一种基于假数据的新型轨迹隐私保护模型

董玉兰 皮德常
(南京航空航天大学计算机科学与技术学院 南京 211106)

摘 要 基于位置服务的普及给人们的生活带来了极大的便利,但同时也带来了严重的隐私泄露问题。基于假轨迹的隐私保护技术是目前比较流行的一种方法,但是现有的大多数假轨迹方法没有考虑到用户的个性化需求。基于此问题,提出了一种改进的隐私保护模型,并按照这个模型设计了一个假轨迹生成算法。该模型包含5个参数,分别命名为短期位置暴露概率、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,用户可以通过自身需求自定义这些度量,并通过假轨迹生成算法来生成假轨迹,从而避免隐私的泄露。实验结果表明,该算法可以在满足相同隐私的条件下生成较少的假轨迹,尤其是考虑了服务请求概率这一背景信息,该模型在保护移动对象轨迹隐私方面比之前的方案更有效。

关键词 基于位置的服务,轨迹隐私,隐私模型,假轨迹
中图法分类号 TP309 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.08.022

Novel Trajectory Privacy Preserving Mechanism Based on Dummies

DONG Yu-lan PI De-chang
(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract The popularity of location-based services(LBS) has brought great convenience to people's life, but it also brings serious privacy leakage at the same time. Dummy is a popular technology at present, but most existing methods do not take the individual needs of users into account. To address this problem, a improved privacy model was proposed, guided by which we designed a dummy trajectories generation algorithm. The model includes five reasonable parameters, namely short-term disclosure, long-term disclosure, trajectories distance deviation, trajectories local similarity and services request probability. People can customize these metrics through their own needs and generate dummies by dummy trajectories generation algorithm to avoid leakage of privacy. The experiment results show that the algorithm can generate fewer dummy trajectories to satisfy the same privacy-preserving requirement. So it's more effective than existing works in preserving movement trajectories, especially the probability of services request is discussed.

Keywords Location-based services(LBS), Trajectory privacy, Privacy model, Pseudonyms-based anonymization

1 引言

随着无线技术和带有GPS定位系统的移动设备的普及,出现了很多基于位置服务(Location Based Service, LBS)[1]的应用。各大厂商都抓住了巨大的商机,推行出各具特色的app,用户通过装载在移动终端上的app便可以随时随地地通过自己的位置信息获取相应的服务。例如,出行时获取的导航服务(高德地图),查找娱乐餐饮场所的生活服务(百度糯米),以及查找附近的人的社交服务(微信)等。

当使用LBS服务时,用户需要先将自己的精确位置和查询请求发送给位置服务商,位置服务商会将查询到的符合用户需求的结果返回给用户。如今的位置服务器已经收集了大量的位置信息,而位置服务器被认为是第三方不可信的服务器,因此一旦这些位置信息遭到泄露,用户的隐私也将随之暴露。例如,攻击者通过获取用户每天的轨迹信息了解到用户经常去的场所,从而分析出经济能力较强的用户,再根据该用户的运动规律选择地点、时间对其展开非法活动。因此,目前LBS服务确实存在着很大的安全隐患,并且用户对位置服务的使用规模还在不停的增长,所以我们必须高度重视轨迹隐私的安全问题。

研究学者们通过不断研究,获得了很多的研究成果。但是这些方法中对满足用户个性化需求的研究很少,而每个用户对隐私的定义是不同的,因此对于相同的信息,用户需要保护的程度也是不相同的。如果研究方法中不考虑用户的个性化需求,则会降低轨迹隐私保护的质量。例如,算法保护了用户特定的轨迹信息,而有些用户认为这些轨迹不是隐私数据;相反,有一些用户认为是隐私数据的运动轨迹,而算法却没有将其保护起来。文献[2]针对此问题提出了一种基于轨迹抑

到稿日期:2016-07-19 返修日期:2016-11-05 本文受国家自然科学基金(U1433116),中央高校基本科研业务费专项资金(NP2017208),南京航空航天大学研究生创新基地(实验室)开放基金(kfjj20171603)资助。
董玉兰(1994—),女,硕士生,主要研究方向为数据挖掘,E-mail:734823175@qq.com;皮德常(1971—),男,博士,教授,博士生导师,CCF会员,主要研究方向为数据挖掘、大数据管理与分析。

制的方法，该方案是在满足用户隐私需求的基础上选择合适的轨迹抑制点的数量来保证数据的有效性的。

现有的轨迹匿名策略大多是根据特定的需求制定的，比如针对移动社交网络的轨迹隐私保护[3]。目前对用户自身行为模式的研究较为成熟，但对满足用户个性化需求的算法研究得较少。本文在综合分析轨迹隐私保护技术的基础上，针对目前已有的轨迹隐私模型，提出了以背景知识为隐私参数的改进模型，用户可以通过自身的需求来设置参数，并且根据这个模型提出了新的假轨迹生成算法。

2　相关工作

国内外对这个问题的研究成果颇丰，已涌现出了大量优秀的解决方案。这些方案按照隐私保护系统结构可以分为两类：1）基于第三方匿名服务器的中心服务器结构，包括客户端、第三方匿名服务器端和服务提供商，第三方匿名服务器负责对客户端的请求数据进行隐私保护并对服务器端的返回结果进行处理；2）基于客户端的分布式结构，仅包含客户端和服务提供商。从技术上来看，这些解决方案大致分为 3 类：空间匿名、区域混合和假轨迹法。

空间隐匿借助于第三方匿名服务器，首先会收集用户的轨迹信息，然后对每个采样点都构造一个至少包括 k 个用户的空间隐匿区域作为服务请求发送给 LBS 提供商，从而实现 k-匿名来保护用户的轨迹隐私。k-匿名技术是目前应用最广泛的一种方法，文献[4]利用用户的历史轨迹充当假轨迹以达到 k-匿名的效果，由于使用的是用户的历史轨迹，因此保证了假轨迹的真实有效性，从而有效保护了用户的隐私。文献[5]基于 k-匿名的原理提出了一种基于贪心法的位置 k-匿名算法，该算法对目标用户周围的用户进行分析，从而选择最优的相邻区域并用贪心法构建匿名区域，这种方法充分考虑到了周围用户的分布情况，并对生成的匿名区域去除冗余以达到最优区域。区域混合技术是构造一个混合区域（Mix-zone），用户一旦进入该区域就会与其他用户混合起来，无法区分。文献[6]提出了基于混合区域的算法，该算法充分结合了路网的特性、时间以及用户运行的规律来构建一个混合区域，用户在进入混合区域后会周期性地改变假名，从而达到隐私保护的效果。而假轨迹方法是不需要借助于值得信赖的第三方服务器或者 LBS 提供商的，Kido 等[7]首先提出了一种新的匿名通信技术，用户可以将其真实位置和假位置一起发送给 LBS 提供商，从而保护用户的位置隐私。文献[8-10]专注于真实环境的物理约束，分别提出了两种新的轨迹生成算法——Dum-Grid[8]和 Dum-P[9]。前者在网格模式下围绕用户轨迹生成假轨迹，后者按照用户移动生成更加真实的假轨迹。接着，文献[10]提出了一种改进的 Dum-P-Cycle 来弥补位置隐私需求的不足。You 等[11]提出了一种新的轨迹隐私模型，其包含 3 个参数即短期泄露概率、长期泄露概率和轨迹偏移距离，并提出了两种假轨迹生成模式，分别为随机生成模式和旋转生成模式，用这两种模式生成了 k 条假轨迹，从而实现轨迹 k-匿名。Wu 等[12]改进了 You 等[11]提出的轨迹隐私模型，并提出了一个自适应轨迹生成算法（Adaptive Dummy Trajectories Generation Algorithm，ADTGA）。该算法可以根据用户自定义的轨迹隐私参数生成假轨迹。Lei 等[13]受 You 等[11]提出的假轨迹生成模式的启发，提出了 k-交叉轨迹生成算法，将用户移动模式考虑在内，并提出增加轨迹间交点的想法，从而使生成的假轨迹更难以区分。

3　新型隐私保护模型

3.1　条件假设

假设没有信赖的第三方服务器来做位置匿名处理，同时利用无线网络进行通信，移动端可以通过 GPS 或其他定位技术定位。为了简化讨论，本文把移动空间划分为网格，每个网格用(x,y)来表示，其中 x 代表横坐标，y 表示纵坐标。用户 u_i 发送给 LBS 服务器的请求信息用 M 表示，$M=\{u_i,\langle L_i^t, L_{d1}^t, L_{d2}^t,\cdots,L_{dn}^t\rangle\}$，其中 u_i 是用户的一个 ID，L_i^t 是用户在 t 时刻的真实位置，$L_{d1}^t, L_{d2}^t,\cdots,L_{dn}^t$ 表示用户在 t 时刻的 n 个假位置。因此，连续发送 m 条请求信息，用户 U_i 的真实轨迹为 $\{L_i^1, L_i^2,\cdots,L_i^m\}$，第 x 条假轨迹表示为 $\{L_{dx}^1, L_{dx}^2,\cdots,L_{dx}^3\}$，以上 L_i^j 表示用户 U_i 在 j 时刻的位置，L_{dx}^j 表示假用户 d_x 在 j 时刻的位置。

3.2　隐私保护模型

本文提出的轨迹隐私保护模型共有 5 个隐私度量，这些度量可以由用户根据自身需求来自定义设置。一个严格的隐私模型能大大降低用户轨迹隐私泄露的概率，从而更好地保护用户隐私。轨迹隐私模型的定义如下。

定义 1（位置泄露概率 Δt-SD）　该参数表示用户位置被成功识别的概率，即在一段时间 Δt 内泄露真实位置的概率。不同于之前部分工作在事件点 t 上考虑位置泄露的概率，本文更加合理地在 Δt 时间内考虑短期位置泄露的概率。例如，假设真实轨迹 $T_r=\{L^1,L^2,\cdots,L^n\}$，假轨迹 T_{d1} 中第 i 个位置是真实轨迹里的第 $i-1$ 个位置，因此假轨迹可以表示为 $T_{d1}=\{L^0,L^1,\cdots,L^{n-1}\}$，另一假轨迹 T_{d2} 与 T_{d1} 和 T_r 不相关，记为 $T_{d2}=\{L^{1'},L^{2'},\cdots,L^{n'}\}$。假如我们考虑在一个时间点上而不是在一段时间内计算位置泄露的概率，那么轨迹集 $\{T_r,T_{d1}\}$ 和 $\{T_r,T_{d2}\}$ 的位置泄露概率相同，都为 1/2。显然，这不符合我们的判断，因为前者泄露的概率明显要比后者大。因此，本文在考虑位置泄露概率时将计算一段时间内的泄露概率，具体计算如式(1)所示：

$$\Delta t\text{-}SD=\frac{1}{m}\times\left[\sum_{i=0}^{\lfloor\frac{m}{\Delta t}\rfloor-1}\frac{\Delta t}{D_{1+\Delta t\times i,(1+i)\times\Delta t}}+\frac{1}{\frac{D_{end+1,m}}{m-end}}\right] \tag{1}$$

其中，$end=\left\lfloor\frac{m}{\Delta t}\right\rfloor\times\Delta t$，$D_{i,j}$ 表示所有轨迹包括真实轨迹和假轨迹从 i 时刻到 j 时刻的位置总数。

定义 2（轨迹泄露概率 LD）　该参数表示用户轨迹被成功识别的概率。n 表示轨迹的数量，假设这 n 条轨迹中有 k 条轨迹与其他轨迹有交点，则剩余 $n-k$ 条轨迹与任何其他轨迹都不存在交点。其中，k 条相交轨迹组成的可能轨迹数量记为 T_k，则有：

$$LD=\frac{1}{T_k+(n-k)} \tag{2}$$

定义 3（轨迹距离偏差 TDD）　该参数表示假轨迹的分布情况。它表示在每个时间点上所有轨迹（包括假轨迹与真实轨迹）位置之间的距离。假设有 n 条假轨迹，记作 T_{dx}（$x\in$

$[1,n]$），真实轨迹记作 T_r，每条轨迹包含 m 个位置。则定义 TDD 为：

$$TDD=\frac{1}{m}\times\frac{2}{n\times(n-1)}\times\sum_{i=0}^{n}\sum_{k=i+1}^{n}\sum_{j=1}^{m}dist(L_{di}^{j},L_{dk}^{j}) \quad (3)$$

其中，m 是每条轨迹的位置数，n 表示假轨迹数目，L_{di}^{j} 表示假轨迹 T_{di} 的第 j 个位置，$dist(L_{di}^{j},L_{dk}^{j})$ 表示两位置之间的欧氏距离。

定义 4（轨迹局部相似度 TLS） 该参数表示假轨迹在局部时刻的相似情况，它表明在某个时间点上，假轨迹与真实轨迹之间的偏移角度之差。设轨迹 T 的第 $i(i\in(1,n))$ 个位置为 L^i，则其前驱位置为 L^{i-1}，后继位置为 L^{i+1}，则它们之间的夹角为向量(L^i,L^{i-1})和向量(L^i,L^{i+1})之间的夹角，记为 θ^i $(\theta\in[0,\pi])$。假设现在有 n 条假轨迹 $T_{dx}(x\in[1,n])$和真实轨迹 T_r，每条轨迹含有 m 个位置，则参数轨迹局部相似度定义为：

$$TLS=\frac{1}{m-2}\times\frac{1}{n}\left[\sum_{i=2}^{m-2}\sum_{r=1}^{n}|\theta_r^i-\theta_{dr}^i|\right] \quad (4)$$

其中，θ_r^i 表示轨迹 T_r 的第 i 个夹角，θ_{dr}^i 表示假轨迹 T_{dr} 的第 i 个夹角。

定义 5（请求概率 RP 与奇点 SR） 该参数表示轨迹空间的背景信息。RP 表明在该轨迹空间下每个网格向 LBS 服务器发送请求的概率。假设一个轨迹空间可以均匀划分为 $n\times n$网格空间，则每个区域的请求概率如式(5)所示：

$$q_{(i,j)}=\frac{\text{训练集中网格}(i,j)\text{发送请求数}}{\text{训练集中总请求数}} \quad (5)$$

其中，(i,j)表示网格，$i\in[1,n]$，$j\in[1,n]$，满足$\sum_{i=1}^{n}\sum_{j=1}^{n}q_{(i,j)}=1$。

在此基础上，定义奇点（Strange Location）表示真实轨迹与假轨迹对应时刻位置点服务请求概率相差超过一个阈值或假轨迹上服务请求概率为 0 的位置点。

图 1 是一个简单的轨迹示意图。T_r 表示用户真实轨迹，T_{d1} 和 T_{d2} 表示两条假轨迹。而每条轨迹各时刻位置点的坐标如表 1 所列，可以看出每条轨迹有 6 个位置，若 $\Delta t=2$，则在每个位置计算泄露概率（见表 1）时应考虑前后 Δt 时间段内总的泄露概率，因此图中 Δt-SD 为：$\frac{1}{6}\times(\frac{1}{3}+\frac{2}{5}+\frac{2}{5}+\frac{2}{5}+\frac{1}{3}+\frac{1}{3})=\frac{11}{30}$。而 T_r 与 T_{d2} 在 $t=4$ 时相交于(4,3)处，注意 T_r，T_{d1} 在(3,3)处的交点在时空间里是不相交的，所以 $k=2$，$N_k=4$，$LD=\frac{1}{4+3-2}=\frac{1}{5}$。同样，每个位置点的位置偏移距离和位置点的角度偏移分别如表 1 所列，则 $TDD=\frac{1}{6}\times(2.4+2.4+1.6+0.7+2.0+2.0)=1.7$，$TLS=\frac{1}{6-2}\times(22.5+45+0+22.5)=22.5$。

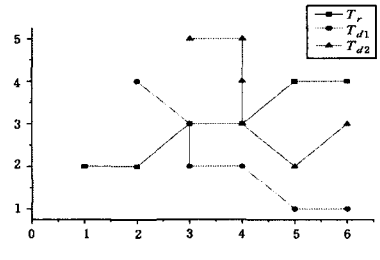

图 1 轨迹示意图

表 1 图 1 中的轨迹参数

时间点 t	1	2	3	4	5	6
T_r	(1,2)	(2,2)	(3,3)	(4,3)	(5,4)	(6,4)
T_{d1}	(2,4)	(3,3)	(3,2)	(4,2)	(5,1)	(6,1)
T_{d2}	(3,5)	(4,5)	(4,4)	(4,3)	(5,2)	(6,3)
Δt-SD	1/3	2/5	2/5	2/5	1/3	1/3
TDD	2.4	2.4	1.6	0.7	2.0	2.0
TLS	—	22.5	45	0	22.5	—

4 假轨迹生成算法

4.1 算法思路

算法开始前，首先需要一个数据训练集来训练每个网格的服务请求概率。算法开始，将两个轨迹集 $Dummies$，$Candidate$ 分别置空。接着对真实轨迹 T 的旋转位置和旋转角度进行遍历来生成假轨迹，其中满足轨迹偏移距离 TDD 的轨迹作为备选继续执行接下来的步骤。然后在调整轨迹时，选择假轨迹 $dummy$ 中$(|T|\times pk)$（其中$|T|$表示真实轨迹 T 的位置数目，pk 为调整参数）个位置，将其按轨迹位置点分布的疏密情况移动一个可接受的距离。这个距离不可过大，否则会破坏轨迹位置近邻原则，导致轨迹很容易被攻击者识别，从而无法保护用户轨迹隐私；同时这个距离也不能太小，否则无法达到对轨迹扰动的目的。因此本文选择一个阈值 R，R 是通过训练得到的同一轨迹相邻位置点的最大间距，随机在阈值 R 范围内选择一个值作为调整距离，将位置点移动至相对稀疏的空间，从而达到更均匀的轨迹分布。同时，需要对不满足服务的请求概率参数和轨迹相似度参数的轨迹点进行调整，将调整后满足需求的假轨迹加入到备选集 $Candidate$ 中，并计算其 TSD 和 LD，将参数最小的加入到 $Dummies$ 中。重复该步骤直到 $Dummies$ 中假轨迹的隐私参数值满足用户自定义要求为止。最后，输出假轨迹集 $Dummies$。

4.2 算法步骤

本文提出的算法包括两个步骤：1）通过真实轨迹的旋转得到备选的假轨迹集，然后根据隐私模型下的 Δt-SD 和 LD 参数对备选集进行筛选；2）对假轨迹进行基于轨迹分布和请求概率参数的调整。具体算法如下。

算法 1 自定义假轨迹生成算法（Custom Dummy Trajectories Generation Algorithm）

输入：用户真实轨迹 T，调整参数 pk，隐私模型参数（TSD，LD，TDD，TLS，RP）

输出：假轨迹集 Dummies

步骤 1 将 Dummies 置为空集；

步骤 2 将 Candidate 置为空集；

步骤 3 遍历轨迹 T 的 m 个位置和角度，旋转生成一条假轨迹 dummy=generate(T,t,θ)；

步骤 4 计算该假轨迹与真实轨迹和已生成假轨迹之间的轨迹距离偏差，若超过阈值，则返回步骤 2，否则继续；

步骤 5 perturbation(dummy,pk)，调整假轨迹，若满足 TDD 和 RP，将 dummy 加入到备选集 Candidate 中，否则返回步骤 2；

步骤 6 对 Candidate 中的每条轨迹，计算其与已生成假轨迹之间的 TSD 和 LD；

步骤 7 将 Candidate 中隐私参数值最小的轨迹加入到 Dummies 集中；

步骤 8 计算 Dummies 的 TSD，LD，并将其分别记为 tsd 和 ld；

步骤 9　若 tsd≤TSD∩ld≤LD，则继续；否则返回步骤 2；

步骤 10　返回假轨迹集 Dummies。

5　实验及仿真

5.1　基本设置

本次实验采用的数据集是由 Thomas Brinkhoff 移动对象生成器[14]基于德国古登堡市地图生成的。将地图按单位长度为 100 进行网格划分，每个网格用网格中心坐标表示。本实验所有的代码都是由 Java 编写的，实验环境为Windows 7 操作系统，内存空间是 4GB，其处理器为 2.27GHz 的 Intel (R) Core(TM) i3 CPU。编程环境为 MyEclipse 10。

本次实验主要涉及到的参数有：

1)Δt-SD 和 LD，分别表示短期位置暴露概率和长期轨迹暴露概率，与假轨迹数量 k 息息相关；

2)k，即假轨迹数量，表示轨迹匿名程度和隐私保护程度；

3)TLS，RP 和 TDD 等，描述生成假轨迹的运动模式以及分布情况，也在一定程度上表示隐私保护程度。本文在设置参数时采用了实验设计法[15]，从 TLS={10,20,30,40,50,60,70}，TDD={5,10,15,20,25,30,35}，RP={0.00005,0.0001,0.00015,0.0002,0.00025,0.0003,0.00035}的 343 种组合中选出了最优参数设置。实验表明，当 $TDD\in[15,20]$，TLS=40，RP=0.0001 时，假轨迹的效果最优，因此本文实验数据都取了最优参数。

4)调整参数 pk，描述调整过程的力度，通常设置为百分数。而文献[12]中的实验指出，pk 值过大和过小都会产生大量假轨迹，而 pk=0.5 时达到最好的效果。因此本文在后面的实验中将 pk 取值置为 0.5。

实验主要比较了轨迹隐私模型各参数与假轨迹生成数量的关系；同时，为了显示该假轨迹生成算法在生成较少的假轨迹数量的情况下就可以满足轨迹隐私需求，本次实验同时实现了随机生成算法、旋转生成算法以及文献[10]提出的算法，并进行纵向对比实验。

5.2　实验结果

5.2.1　TSD 对假轨迹数量 k 的影响

图 2 验证并给出了 TSD 对假轨迹数量 k 的影响，其中设置 ld=0.4。从图 2 可以看出，随着 TSD 值的减小，3 种算法的假轨迹生成数量都在增长，这是由于 TSD 在一定程度上代表着轨迹隐私保护要求，TSD 值越小，轨迹隐私保护程度就越高，假轨迹数量 k 也越大。可以看到，由于假轨迹生成的随机性和位置点的低概率重合性，随机模式生成的假轨迹数量较少。得益于轨迹生成后的调整，本文方案生成的假轨迹数量仅多于随机模式，而旋转模式的表现则差强人意。

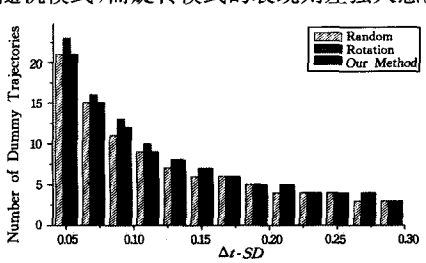

图 2　短期位置概率对假轨迹数量 k 的影响

5.2.2　LD 对假轨迹数量 k 的影响

图 3 验证并给出了 LD 对假轨迹数量 k 的影响，设置 TSD=0.5。从图 3 可以看出，随着 LD 值的减小，3 种算法的假轨迹生成数量都在增长，这是由于 LD 在一定程度上代表着轨迹隐私保护要求，LD 值越小，轨迹隐私保护程度就越高，假轨迹数量 k 也越大。可以看到，由于位置点的随机性，随机模式很难产生交点，长期轨迹隐私概率较高，因此在同样的轨迹隐私概率参数上，随机模式要比其他两种方案生成的假轨迹数量多。得益于尽可能多地创造出更多的交点这一初衷，本文方案生成的假轨迹数量最少，表现较好。旋转模式生成的假轨迹数量仅多于本文的方案，而随机模式的表现则差强人意。

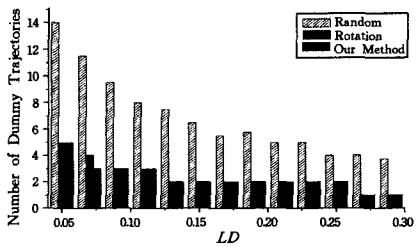

图 3　长期轨迹暴露概率对假轨迹数量 k 的影响

5.2.3　假轨迹数量对 TSD 的影响

如图 4 所示，设置参数 $TDD\in[15,20]$，TLS=40，RP=0.0001，pk=0.5。从图 4 可以明显看出，随机模式生成的假轨迹集短期位置暴露概率最低，因为它是在真实位置周围随机生成 k 个位置，所以这 k 个位置难以发生重合的情况，从而使位置暴露概率相对较低。旋转模式和本文方案在本质上都是真实轨迹围绕某个位置旋转得到假轨迹，而且在设计之初，希望增加交点的数量来降低轨迹暴露概率，因此其位置暴露概率都比随机模式的高。但本文方案在假轨迹生成后进行了位置调整，在一定程度上将一个重合点调整至较稀疏处，因此其位置暴露概率低于旋转模式。

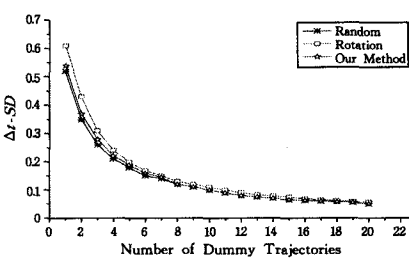

图 4　假轨迹数量对短期位置暴露概率参数的影响

5.2.4　假轨迹数量 k 对 LD 的影响

图 5 验证并给出了假轨数量 k 对长期轨迹暴露概率的影响。其中本文设置参数 TLS=40，$TDD\in[15,20]$，RP=0.0001，pk=0.5。从图 5 可以看出，随机模式由于位置点的随机性，很难产生交点，因此它生成的假轨迹集暴露概率偏高。而旋转模式和本文提出的方案都采用旋转真实轨迹的方法生成假轨迹，并且在设计之初，通过尽可能合理地增加交点的数量来增加相交轨迹的数量，使假轨迹集产生尽可能多的轨迹，从而使本文方案和旋转模式在轨迹暴露概率上有着优良的表现。

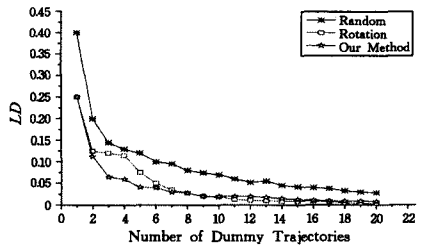

图 5 假轨迹数量对长期位置暴露概率参数的影响

5.2.5 假轨迹数量 k 对 TDD 的影响

图 6 验证并给出了假轨迹数量 k 对轨迹偏移距离参数的影响。从直观上讲，轨迹偏移距离不应太大，否则会破坏速度一致性原则；同时也不应太小，否则假轨迹分布较为密集，导致用户隐私很容易被恶意攻击者窃取。因此我们希望轨迹偏移距离在一个可接受的范围内，并且此时的轨迹位置分布保留上一时刻的轨迹位置分布特征。实验表明，当 TDD 在 15～20 之间时可以达到最好的效果。如实验结果所示，随机模式与旋转模式随着轨迹生成数量的增加，轨迹偏移距离变大，但当 $k>10$ 时，旋转模式值趋于平稳，并稳定于 20 左右；而本文方法由于可以自定义控制轨迹偏移参数的量，因此取得了不错的表现。

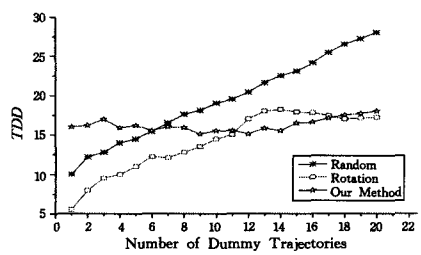

图 6 假轨迹数量对轨迹偏移距离的影响

5.2.6 假轨迹数量 k 对 TLS 的影响

图 7 验证并给出了假轨迹数量 k 对轨迹局部相似度参数的影响。其中设置参数 $TLS=40$，$TDD\in[15,20]$，$RP=0.0001$，$pk=0.5$。而 TLS 代表两条轨迹中位置角度偏移的大小，角度偏移越大则两条轨迹相似程度越小，角度偏移越小则轨迹的相似程度越大，因此 TLS 越小，假轨迹与真实轨迹的相似度越高，轨迹特征破坏得就越小。从图 7 可以得出结论，旋转模式中真实轨迹自身旋转生成假轨迹的特性较好地保留了原始轨迹的基本轨迹特征，所以假轨迹集与真实轨迹的相似度很高。由于本次实验是基于均匀网格划分的思想，每个网格中点都用网格中心泛化，因此对轨迹旋转过程后的位置点都进行了泛化处理，从而导致即使是在旋转模式下依然有相似度误差，属正常现象。随机模式中，位置点的随机性导致生成的假轨迹几乎完全抛弃了真实轨迹特征，与真实轨迹相差较大也在情理之中。文献[12]提到的自适应轨迹生成算法虽然在假轨迹生成时采用旋转模式，但对每条假轨迹做扰动的过程严重破坏了真实轨迹的轨迹特征，因此会使假轨迹与真实轨迹产生相对较大的误差，甚至在假轨迹数量 k 较大时，其轨迹相似度要低于随机模式。本文方案吸取了其教训，考虑加入轨迹相似度参数来控制假轨迹的生成，并对调整过程中可能产生的局部相似度相差较大的情况进行了处理。因此本文方案保留了一定的真实轨迹特征，并且可以根据用户隐私保护要求进行自定义控制。

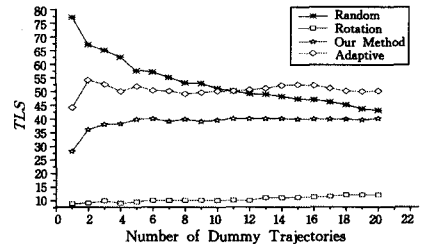

图 7 假轨迹数量对长期轨迹暴露概率的影响

5.2.7 轨迹数量 k 对奇点 SR 数目的影响

图 8 验证并给出了假轨迹数量 k 对奇点数量的影响。其中设置参数 $TLS=40$，$TDD\in[15,20]$，$RP=0.0001$，$pk=0.5$。由于本文方案在每条假轨迹生成后都对其进行调整和筛选，即对假轨迹中可能的奇点进行调整，若无法调整至符合隐私需求的位置，则舍弃这条假轨迹，因此相比其他 3 种算法，本文算法的奇点数量明显减少。另外 3 种算法由于没有考虑背景知识，在奇点数量的控制上都难以让人满意。

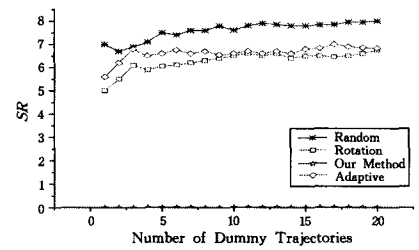

图 8 假轨迹数量对奇点数量的影响

结束语 本文对基于假轨迹的轨迹隐私保护方案展开了深入的研究，并提出了一种改进的轨迹隐私保护模型以及轨迹生成算法。本文采用的算法基于旋转生成及创造交点的思想，生成的假轨迹能较好地保留真实轨迹的特征，因此在轨迹暴露概率和轨迹相似度方面有较好的表现；同时对生成的假轨迹进行适当调整，增加假轨迹的随机性，因此在位置暴露概率上的表现比纯粹的旋转模式更优秀；另外，本文在隐私度量上考虑了服务请求概率这一背景知识，对不满足隐私需求的假轨迹进行调整甚至舍弃，因此本文方案不仅能满足用户的个性化需求，而且使得生成的假轨迹更加真实。

参 考 文 献

[1] HUO Z, MENG X F. A Survey of Trajectory Privacy-Preserving Techniques[J]. 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[J]. 计算机学报，2011，34(10)：1820-1830.

[2] ZHAO J, ZHANG Y, LI X, et al. A Trajectory Privacy Protection Approach via Trajectory Frequency Suppression[J]. Chinese Journal of Computers, 2014, 37(10): 2096-2106. (in Chinese)
赵婧，张渊，李兴，等. 基于轨迹频率抑制的轨迹隐私保护方法[J]. 计算机学报，2014，37(10)：2096-2106.

（下转第 139 页）

cations and Software，2014，31(8)：17-19. (in Chinese)
刘洪伟，梁飞，朱慧. 面向隐私保护推荐系统的安全两方协议研究[J]. 计算机应用与软件，2014，31(8)：17-19.

[4] CRONJ S，MANDALA N. Fully homomorphic encryption over the integers with shorter public keys [M]// Advances in Cryptology-CRYPTO 2011. Berlin：Springer，2011：94-145.

[5] DWORK C. Difference Privacy[C]// Proc of the 33rd International Colloquium on Automata，Languages and Programming，Part II. 2006：1-12.

[6] PENG F，ZENG X W，LIU L，et al. Privacy preserving recommendation method based on groups[J]. Application Research of Computers，2015，32(3)：869-871. (in Chinese)
彭飞，曾学文，刘磊，等. 一种基于群组推荐的用户隐私保护方法[J]. 计算机应用与研究，2015，32(3)：869-871.

[7] ZHANG F Z，LIU T，FENG S S. Improved Privacy-preserving Collaborative Filtering Recommendation Algorithm [J]. Computer Engineering，2010，36(16)：126-134. (in Chinese)
张付志，刘亭，封素石. 一种改进的隐私保持协同过滤推荐算法[J]. 计算机工程，2010，36(16)：126-134.

[8] DWORK C. Difference privacy：a survey of results[C]// Theory and Applications of Models of Computation. 2008：1-9.

[9] XIONG P，ZHU T Q，WANG X F. A Survey on Differential Privacy and Applications[J]. Chinese Journal of Computers，2014，37(1)：101-122. (in Chinese)
熊平，朱天清，王晓峰. 差分隐私保护及其应用[J]. 计算机学报，2014，37(1)：101-122.

[10] ZHOU J，CHEN C，YU N H. Tag Clustering Algorithm Using Object-based Feature Vector[J]. Journal of Chinese Computer Systems，2012，33(3)：525-530. (in Chinese)
周津，陈超，俞能海. 采用对象特征向量表示法的标签聚类算法[J]. 小型微型计算机系统，2012，33(3)：525-530

[11] AGGARWAL C C. On randomization，public information and the curse of dimensionality[C]// Proc of the 23rd IEEE International Conference on Data Engineering. IEEE Computer Society，2007：136-145.

[12] ZHANG J C，TANG X，SUN Y. Study of Crowds Anonymous Communication System[J]. Microcomputer Information，2012，28(7)：128-130. (in Chinese)
章敬崇，唐旭，孙宇. Crowds 匿名通信系统研究[J]. 微计算机信息，2012，28(7)：128-130.

[13] DWORK C. A Firm Foundation for Private Data Analysis[C]// Communications of the ACM. 2011：86-95.

[14] LI Y，HAO Z F，WEN W，et al. Research on Differential Privacy Preserving k-means Clustering[J]. Computer Science，2013，40(3)：287-290. (in Chinese)
李杨，郝志峰，温雯，等. 差分隐私保护 k-means 聚类方法研究[J]. 计算机科学，2013，40(3)：287-290.

[15] LONG J. Research on Hybrid Privacy Models and Algorithms for Collaborative Filtering [D]. Guilin：Guangxi Normal University，2015. (in Chinese)
龙军. 面向协同过滤推荐的混合隐私保护技术和算法研究[D]. 桂林：广西师范大学，2015.

[16] MCSHERRY F，MIRONOV I. Differential private recommender system：building privacy into Netflix prize conteders[C]// Proc. of KDD. 2009：627-636

[17] XIAN Z Z，LI Q L. Research on application of differential privacy in recommender system[J]. Application Research of Computers，2016，33(5)：1549-1557. (in Chinese)
鲜征征，李启良. 差分隐私保护在推荐系统中的应用研究[J]. 计算机应用研究，2016，33(5)：1549-1557.

(上接第 128 页)

[3] HUO Z，MENG X F，HUANG Y. PrivateCheckIn：Trajectory Privacy-Preserving for Check-In Services in MSNS[J]. 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[J]. 计算机学报，2013，36(4)：716-726.

[4] XU T，CAI Y. Exploring Historical Location Data for Anonymity Preservation in Location-Based Services[C]// The 27th Conference on Computer Communications. 2008：547-555.

[5] SHI M Y. Research on Trajectory Privacy Protection in Location Based Service[D]. Nanjing：Nanjing University of Posts，2014. (in Chinese)
史敏仪. 面向位置服务的轨迹隐私保护技术研究[D]. 南京：南京邮电大学，2014.

[6] PALANISAMY B，LIU L. Attack-resilient mix-zones over road-networks：architecture and algorithms[J]. IEEE Transactions on Mobile Computing，2015，14(3)：495-508.

[7] KIDO H，YANAGISAWA Y，SATOH T. An anonymous communication technique using dummies for location-basedservices [C] // International Conference on Pervasive Services，2005 (ICPS'05). 2005：88-97.

[8] SUZUKI A，IWATA M，ARASE Y，et al. A user location anonymization method for location based services in a realenvironment [C]// Sigspatial International Conference on Advances in Geographic Information Systems. ACM，2010：398-401.

[9] KATO R，IWATA M，HARA T，et al. A dummy-based anonymization method based on user trajectory with pauses[C]// International Conference on Advances in Geographic Information Systems. 2012：249-258.

[10] KATO R，IWATA M，HARA T，et al. User Location Anonymization Method for Wide Distribution of Dummies[M]// Database and Expert Systems Applications. Springer Berlin Heidelberg，2013：259-273.

[11] YOU T H，PENG W C，LEE W C. Protecting Moving Trajectories with Dummies[C] // International Conferenceon Mobile Data Management. IEEE，2007：278-282.

[12] WU X，SUN G. A Novel Dummy-Based Mechanism to Protect Privacy on Trajectories[C]// IEEE International Conference on Data Mining Workshop. IEEE，2014：1120-1125.

[13] LEI P R，PENG W C，SU I J，et al. Dummy-Based Schemes for Protecting Movement Trajectories [J]. Journal ofInformation Science & Engineering，2012，28(2)：335-350.

[14] BRINKHOFF T. Generating Network-Based Moving Objects [C]// International Conference on Scientific and Statistical Database Management. IEEE Computer Society，2000：253-255.

[15] MONTGOMERY D C，KOWALSKI S M. Design and Analysis of Experiments：Minitab Companion[M]. Wiley，2010.