计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(12):33−39 [doi: 10.15888/j.cnki.csa.006653]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于相似路径的位置隐私保护方法①

解　瑾, 孙小婷

(中国石油大学 (华东) 计算机与通信工程学院, 青岛 266580)
通讯作者: 解　瑾, E-mail: xiejin_@163.com

摘　要: 目前, 基于位置隐私的保护技术大多针对用户进行单次 LBS 请求进行设计, 只考虑保护当前真实用户所在位置, 而忽略了真实用户连续多次查询时存在的协作用户交叠导致真实用户位置泄露的情况, 进而攻击者可根据真实用户位置点进行轨迹预测, 最终获取真实用户运动轨迹, 导致真实用户位置隐私的泄露. 本文针对上述情况, 在用户发起连续 LBS 请求时, 提出了基于相似路径的位置隐私保护方法 (LPBSP), 首先通过网格结构中历史用户密度进行一定均衡处理, 使之符合真实的环境条件; 然后对前后相邻时刻构造的相似路径进行轨迹偏移度、速度相似度等进行一定条件约束, 使其更加贴近真实用户, 从而混淆攻击者, 达到位置隐私保护的目的, 最后本文通过实验对比验证了本文在匿名成功率、执行时间及位置隐私保护度方面的可行性.

关键词: 位置隐私; 相似路径; 轨迹偏移度; 速度相似度; 隐私保护

引用格式: 解瑾,孙小婷.基于相似路径的位置隐私保护方法.计算机系统应用,2018,27(12):33–39. http://www.c-s-a.org.cn/1003-3254/6653.html

Location Privacy Protection Method Based on Similar Path

XIE Jin, SUN Xiao-Ting

(College of Computer & Communication Engineering, China University of Petroleum, Qingdao 266580, China)

Abstract: At present, most of the protection technology based on location privacy is designed for the user to carry out a single LBS request, it only protects the location of the current real user, but ignores the situation where the real user location is leaked by the cooperative user's overlapping when the real user is repeatedly queried. In this scenario, location prediction based on the real user position is used by the attacker to track the real user trajectories, resulting in the leakage of real user location privacy. In this study, a Location Privacy Protection Method Based on Similar Path (LPBSP) is proposed when the user initiates a continuous LBS request. Firstly, a certain equilibrium process is carried out through the history user density in the grid structure to make it conform to the real environment conditions, and then the similar path constructed in the adjacent time is carried out. The trajectory offset and speed similarity are constrained to make it closer to the real users, so as to confuse the attackers and achieve the purpose of location privacy protection. Finally, the feasibility of anonymous success rate, execution time, and location privacy protection is verified by simulation experiments.

Key words: location privacy; similarity path; trajectory offset; speed similarity; privacy protection

近年来, 随着科技的不断发展, 基于位置的服务 (Location Based Service, LBS) 作为移动互联网应用得到越来越广泛的传播, 为人们生活提供了便利. 但是由于 LBS 服务器是通过获取用户位置向用户发布相应的查询信息结果, 这就导致了用户在享受位置服务的同时, 也更容易遭受个人位置隐私信息泄露的风险.

① 收稿时间: 2018-04-15; 修改时间: 2018-05-08; 采用时间: 2018-05-18; csa 在线出版时间: 2018-12-03

© 中国科学院软件研究所 http://www.c-s-a.org.cn

现阶段位置隐私保护大致分为两部分: 基于快照 LBS 的位置隐私保护和基于连续 LBS 的位置隐私保护. 快照 LBS 是指用户向位置服务提供商发出单次 LBS 请求, 获取相应查询结果. 连续 LBS[1]是指用户按照一定频率将自己的位置信息周期性的发送给 LBS 服务器, LBS 服务器通过用户周期性的位置信息和搜索内容, 实时将最新的结果返回给用户. 然而, 在连续查询过程中某些可以关联推断出的背景因素结合特定场景可能会给用户带来隐私威胁. 如图 1, 用户 A 在连续发送三次 LBS 请求时, 分别处于三个不同匿名集{A,B,C,D}{A,E,F,G}{A,H,I,J}, 攻击者以根据用户的轨迹关联重构用户 A 的过程轨迹. 针对此类问题, 文献[2]首次提出了 KAA 匿名算法, 保证匿名用户在 n 次查询之后还保持在初始匿名集中. 这也就要求在构建连续查询匿名集时, 需要尽可能的将用户附近运动趋势相同的匿名用户加入到匿名集中, 以避免推断攻击.

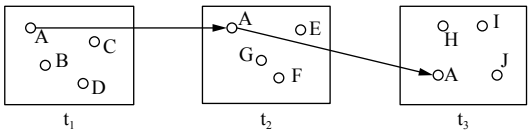


图 1 连续 LBS 请求轨迹过程

目前, 现有连续 LBS 请求中随机生成的虚假位置大多忽略攻击者已知的地理背景信息[3]和真实用户的运动模式. 即使用户采用了一定的匿名方法保护位置隐私, 攻击者还是可以通过相应背景知识攻击用户位置隐私. 针对现有连续 LBS 请求形成虚假轨迹可信性过低造成的位置隐私泄露问题, 提出了一种更加真实的相似路径形成算法, 主要工作如下:

(1) 利用网格单元对历史用户请求密度进行划分, 通过每次采样时刻时真实用户周围的历史用户请求密度计算相对请求概率, 使匿名组的熵达到最大, 从而切断请求与位置间的联系.

(2) 通过真实用户运动轨迹的方向、速度对相似轨迹集合建立约束, 关注用户移动位置间的时空相关性, 使其更加贴近真实用户运动轨迹以达到位置隐私保护目的.

本文第 1 节讲述相关工作, 第 2 节讲述本文研究的预备知识, 第 3 节讲述 LPBSP 位置隐私保护方法, 第 4 节分析方法的安全性, 第 5 节讲述本文实验结果及分析, 最后第 6 节总结.

1 相关工作

1.1 抑制法

基于轨迹数据抑制法的隐私保护技术是指在真实用户轨迹中抑制某些用户的敏感信息, 使攻击者无法将其与用户相关联, 从而达到隐私保护的目的. Gruteser 等[4]将地图上的地理区域划分成敏感区域和非敏感区域, 通过延迟或抑制敏感区域中真实用户的位置更新信息来保护位置隐私. 文献[5]通过 MASKIT 系统通过过滤保护用户位置隐私的上下文信息流进行隐私检查, 从而限制攻击者获得用户的敏感位置信息. 在匿名过程中只考虑状态的抑制, 隐藏会导致隐私泄露的应用程序. 然而, 单纯的抑制法通过抑制敏感位置或访问频次高的数据进行位置隐私保护, 实现简单但是信息丢失率较大.

1.2 历史数据泛化法

历史数据泛化法会引入可信的第三方, 将轨迹信息中的采样点泛化成对应的匿名区域, 以达到隐私保护的目的. 泛化技术可分为空间泛化技技术和时间泛化技术[6]. 空间泛化是指在空间区域范围内降低真实用户的位置精度, 从而增加位置区域上的不确定性. Wang 等[7]为确保用户在每次提交 LBS 请求时构造的匿名区中包含的匿名用户是相同的, 提出了一种基于贪心算法的匿名区构造方法, 但这样会使匿名区面积随着查询次数的增加而增加, 造成严重的通信和计算开销. Pan X 等[8]提出了一种 ICliqueCloak 的隐身算法, 在连续状态中, 采用位置 k-匿名性和隐身粒度作为隐私度量, 划分区域候选集合, 当用户发送请求时, 可以快速识别并生成相应隐藏区域. 2013 年 Latha K[9]在 IClique Cloak 的基础上考虑与用户位置相关的处理延迟和匿名化成本, 提出 KRUPTO 算法, 增加匿名区域覆盖形成最大团.

时间泛化是指在增加时间范围内用户精确位置的不确定性. Hwang RH[10]提出 r-匿名机制用来模糊用户真实轨迹, 将 k-匿名与 s-路段合并, 引入时间混淆技术打破用户进行 LBS 请求提出的时间序列, 运用模糊过程的随机性提供轨迹隐私保护. Palanisamy[11]在路网结构中使用 mix-zone 模型, 使 k 个用户在混合区等待时间相等, 并打乱用户进出混合区的关联顺序, 以确保隐私保证. 历史数据泛化法对实时性要求很高.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

1.3 假数据法

假数据法多指利用 k-匿名技术[12,13]向真实用户运动轨迹中添加假数据, 通过生成 k–1 条假轨迹, 将其一起发送给 LBS 服务器, 以达到匿名效果. 假数据的添加可充分考虑现实环境约束, 由用户定义虚假数据产生方式, 从而提出更加适用于用户的特定需求位置隐私保护算法. 其中, Kido[12]等针对用户在地理环境的分布情况利用普遍性, 拥挤度和分布均匀性三方面设计 MN 和 MLN 算法, 首次提出添加假位置达到隐私保护的目的. Gao 等[14]提出了用于参与式感知的轨迹隐私保护框架, 从图论的角度出发, 提出考虑时间因素的理论混合区模型. Dong 等[15]考虑用户个性化隐私需求, 通过短期位置暴露概率、长期轨迹暴露概率、轨迹偏移距离、轨迹局部相似度和服务请求概率等五个参数让用户进行自定义生成虚假轨迹, 但此方法没有考虑攻击者背景信息. Ye AY 等[16]通过可信的第三方将伪查询插入真实查询中, 防止从用户标识到查询内容的反向映射. 值得注意的是, 在添加假轨迹时要保证添加的假的干扰数据不能对真实轨迹产生严重后果.

本文提出的相似轨迹 LPBSP 方法可以更好地符合真实用户轨迹模式, 使相似轨迹在地图上分布更加合理, 在一定程度上避免了推理攻击和最大移动边界攻击.

2 预备知识

2.1 系统结构

本文提出的基于相似路径的位置隐私保护方法 (LPBSP) 系统采用中心式服务器结构, 包括移动终端、可信第三方匿名服务器 (TTP, Trusted Third Party) 以及 LBS 服务器三部分组成, 系统架构如图 2 所示. 其优点在于具有用户全局信息, 隐私保护程度高、移动终端和 TTP 间的通信和计算开销较少.

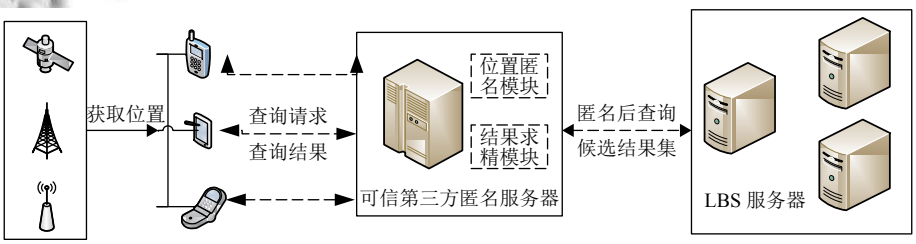


图 2 中心服务器系统结构

移动终端中的 GPS 模块获取移动用户的自身位置坐标, 将查询请求发送给 TTP; 经过 TTP 中的位置匿名模块做匿名处理生成相似轨迹, 并将查询请求发送给 LBS 服务器; LBS 服务器根据查询后选结果将候选集发送给 TTP, 通过 TTP 的结果求精模块过滤精确结果并把结果返回给用户.

2.2 相关定义

本文采用基时空网格划分方法, 将城市地理区域预划分为网格区域, 每个网格都有自己的唯一标识 gidgid.

定义 1(保护的用户属性). $P=\{ID,l_i,v_i,t_i,POI\}$, ID为用户请求 LBS 服务的唯一标识符, l_i 为用户所在的位置信息, v_i 为用户历史速度向量集合, t_i为采样时间信息, POI为用户所请求的兴趣点.

定义 2(k-匿名组). $G\{u\}=\{k,m\}$, k表示用户需要的匿名轨迹数量, k 值隐私保护效果越好. m表示当前匿名组内的用户数量.

定义 3(整体轨迹方向偏移度). 表示采样时刻t_i真实轨迹和相似轨迹的偏移情况. 假设用户真实轨迹为 $T_r=\{ID_i,(l_0,t_0),(l_1,t_1),\cdots,(l_n,t_n)\}$, t_i 表示采样时刻, $l_i=(x_i,y_i)$表示用户在t_i时刻所处位置. 假设真实用户在t_i时刻的位置为$l_i=(x_i,y_i)$,则相对上一位置$l_{i-1}=(x_{i-1},y_{i-1})$轨迹偏移向量为$\overrightarrow{p_i}$, $\overrightarrow{p_i}=(x_i-x_{i-1},y_i-y_{i-1})$. 相似的, 相似轨迹$FT_j=\left\{ID_j^i,\left(l_0^j,t_0\right),\left(l_1^j,t_1\right),\cdots,\left(l_n^j,t_n\right)\right\}$在$t_i$时刻的轨迹偏移向量为$\overrightarrow{p_i^j}=\left(x_i^j-x_{i-1}^j,y_i^j-y_{i-1}^j\right),(1\leqslant j\leqslant k-1)$. 则在$t_i$时刻, 真实轨迹$T_r$和相似轨迹$FT_j$的余弦相似度可表示为:

$$\cos\Delta\theta_i=\frac{dot\left(\overrightarrow{p_i},\overrightarrow{p_i^j}\right)}{|\overrightarrow{p_i}|\cdot|\overrightarrow{p_i^j}|}=\frac{(x_i-x_{i-1})(y_i-y_{i-1})+\left(x_i^j-x_{i-1}^j\right)\left(y_i^j-y_{i-1}^j\right)}{\sqrt{(x_i-x_{i-1})^2+(y_i-y_{i-1})^2}\sqrt{\left(x_i^j-x_{i-1}^j\right)^2+\left(y_i^j-y_{i-1}^j\right)^2}}$$

© 中国科学院软件研究所 http://www.c-s-a.org.cn

所以

$$\Delta\theta_i = \cos^{-1}\frac{(x_i - x_{i-1})(y_i - y_{i-1}) + \left(x_i^j - x_{i-1}^j\right)\left(y_i^j - y_{i-1}^j\right)}{\sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}\sqrt{\left(x_i^j - x_{i-1}^j\right)^2 + \left(y_i^j - y_{i-1}^j\right)^2}}$$

在整个采样阶段相似轨迹FT_j相对真实轨迹 T_r 的偏移角度序列为$P = \{\Delta\theta_1, \Delta\theta_2, \cdots, \Delta\theta_n\}$. 那么, 真实轨迹$T_i$和相似轨迹$FT_j$的轨迹方向偏移度$d \cdot sim$为:

$$d \cdot sim = \frac{1}{n}\sum_{i=1}^{n}\frac{|\Delta\theta_i|}{2\pi} \tag{1}$$

显而易见, $d \cdot sim \in [0,1]$, 越趋近于 0, 设置的相似轨迹整体相对真实轨迹偏移度越小, 整体走向越相似, 从而更难被区分.

定义 4(局部速度相似度). 受真实环境影响, 用户移动速度不是一成不变的. 假设真实用户移动速度序列为$V = \{v_1, v_2, \cdots, v_n\}$在$t_i$时刻的速度为$v_i$, 数值大小为$|v_i|$, 则相应的相似轨迹$FT_j$在$t_i$时刻的速度数值大小为$|v_i^j|$. 那么, 在$t_i$时刻轨迹间的速度相似性$v \cdot sim$为:

$$v \cdot sim = \frac{1}{1 + \|v_i^j| - |v_i\|} \tag{2}$$

可知$v \cdot sim \in (0,1]$, 越趋近于 1, 相似轨迹在第i个采样时刻的速度相似度越高, 在速度上越接近真实轨迹.

定义 5(匿名区域ASR_i). 根据t_i时刻形成的$k-$匿名组内真实用户u_i和$k-1$个虚假用户u_i^j坐标组成的最小矩形区域.

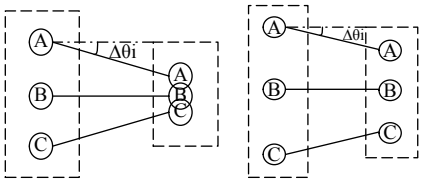
3 LPBSP 位置隐私保护方法

3.1 初始位置生成阶段

在连续请求状态下,轨迹方向偏移角度$d \cdot sim$, 生成的虚假轨迹在极端情况下可能会出现在某个采样时刻集中在一起的情况, 从而造成轨迹间距过小匿名效果不佳, 如图 3(a). 为保证生成的虚假轨迹在连续状态下更加分散的匿名效果, 需要在初始位置生成阶段进行一定的处理. 通过设置每个用户之间的距离限定, 扩大初始匿名区用户之间的间距, 使初始匿名位置点更加分散, 如图 3(b).

3.2 候选位置筛选阶段

若在某个采样点时刻匿名区域内用户数量发生了很大的变化, 则很有可能是真假用户的位置数据进行了某种变化, 从而增加真实用户位置泄露风险. 因此, 为避免虚假位置点急剧变化情况, 在生成候选位置点时需要对其进行筛选处理. 本文采用历史位置点和生成虚假位置点相结合的方式, 以用户密度为基础, 使真实轨迹和相似轨迹在采样时刻更加均匀, 提高匿名效果. 具体过程如下:



(a) 初始位置 (b) 扩大初始位置间距

图 3 初始位置生成状态

1) TTP 记录每个网格内用户请求密度, 并预先设置密度阈值σ. 对用户请求密度低于σ的区域剔除, 因为这些区域往往代表真实环境中不可到达区域 (如湖泊, 森林等).

2) 在进行候选位置筛选阶段前, 需要根据预设参数计算相似轨迹生成区域. 假设在t_i时刻真实用户在$l_i = (x_i, y_i)$的速度为v_i, 根据 TTP 预先设置的$d \cdot sim$和$v \cdot sim$确定生成区域范围. 以真实用户所在网格的请求密度为标准, 请求密度大的区域代表历史用户在该区域中进行频繁的 LBS 请求, 则在该区域中发生请求的概率相对较大. 在t_i时刻根据用户所在位置$l_i = (x_i, y_i)$判断此网格的历史请求密度为$d_i = \frac{N_i}{S_c}$(N_i: 真实用户所在网格内用户请求数量, S_c: 网格面积), 当 TTP 提交的$k-$匿名组内成员的密度相等时, 不易攻击者发现. 因此用户需要获得生成区域范围内所有网格用户请求密度, 然后进行排序. 在排序列表中, 选择与真实用户所在区域请求密度相似的网格, 并在此基础上随机确定数量为k个历史用户作为候选对象集C_i, 每个历史用户代表一个实际位置点. 将这些用户的请求密度进行归一化, 可得到在t_i时刻候选对象集C_i相对请求概率:

真实用户相对请求概率:

$$q_i = \frac{d_i}{d_u + \sum_{j=1}^{k-1} d_i^j}, (j = 1,2,\cdots,k-1) \tag{3}$$

虚假用户相对请求概率:

© 中国科学院软件研究所 http://www.c-s-a.org.cn

$$q_i^j = \frac{d_i^j}{d_u + \sum_{j=1}^{k-1} d_i^j} \tag{4}$$

其中, d_i^j表示t_i时刻第j个虚假用户u_i^j所在区域的历史请求密度, 则在该采样时刻该匿名组的熵为:

$$H_i = -\left(q_i \log_2 q_i + \sum_{j=1}^{k-1} q_i^j \log_2 q_i^j\right) \tag{5}$$

为使每次采样时刻匿名组用户所在区域中请求密度更加均匀, 我们需要一个衡量标准, 在理想情况下当真假用户所在区域密度完全相等时, 熵的值最大有:

$$H_c = -\log_2 \frac{1}{k} = \log_2 k \tag{6}$$

当 H 的值越趋近于H_c时, 历史用户请求分布越均匀, 攻击者越无法将真实用户从匿名用户中分离出来, 此时得到的H_i的匿名组用户最佳:

$$C_i = \arg\max H_i \tag{7}$$

确定此时候选对象集C_i的匿名区域范围ASR_i.

CLS 算法如下:

算法 1. CLS 算法

输入: 每个单元格内历史用户请求密度, t_i时刻用户位置坐标$l_i=(x_i,y_i)$
输出: t_i时刻$H_{\max}$, 候选对象集C_i, 匿名区域ASR_i

初始化m=0,H=0;
1. 将区域中各个单元格中的用户请求密度进行排序;
2. 剔除不可到达区域;
3. 从排序列表中选取与真实用户请求区域密度相似的3k 个候选点;
4. while $j \leqslant k-1$ //随机从中选取2k 个候选点
5. 当t_i时刻虚假用户位置数量小于 k–1 时;
6.计算t_i时刻的q_i, q_i^j和$H_i \leftarrow -\left(q_i \log_2 q_i + \sum_{j=1}^{k-1} q_i^j \log_2 q_i^j\right)$;
7. 进行H_i更新;
8. end while
9. 选取$H_{\max}$;
10. 确定t_i时刻熵最大的最佳候选对象集C_i;
11. 确定候选对象集C_i的匿名区域ASR_i;
12.输出ASR_i.

3.3 相似轨迹生成阶段

LPBSP 方法中相似轨迹是 TTP 根据真实用户的当前位置和前一次匿名区域内所有的匿名用户为基础, 通过预先设置的参数对生成的虚假轨迹进行一定约束, 使设置的相似轨迹更加贴近真实用户轨迹状态. 在采样时刻, 用户发起连续 LBS 请求, 真实用户 u 在t_i时刻以位置$l_i=(x_i,y_i)$发起匿名请求, 则在$[t_{i-1},t_i]$时间段内用户 u 的轨迹距离为$s=v_i \triangle t$, 偏移向量为$\overrightarrow{p_i}=(x_i-x_{i-1},y_i-y_{i-1})$. 虚假用户$u_j$在$t_{i-1}$时刻的位置为$l_{i-1}^j=\left(x_{i-1}^j,y_{i-1}^j\right)$, TTP 根据预先设置的轨迹方向偏移度$d\cdot sim$对要生成的相似轨迹进行最大偏移设置, 确定最大偏移角$\triangle\theta_i$, 用户u_i在t_i时刻的行驶速度$|v_i|$; 根据局部速度相似度$v\cdot sim$确定相似轨迹与真实轨迹的相似速度$|v_i^j|$. 在候选位置生成区域ASR_i内选择合适位置作为t_i时刻的虚拟位置点, TTP 获取此时虚拟点坐标$l_i^j=\left(x_i^j,y_i^j\right)$, 连接$l_{i-1}^j$, l_i^j作为虚假用户u_j^i在$[t_{i-1},t_i]$时间段相对真实用户u_i的相似轨迹.

DVS 算法如下:

算法 2. DVS 算法

输入: P={ID,l_i,v_i,t_i,POI}, $d\cdot sim$, $v\cdot sim$
输出: 相似路径集合 T={FT_1,FT_2,$\cdots$,FT_j,$\cdots$,FT_{k-1},T_r}

设置中间集合U=$null$
1. if u_i is fresh //出现用户请求u_i;
2.将u_i压入集合 U 中;
3. 从C_i中挑选u_j^i并计算每一个u_j^i加入后ASR_i面积;
4. If ASR_i< ω // ω 为预设值
5. 将 $u_i^j u_j^i$加入集合U中;
6. end if
7. end if
8. if (satisfy $d\cdot sim$ (u_i, u_i^j) &&$v\cdot sim$ (u_i, u_i^j) //对集合 U 中的每个虚假用户u_i^j进行$d\cdot sim$ && $v\cdot sim$期望范围检查
9. else
10. 将不符合的u_i^j从集合U中删除;
11. 计算 U 内用户数量$|U|$;
12. if $|U|\geqslant k$
13. 从中随机选取k个;
14. end if
15. end if
16. 确定t_i时刻真实用户u_i的用户轨迹T_r和虚假用户u_i^j的相似轨迹FT_j.

重复步骤 2 到 16 直到u_i停止进行 LBS 请求.

4 安全性分析

针对轨迹隐私保护的攻击者可以根据发起连续 LBS 请求的时间序列归纳用户大致行动方向, 根据相应已知条件发起推理攻击、最大移动边界攻击等攻击模式. 这是由于真实连续 LBS 的请求位置序列有一定的上下文联系, 攻击者可以分析真假轨迹的差异性推断出真实用户.

针对推理攻击, 若存在: $p_i\{u_i \in C_i\} = p_i^j\left\{u_i^j \in C_i\right\}$, $\forall(1 \leqslant i \neq j \leqslant k)$则, 可以抵抗推理攻击.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

证明: 在t_i时刻, 真实用户u_i被识别的概率$p_i\{u_i \in C_i\} = \frac{q_i}{q_i + \sum_{j=1}^{k-1} q_i^j}$, 此时, 根据公式 (3) 可得真实用户相对请求概率q_i, 可知q_i与u_i所在区域请求密度d_i有关, 有$p_i\{u_i \in C_i\} = f(d_i)$; 相似的, t_i时刻虚假用户u_i^j被识别的概率$p_i^j\{u_i^j \in C_i\} = f(d_i^j)$. 若$d_i = d_i^j$, 可保证$p_i\{u_i \in C_i\} = p_i^j\{u_i^j \in C_i\}$, 此时$H_i = H_c$.

针对最大移动边界攻击, 本文以真实用户运动方向和速度为基础, 由用户自定义参数范围, 充分考虑真实情况虚假位置的不可到达性, 通过速度和方向偏移相似程度限制抵抗最大移动边界攻击.

5 实验结果及分析

本文在 Windows sever 2008 服务器下采用 Java 语言开展 LPBSP 实验, 实验采取 Thomas Brinkhoff 路网生成器生成 Oldenburg 城市路网中的用户及兴趣点的实验数据, 并与文献[12]做对比实验, 验证本文 LPBSP 方法在匿名成功率、执行时间、熵三个维度上具有一定的贡献. LPBSP 试验参数如表 1 所示.

表 1　LPBSP 实验默认参数

参数名称	默认值
移动用户数量	1000
k 值	5
n	10
$v \cdot sim$	0.5

在匿名成功率方面, 如图 4 所示文献[12]基于虚假路径的设置, 只要有匿名需求便可构建虚假用户进行协作匿名, 没有考虑到虚假用户的真实性, 假如真实用户在海边, 构建的虚假用户很可能会在海里, 从而导致匿名效果的损失, 而本文 LPBSP 方法是基于用户查询的历史数据来构建协作匿名, 在真实用户发起查询时, 周围的历史数据可能会存在不足而导致匿名组构建失败, 因而在匿名成功率方面略低于文献[12], 但是本文方法仍然具有较高的匿名成功率.

在方法执行时间方面, 本文 LPBSP 方法由于需要获得生成区域范围内所有网格用户请求密度, 然后进行排序, 时间复杂度为 $n\log n$. 如图 5 所示, 文献[12]由于不考虑其他因素, 仅以构建虚假路径为核心, 因而执行效率高, 执行时间较短. 本文方法由于在构建协作用户的虚假路径时综合考虑了协作用户真实性, 协作路径相似性等诸多因素, 因而在执行时间方面略长与文献[12], 但是总体来看执行时间仍旧保持在较快的速度, 毫秒的差距不影响用户体验.

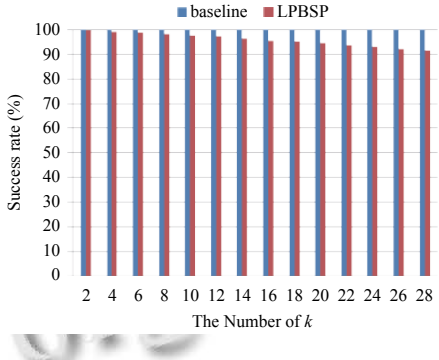


图 4　匿名成功率

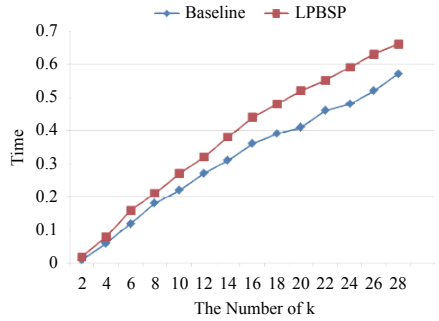


图 5　执行时间

在探究位置隐私保护度方面, 如图 6 所示, 本文采用熵的形式与文献[12]进行对比实验, 通过实验结果图所示, 可以发现, 熵随着 k 值的增大而逐渐增大, 本文 LPBSP 方法相较于文献[12]中具有一定优势, 原因在于本文考虑了真实用户的历史位置数据进行匿名, 并对用户的轨迹及移动速度进行了相似度的设置, 相较于文献[12]单纯的构建虚假路径更具有真实性, 使得匿名虚假轨迹与真实轨迹具有较高的相似性, 因而具有较好的隐私保护度. 当然相对于理想状态熵的最优值 $\log_2 k$ 还是有一定差距, 仍需要进行继续的优化.

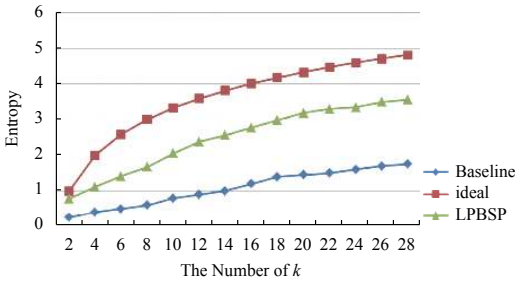


图 6　熵

© 中国科学院软件研究所 http://www.c-s-a.org.cn

6 结论与展望

本文针对连续查询位置隐私保护问题中可能存在的因协作用户交叠而暴露真实查询用户的问题, 提出了基于相似路径的位置隐私保护方法 (LPBSP), 首先采用用户历史数据构建初始协作匿名组, 然后利用用户历史位置数据、速度及轨迹相似度等对协作路径加以约束, 使得协作路径更具有真实性, 加以迷惑攻击者, 最后通过实验验证本文方法虽然在匿名成功率、执行时间上略逊色于文献[12], 但是在位置隐私保护度方面有了较好的应用, 在研究位置隐私方面有一定价值.

参考文献

1 胡德敏, 郑霞. 基于连续查询的用户轨迹 *k*-匿名隐私保护算法. 计算机应用研究, 2017, 34(11): 3421–3423, 3427. [doi: 10.3969/j.issn.1001-3695.2017.11.049]

2 Xu T, Cai Y. Location anonymity in continuous location-based services. Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems. Seattle, WA, USA. 2007. 39.

3 Gustav YH, Wang Y, Domenic MK, *et al*. Velocity similarity anonymization for continuous query location based services. Proceedings of 2013 International Conference on Computational Problem-Solving. Jiuzhai, China. 2013. 433–436.

4 Gruteser M, Liu X. Protecting privacy, in continuous location-tracking applications. IEEE Security & Privacy, 2004, 2(2): 28–34.

5 Götz M, Nath S, Gehrke J. MaskIt: Privately releasing user context streams for personalized mobile applications. Proceedings of 2012 ACM SIGMOD International Conference on Management of Data. Scottsdale, AZ, USA. 2012. 289–300.

6 张学军, 桂小林, 伍忠东. 位置服务隐私保护研究综述. 软件学报, 2015, 26(9): 2373–2395.

7 Wang Y, Xu DB, He X, *et al*. L2P2: Location-aware location privacy protection for location-based services. Proceedings of 2012 IEEE INFOCOM. Orlando, FL, USA. 2012. 1996–2004.

8 Pan X, Xu JL, Meng XF. Protecting location privacy against location-dependent attacks in mobile services. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(8): 1506–1519. [doi: 10.1109/TKDE.2011.105]

9 Latha K, Jayanthi S, Elavenil V. KRUPTO: Supporting privacy against location dependent attacks in wireless sensor network. Proceedings of 2013 International Conference on Communication and Signal Processing. Melmaruvathur, India. 2013. 908–912.

10 Hwang RH, Hsueh YL, Chung HW. A novel time-obfuscated algorithm for trajectory privacy protection. IEEE Transactions on Services Computing, 2014, 7(2): 126–139. [doi: 10.1109/TSC.2013.55]

11 Palanisamy B, Liu L. MobiMix: Protecting location privacy with mix-zones over road networks. Proceedings of the 2011 IEEE 27th International Conference on Data Engineering. Hannover, Germany. 2011. 494–505.

12 Kido H, Yanagisawa Y, Satoh T. An anonymous communication technique using dummies for location-based services. Proceedings of 2005 International Conference on Pervasive Services. Santorini, Greece. 2005. 88–97.

13 Wu XC, Sun GZ. A novel dummy-based mechanism to protect privacy on trajectories. Proceedings of 2014 IEEE International Conference on Data Mining Workshop. Shenzhen, China. 2015. 1120–1125.

14 Gao S, Ma JF, Shi WS, *et al*. TrPF: A trajectory privacy-preserving framework for participatory sensing. IEEE Transactions on Information Forensics and Security, 2013, 8(6): 874–887. [doi: 10.1109/TIFS.2013.2252618]

15 董玉兰, 皮德常. 一种基于假数据的新型轨迹隐私保护模型. 计算机科学, 2017, 44(8): 124–128, 139.

16 Ye AY, Li YC, Xu L. A novel location privacy-preserving scheme based on l-queries for continuous LBS. Computer Communications, 2017, 98: 1–10. [doi: 10.1016/j.comcom.2016.06.005]

© 中国科学院软件研究所 http://www.c-s-a.org.cn