p}_1$,在此基础上对攻击者的背景知识做

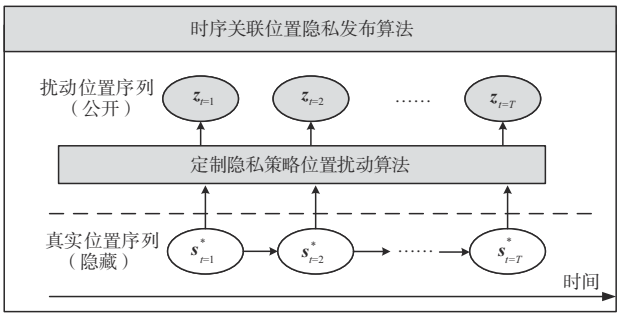


图5 时序关联位置隐私发布示意图

最大假设，假设攻击者的背景知识还包括定制隐私策略位置扰动算法CPLP，在这样的情况下，攻击者先验知识（如时序关联域）的计算可以看作隐马尔可夫模型的推理问题（Inference Problem），即攻击者试图结合定制隐私策略位置扰动算法CPLP、当前时刻的马尔可夫模型和当前时刻之前的所有扰动输出推测出当前时刻的真实位置. 为了抵御拥有强大背景知识的攻击者对用户位置的推测，本节设计了时序关联位置隐私发布算法TRLP，具体流程如算法2所示.

算法2 时序关联位置隐私发布算法(TRLP)

输入:隐私预算ε,时序关联隐私策略G,状态转移矩阵$\boldsymbol{M}$,前一时刻后验概率$\boldsymbol{p}^+_{t-1}$,当前时刻位置$\boldsymbol{s}^*_t$

输出:每一时刻经算法CPLP扰动后的位置

1. $\boldsymbol{p}^-_t=\boldsymbol{p}^+_{t-1}\boldsymbol{M}$
2. $C_t\leftarrow\{\boldsymbol{s}_i|\boldsymbol{p}^-_t[i]>0\}$
3. $G^C_t\leftarrow G\wedge C_t$
4. $\boldsymbol{z}_t\leftarrow \mathrm{CPLP}(\varepsilon,G^C_t,\boldsymbol{s}^*_t)$
5. 根据式(1)计算$\boldsymbol{p}^+_t$
6. $\mathrm{TRLP}(\varepsilon,G^C_t,\boldsymbol{M},\boldsymbol{p}^+_t,\boldsymbol{s}^*_t)$//递归调用本算法

算法第1步计算当前时刻的先验概率，每一时刻的先验概率由前一时刻的后验概率与马尔可夫状态转移矩阵$\boldsymbol{M}$相乘得来；第2步根据先验概率计算t时刻的时序关联域C_t，即攻击者根据历史发布的数据所推测出该时刻用户的所有可能位置集合，第3步将此时的时序关联域和用户的定制隐私策略求交集得到时序关联域隐私策略G^C_t，第4步中将隐私预算、当前时刻的时序关联域隐私策略G^C_t和当前时刻真实位置状态$\boldsymbol{s}^*_t$代入定制隐私策略位置扰动算法CPLP中，对此时的真实位置进行扰动得到$\boldsymbol{z}_t$，第5步中将扰动位置$\boldsymbol{z}_t$代入式(1)计算t时刻的后验概率$\boldsymbol{p}^+_t$，最后在第6步中将时序关联域隐私策略G^C_t、t时刻的后验概率$\boldsymbol{p}^+_t$和$t+1$时刻的真实位置$\boldsymbol{s}^*_{t+1}$代入本算法，即递归调用，实现下一时刻扰动位置的发布. 该算法的输出为每一时刻经算法CPLP扰动后的位置（算法2中第4步）.

4.4 算法分析

本节分别从隐私安全性和时间复杂度两方面对所提算法进行理论分析.

4.4.1 算法隐私安全性分析

首先证明单一时刻定制隐私策略位置扰动算法CPLP满足$\{\varepsilon,G\}$-位置差分隐私，由于算法CPLP基于对攻击者能力的最大假设，因此证明算法CPLP满足$\{\varepsilon,G\}$-位置差分隐私即可保证该时刻所发布位置的隐私性.

定理1 算法CPLP满足$\{\varepsilon,G\}$-位置差分隐私.

证明 取$\forall\boldsymbol{s}_i,\boldsymbol{s}_j\in N^P(\boldsymbol{s})$，对于同样的扰动输出$\boldsymbol{z}$，其概率分布如下：

$$\Pr(\boldsymbol{z}|\boldsymbol{s}_i)=\frac{1}{\Gamma(3)\mathrm{Vol}(K_I/\varepsilon)}\mathrm{e}^{(-\varepsilon\left\|\boldsymbol{z}-\boldsymbol{s}_i\right\|_{K_I})},$$

$$\Pr(\boldsymbol{z}|\boldsymbol{s}_j)=\frac{1}{\Gamma(3)\mathrm{Vol}(K_I/\varepsilon)}\mathrm{e}^{(-\varepsilon\left\|\boldsymbol{z}-\boldsymbol{s}_j\right\|_{K_I})}$$

比较两个概率分布可得：

$$\frac{\Pr(\boldsymbol{z}|\boldsymbol{s}_i)}{\Pr(\boldsymbol{z}|\boldsymbol{s}_j)}=\frac{\frac{1}{\Gamma(3)\mathrm{Vol}(K_I/\varepsilon)}\mathrm{e}^{(-\varepsilon\left\|\boldsymbol{z}-\boldsymbol{s}_i\right\|_{K_I})}}{\frac{1}{\Gamma(3)\mathrm{Vol}(K_I/\varepsilon)}\mathrm{e}^{(-\varepsilon\left\|\boldsymbol{z}-\boldsymbol{s}_j\right\|_{K_I})}}$$

$$=\mathrm{e}^{\varepsilon(\left\|\boldsymbol{z}-\boldsymbol{s}_i\right\|_{K_I}-\left\|\boldsymbol{z}-\boldsymbol{s}_j\right\|_{K_I})}$$

根据三角不等式，有：

$$\mathrm{e}^{\varepsilon(\left\|\boldsymbol{z}-\boldsymbol{s}_i\right\|_{K_I}-\left\|\boldsymbol{z}-\boldsymbol{s}_j\right\|_{K_I})}\le\mathrm{e}^{\varepsilon\left\|\boldsymbol{s}_i-\boldsymbol{s}_j\right\|_{K_I}}$$

$$\because\ \left\|\boldsymbol{s}_i-\boldsymbol{s}_j\right\|_{K_I}\le 1$$

$$\therefore\ \frac{\Pr(\boldsymbol{z}|\boldsymbol{s}_i)}{\Pr(\boldsymbol{z}|\boldsymbol{s}_j)}=\mathrm{e}^{\varepsilon\left\|\boldsymbol{s}_i-\boldsymbol{s}_j\right\|_{K_I}}\le\mathrm{e}^{\varepsilon}$$

因此可知对于时序关联域隐私策略中任意两个相邻的位置点，算法CPLP满足ε-不可区分性，根据引理1可知算法CPLP满足$\{\varepsilon,G\}$-位置差分隐私.

证毕.

其次，分析连续时刻位置隐私发布算法TRLP的隐私安全性. 根据差分隐私的序列组合性[4]，记$A_1,\cdots,A_T$为T个独立的随机化算法，分别表示每一时刻对真实位置的扰动处理，若分别$A_1,\cdots,A_T$满足$\{\varepsilon,G\}$-位置差分隐私，则其组合$\{A_1,\cdots,A_T\}$满足$\{T_\varepsilon,G\}$-位置差分隐私，其中T_ε表示所有时刻隐私预算的总和，即连续时刻的位置发布算法TRLP同样满足$\{\varepsilon,G\}$-位置差分隐私，因此根据连续时刻位置发布算法所得到的轨迹数据具有一定的隐私安全性.

4.4.2 算法时间复杂度分析

关于算法的时间复杂度，TRLP算法最耗时的地方在于计算每个时刻扰动输出的位置点$\boldsymbol{z}$，即CPLP算法的输出，而在CPLP算法中，根据每个时刻的时序关联域G^C_t计算凸包敏感度耗时最大，记凸包的顶点个数为n，时序关联域大小为h，则算法的时间复杂度表示为$O(h\log(n)+n^2\log(h))$.

5　实验与分析

5.1　实验数据与环境

本文所使用的实验平台操作系统为 Windows 10+64 位,开发环境为 Pycharm,编程语言为 Python 3.8,CPU 为 Intel(R) Core(TM) i5-7300HQ,内存为 8 GB. 实验采用两个数据集,分别是微软亚洲研究院的 GeoLife 数据集[27]和斯坦福大学复杂网络分析平台公开的真实数据集 Gowalla 数据集[28]. GeoLife 数据集记录了从 2007 年 4 月到 2012 年 8 月 182 个用户的轨迹数据,包含一系列以时间为序的包含经纬度、海拔等信息位置点信息,共计包含 17621 条轨迹. 本文提取其中位于北京四环内的轨迹数据,将地图分割成 $340\times340\ m^2$ 的网格单元,用于马尔可夫状态转移矩阵 $\boldsymbol{M}$ 的训练. Gowalla 数据集中包含 196586 名用户 20 个月内在 6442890 个位置上签到的数据,本文提取该数据集中所有位于洛杉矶的位置数据,将地图分割成 $370\times370\ m^2$ 的网格单元用于马尔可夫状态转移矩阵 $\boldsymbol{M}$ 的训练.

5.2　算法可用性度量指标

为了评估本文所提位置隐私发布算法 TRLP 的可用性,在实验中使用两个度量指标. 第一个度量指标是原始位置和扰动位置之间的欧几里得距离 E_{eu},即原始位置和扰动位置之间的误差,单位为 m,该指标是位置隐私发布算法中通用一个的度量指标,E_{eu} 值越小,即误差越小,说明位置隐私发布算法的可用性越高. 第二个度量指标是在位置数据集上运行 k 近邻查询的精度,分别在原始位置数据集和发布的扰动位置数据集上运行 k 近邻查询,假设在原始位置数据上运行近邻查询的结果为 R,在扰动位置上运行 k 近邻查询结果为 R',则 k 近邻查询的精度 P 计算方式如下:

$$P=\left|R\cap R'\right|/k$$

k 近邻查询是位置数据发布后常用的一种数据分析方法,其目的是查找距离用户最近的 k 个兴趣点,对于不同的位置扰动算法,k 近邻查询的精度越高,说明经过位置隐私发布算法扰动后所得到的轨迹质量越高.

5.3　算法可用性实验

本节分别从两个方面探究所提时序关联位置隐私发布算法 TRLP 的可用性,一方面探究定制隐私策略的粒度和隐私预算对算法可用性的影响,另一方面探究发布时序位置数据时算法可用性的变化情况.

首先探究定制隐私策略的粒度和隐私预算 ε 对算法 TRLP 可用性的影响. 在 Geolife 数据集上选择 20 个用户 150 个时刻下的位置进行隐私发布,构造不同粒度的隐私策略(PG_{k9},PG_{k16},PG_{k25}),在这三种隐私策略上分别运行 TRLP 算法 30 次,以原始位置和扰动位置之间的欧几里得距离 E_{eu} 作为算法可用性的度量标准,在 ε 分别取值 0.3,0.5,0.7,1 时返回 E_{eu} 的平均值,实验结果如图 6 所示.

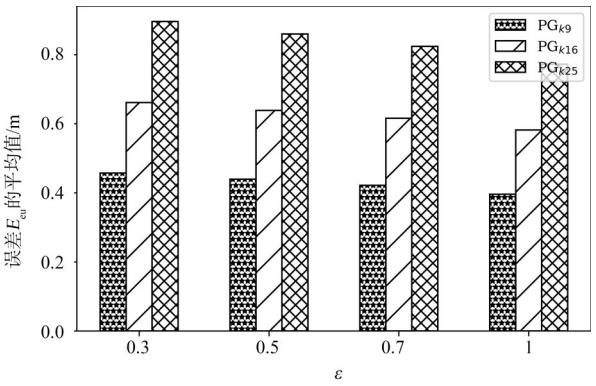


图 6　定制隐私策略粒度对算法可用性影响

实验结果分析如下:

(1) 随着隐私预算的增加,算法 TRLP 的误差逐渐减小,即可用性越高,说明算法 TRLP 的可用性随着隐私预算的增加而增强.

(2) 在相同隐私预算的情况下,隐私策略的粒度越小,则算法 TRLP 的误差越小,即算法 TRLP 的可用性越高,说明可以通过调整隐私策略的粒度来调整算法 TRLP 中隐私保护程度与可用性之间的平衡.

(3) 在不同隐私预算以及不同粒度的隐私策略下,算法 TRLP 的误差范围均在 1 m 内,说明算法 TRLP 具有较高的可用性.

其次探究算法 TRLP 发布连续位置数据时可用性的变化情况. 分别在 Geolife 和 Gowalla 两个数据集上选择 20 个用户 100 个时刻的位置数据,设置隐私预算 $\varepsilon=0.5$,基于三种粒度的隐私策略运行时序关联位置隐私发布算法 TRLP,结果如图 7 所示.

实验结果分析如下:

(1) 在 Geolife 和 Gowalla 两个真实的位置数据集上运行时序关联位置隐私发布算法 TRLP 时,随着隐私策略粒度的减小,算法 TRLP 的误差也随之减小,即可用性逐渐增加. 同样说明算法 TRLP 可以通过调整隐私策略的粒度来调整隐私保护程度与可用性之间的平衡.

(2) 在 Geolife 和 Gowalla 两个真实的位置数据集上运行时序关联位置隐私发布算法 TRLP 时,在三种不同粒度的隐私策略下,连续时刻之间算法 TRLP 的误差值变化幅度较小,且所有时刻误差值均在 1 m 内,同样说明算法 TRLP 可用性较高.

(3) 在三种不同粒度的隐私策略下,算法 TRLP 在 Gowalla 数据集上的误差 E_{eu} 均小于 Geolife 数据集上的误差 E_{eu},原因是与 Geolife 数据集相比,Gowalla 数据集

收集的用户数据具有明显的移动模式，所以训练的马尔可夫状态转移矩阵$\boldsymbol{M}$的准确性更高，因此算法TRLP在Gowalla数据集上运行的可用性更高.

5.4 对比实验

本节以文献[24]中的拉普拉斯方法作为基线算法，采用不同的度量指标对算法TRLP与基线算法的可用性进行比较. 两种算法主要的差别是计算敏感度的方式以及噪声添加的方式不同，算法TRLP通过定制隐私策略的方法计算敏感度并添加满足噪声，基线算法仅通过计算l_1-敏感度给位置数据添加噪声.

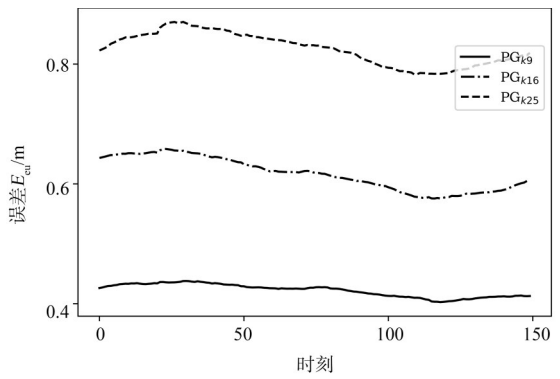
首先以原始位置和扰动后位置之间的欧几里得距离E_{eu}作为度量指标，在Geolife数据集上选择20个用户150个时刻下的位置，基于隐私策略分别运行算法TRLP和基线算法，迭代次数为30次，ε分别取值0.3，0.5，0.7，1时，计算E_{eu}的平均值，实验结果如图8(*a*)所示.

其次以k近邻查询的精度P作为度量指标，在Geolife数据集上选择20个用户150个时刻下的位置，设置隐私预算$\varepsilon=0.5$，基于隐私策略PG_{k9}分别运行算法TRLP和基线算法，迭代次数为30次，在扰动前后的位置数据集上运行k近邻查询，根据5.2中的描述计算k近邻查询的精度P. 在$k\in[50,75,100,125,150]$的情况下运行k近邻查询时算法TRLP和基线算法的精度如图8(*b*)所示.

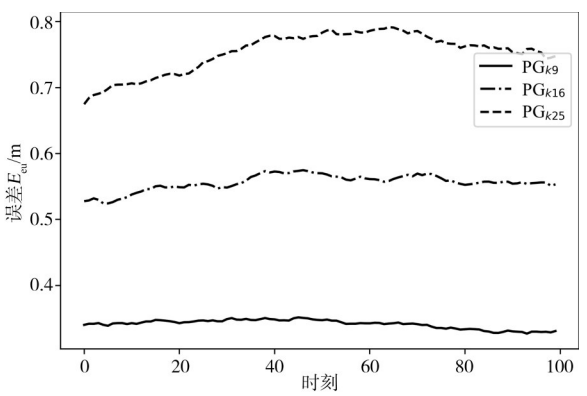
最后比较算法TRLP与基线算法的时效性，同样在Geolife数据集上选择20个用户150个时刻下的位置，基于隐私策略分别运行TRLP算法和基线算法，迭代次数为30次，返回不同隐私预算下两种算法针对单个时刻运行所需时间的平均值，ε分别取值0.3，0.5，0.7，1时，实验结果如图8(*c*)所示.

实验结果分析如下：

（1）随着隐私预算的增加，算法TRLP和基线算法的误差均逐渐减小，说明算法TRLP和基线算法的可用性均随着隐私预算的增加而增加，然而在同样的隐私预算下，算法TRLP的误差E_{eu}小于基线算法，



(*a*) Geolife数据集

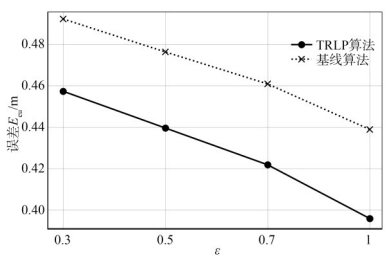


(*b*) Gowalla数据集

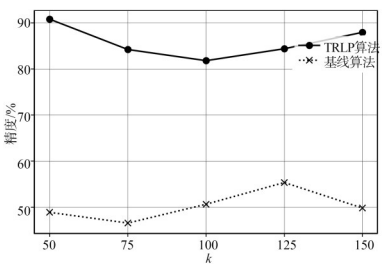
图7 时序关联位置隐私发布

即算法的可用更高，说明算法TRLP计算敏感度和添加噪声的方式与基线算法相比冗余更少，即采用定制隐私策略进行敏感度的计算效果更好，因此算法TRLP可以在满足位置差分隐私的同时保证更好的可用性.

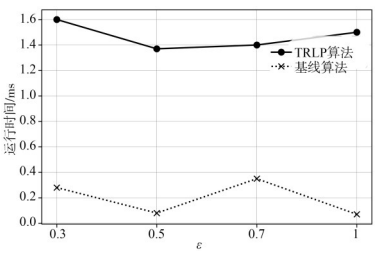
（2）基于不同的k值对两种位置隐私发布算法扰动后的轨迹数据进行k近邻查询时，算法TRLP的精度P均高于基线算法，因此将经过算法TRLP扰动后发布的轨迹数据用于k近邻查询时效果更好，说明经过算法TRLP扰动后发布的轨迹数据质量高于基线算法扰动后所发布的轨迹数据质量.



(a) 误差E_{eu}对比



(*b*) 精度P对比



(c) 运行时间对比

图8 TRLP算法与基线算法对比

（3）在不同的隐私预算下,算法TRLP和基线算法针对单个时刻位置进行隐私发布所需要的时间都在2毫秒内,说明两种算法都具有一定的实时性,但与基线算法相比,算法TRLP的运行时间较长,因为算法TRLP计算敏感度时需要根据每个时刻的时序关联域计算凸包敏感度,这一步耗时较大.

6 结束语

本文提出了一种基于本地化差分隐私的时序位置发布模型,模型采用了灵活的位置隐私保护方案,即由用户选择系统已设定的多种隐私策略或定制隐私策略,在此基础上设计了定制隐私策略位置扰动算法(CPLP),同时提出一种基于定制隐私策略的时序位置发布算法(TRLP),通过保证用户发布位置的不可区分性从而保证用户的位置隐私. 在两个真实的位置数据集上进行实验,验证了与基线相比,算法TRLP具有较好的可用性. 今后的研究将考虑如下两个方面:(1)与基线算法相比,算法TRLP的运行时间较长,因此后续工作中需要研究如何在保证算法TRLP可用性的前提下提高其运行速度,从而将算法TRLP扩展到实时位置服务中;(2)在设计定制隐私策略时没有考虑用户不同的移动模式(如交通方式),因此在后续工作中可以将用户的移动模式引入定制隐私策略的设计,根据用户不同移动模式为用户提供更细粒度的隐私保护选择.

参考文献

[1] BAO J, HE T F, RUAN S J, et al. Planning bike lanes based on sharing-bikes' trajectories[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2017: 1377-1386.

[2] CAO Y, XIAO Y H, TAKAGI S, et al. PGLP: Customizable and rigorous location privacy through policy graph[C]// European Symposium on Research in Computer Security. Cham: Springer, 2020: 655-676.

[3] GEDIK B, LIU L. Protecting location privacy with personalized k-anonymity: Architecture and algorithms[J]. IEEE Transactions on Mobile Computing, 2008, 7(1): 1-18.

[4] DWORK C, KENTHAPADI K, MCSHERRY F, et al. Our data, ourselves: Privacy via distributed noise generation [C]//Annual International Conference on the Theory and Applications of Cryptographic Techniques. Heidelberg, Berlin: Springer, 2006: 486-503.

[5] 康海燕, 朱万祥. 位置服务隐私保护[J]. 山东大学学报(理学版), 2018, 53(11): 35-50.
KANG Hai-yan, ZHU Wan-xiang. Privacy preservation for location-based services[J]. Journal of Shandong University (Natural Science), 2018, 53(11):35-50. (in Chinese)

[6] 冯登国, 张敏, 叶宇桐. 基于差分隐私模型的位置轨迹发布技术研究[J]. 电子与信息学报, 2020, 42(1): 74-88.
FENG D G, ZHANG M, YE Y T. Research on differentially private trajectory data publishing[J]. Journal of Electronics & Information Technology, 2020, 42(1): 74-88. (in Chinese)

[7] 郑孝遥, 罗永龙, 汪祥舜, 等. 基于位置服务的分布式差分隐私推荐方法研究[J]. 电子学报, 2021, 49(1): 99-110.
ZHENG X Y, LUO Y L, WANG X S, et al. Research on location-based distributed differential privacy recommendation method[J]. Acta Electronica Sinica, 2021, 49(1): 99-110. (in Chinese)

[8] TAKAGI S, CAO Y, ASANO Y, et al. Geo-graph-indistinguishability: Protecting location privacy for LBS over road networks[C]//IFIP Annual Conference on Data and Applications Security and Privacy. Cham: Springer, 2019: 143-163.

[9] CORMODE G, PROCOPIUC C, SRIVASTAVA D, et al. Differentially private spatial decompositions[C]//2012 IEEE 28th International Conference on Data Engineering. Piscataway: IEEE, 2012: 20-31.

[10] HAY M, LI C, MIKLAU G, et al. Accurate estimation of the degree distribution of private networks[C]//2009 Ninth IEEE International Conference on Data Mining. Piscataway: IEEE, 2009: 169-178.

[11] ZHANG J, XIAO X K, XIE X. PrivTree: A differentially private algorithm for hierarchical decompositions[C]//Proceedings of the 2016 International Conference on Management of Data. New York: ACM, 2016: 155-170.

[12] XIE H R, TANIN E, KULIK L, et al. Euler histogram tree: A spatial data structure for aggregate range queries on vehicle trajectories[C]//Proceedings of the 7th ACM SIGSPATIAL International Workshop on Computational Transportation Science. New York: ACM, 2014: 18-24.

[13] TO H, NGUYEN K, SHAHABI C. Differentially private publication of location entropy[C]//Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM, 2016: 1-10.

[14] 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, 2012: 213-

221.
[15] 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. (in Chinese)
[16] HUA J Y, GAO Y, ZHONG S. Differentially private publication of general time-serial trajectory data[C]//2015 IEEE Conference on Computer Communications. Piscataway: IEEE, 2015: 549-557.
[17] ZHAO X D, PI D C, CHEN J F. Novel trajectory privacy-preserving method based on prefix tree using differential privacy[J]. Knowledge-Based Systems, 2020, 198: 105940.
[18] 杨高明, 朱海明, 方贤进, 等. 局部差分隐私约束的关联属性不变后随机响应扰动[J]. 电子学报, 2019, 47(5): 1079-1085.
YANG G M, ZHU H M, FANG X J, et al. Invariant post-random response perturbation for correlated attributes under local differential privacy constraint[J]. Acta Electronica Sinica, 2019, 47(5): 1079-1085. (in Chinese)
[19] XIONG S J, SARWATE A D, MANDAYAM N B. Randomized requantization with local differential privacy[C]//2016 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2016: 2189-2193.
[20] SARWATE A D, SANKAR L. A rate-disortion perspective on local differential privacy[C]//2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton). Piscataway: IEEE, 2014: 903-908.
[21] WANG T, ZHANG X F, FENG J Y, et al. A comprehensive survey on local differential privacy toward data statistics and analysis[J]. Sensors (Basel, Switzerland), 2020, 20 (24): 7030.
[22] GARFINKEL S L, ABOWD J M, POWAZEK S. Issues encountered deploying differential privacy[C]//Proceedings of the 2018 Workshop on Privacy in the Electronic Society. New York: ACM, 2018: 133-137.
[23] CORMODE G, JHA S, KULKARNI T, et al. Privacy at scale: Local differential privacy in practice[C]//Proceedings of the 2018 International Conference on Management of Data. New York: ACM, 2018: 1655-1658.
[24] 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, 2015: 1298-1309.
[25] HARDT M, TALWAR K. On the geometry of differential privacy[C]//Proceedings of the Forty-second ACM Symposium on Theory of Computing. New York: ACM, 2010: 705-714.
[26] HE X, MACHANAVAJJHALA A, DING B L. Blowfish privacy: Tuning privacy-utility trade-offs using policies [C]//Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data. New York: ACM,2014: 1447-1458.
[27] ZHENG Y, XIE X, MA W. GeoLife: A collaborative social networking service among user, location and trajectory.[J]. IEEE Data Eng Bull, 2010, 33(2): 32-39.
[28] 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, 2011: 1082-1090.

作者简介

康海燕 男,1971年7月生,河北石家庄人,博士,教授,北京信息科技大学信息管理学院副院长,研究方向为网络安全与隐私保护等.
E-mail: kanghaiyan@126.com

冀源蕊 女,1997年11月生,宁夏银川人,北京信息科技大学网络空间安全专业在读硕士研究生,研究方向为网络安全与隐私保护.