个人移动数据收集中的多维轨迹匿名方法

王丽娜[1,2],彭瑞卿[1,2],赵雨辰[1,2],陈 栋[1,2]
(1.空天信息安全与可信计算教育部重点实验室,湖北武汉 430072;2.武汉大学计算机学院,湖北武汉 430072)

摘 要: 在情景感知位置服务中,移动互联网络的开放性使得个人移动数据面临巨大的安全风险,移动数据的时空关联特性对个人数据的隐私保护提出重大挑战.针对基于时空关联的背景知识攻击,本文提出了一种多维的轨迹匿名隐私保护方法.该方法在匿名轨迹数据收集系统的基础上,基于多用户协作的隐私保护模式,通过时间匿名和空间匿名算法,实现用户的隐私保护.实验结果表明,该方法可以有效的对抗基于位置和移动方式的背景知识攻击,满足了 k-匿名的隐私保护要求.

关键词: 隐私保护; 匿名; 个人移动数据; 基于位置的服务
中图分类号: TP393.0 **文献标识码:** A **文章编号:** 0372-2112 (2013) 08-1653-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2013.08.032

Multi-Dimensional Trajectory Anonymity in Collecting Personal Mobility Data

WANG Li-na[1,2], PENG Rui-qing[1,2], ZHAO Yu-chen[1,2], CHEN Dong[1,2]
(1. *Key Laboratory of Aerospace Information Security and Trusted Computing*, *Ministry of Education*, *Wuhan*, *Hubei* 430072, *China*;
2. *School of Computer*, *Wuhan University*, *Wuhan*, *Hubei* 430072, *China*)

Abstract: In the Context-Aware location services, the opening characteristic of the mobile network brings some security risks for personal mobility data. The spatial-temporal correlation of the mobility data is a great challenge for protecting privacy of the users. To solve the background knowledge attack based on the spatial-temporal correlation, we proposed a privacy protection method based on multi-dimensional trajectory anonymity in this paper. This method is built on the anonymous trajectory data collection system and achieves the protection of user privacy through spatial-temporal anonymity algorithm based on the multi-user collaboration privacy protection mode. The experimental results demonstrated that this method can prevent the background knowledge attack based on the position and moving mode effectively and meets the demands of the privacy protect of k-Anonymity.

Key words: privacy protect; anonymity; personal mobility data; location-based service

1 引言

随着位置服务的发展,基于情景感知的位置服务由于其能够为用户提供更为人性化的服务而成为关注的焦点.研究者需要收集大量的移动位置数据,以提供更为丰富的上下文信息.然而位置数据的收集可能会泄露用户的行为习惯,兴趣爱好等隐私信息[1].因此如何保证位置数据收集过程中的用户隐私安全是目前研究的重点.

简单的删除用户名或使用假名的匿名方法并不能有效地消除隐私威胁.位置作为个人的一种准身份标识[2],具有较强的时空关联特性.攻击者可以利用对手知识攻击,并结合外部数据源来重新标识用户的身份.甚至通过少量的轨迹样本唯一的标识用户并精确的描述用户的轨迹[3].因此一个安全有效的轨迹收集系统既要收集到精确的轨迹信息,又要隐藏用户的身份,切断用户的轨迹与身份之间的时空关联.

随着研究的深入,许多文献提出了基于位置服务的隐私保护方法.这些方法要么是通过时空泛化要么是采用混淆确切的用户位置来实现隐私保护.其核心是 k-匿名模型[4],在该模型中 k 个用户混合在一起,使得攻击者无法很好的区分单个用户.位置隐藏是一种常用的隐私保护方法[5],在位置服务过程中,用户使用一个安全的粗粒度的隐藏区域(Cloaking Region, CR)替代精确的位置坐标,从而满足隐私的需求.同时,用户可以通过自身的偏好来设置应用环境,如空间 k-匿名(Spatial

收稿日期:2013-01-05;修回日期:2013-04-28;责任编辑:梅志强
基金项目:教育部博士点基金优先发展领域(No.20110121130006);国家自然科学基金(No.60970114, No.61103219, No.41104010);中央高校基本科研业务费专项资金(No.2011211020002)

k-Anonymity, SKA)[6]. SKA 是最为典型的位置隐私应用示例之一,其目标是在用户使用空间查询时保护用户的隐私身份. Jianguo H 提出了一种基于认证机制的隐私保护方法,通过使用组签名机制来实现车辆移动网络中匿名信息发布[7].

随着基于情景感知位置服务的发展,许多研究者提出了针对移动对象轨迹数据的隐私保护方法[8~10],这些方法可以较好的保证轨迹数据的隐私安全.其中一种比较典型的轨迹匿名模型,综合考虑了 k-匿名性和 α-多样性的隐私保护方法[8],以实现双重的隐私保护.借助于传统的隐私保护数据挖掘的方法对隐私保护研究具有重要的意义[9]. Yarovoy R 在隐私保护方法中考虑了基于位置的背景知识攻击[10]. 这些方法都较好的实现了离线轨迹数据的隐私保护.然而将其应用于轨迹收集隐私保护的过程中都无法满足客户端和数据收集方可信的要求[11].

面向位置数据发布环境下的轨迹 k-匿名(Trajectory k-Anonymity, TKA)[12]是另一种比较典型的示例. TKA 的目标是保护位置数据的敏感属性与特定用户的关联,有效的保护了用户的身份和位置隐私信息.然而 TKA 必须假设数据的收集和处理方是可信的,因此限制了其实用性. Gidófalvi, G 提出了隐私保护轨迹收集系统(Trajectory Collection System, TCS)[13]. TCS 不仅能有效的对抗基于位置的对手知识攻击,并且为收集者提供精确的轨迹数据.然而, TCS 对基于移动方式的对手知识攻击没有提出有效的保护方案,无法保证客户端安全.考虑这样一种情况,用户 u 在时间 t_A 的 CRs(CRs 表示多个 CR 区域)为 A,在下一步时间 t_B 的 CRs 为 $B(t_B > t_A)$.假设攻击者获得了 u 的移动方式,则可以推测出相应的最大移动速度为 v.攻击者通过 $v\cdot(t_B-t_A)$来计算 CR A 的 Minkowski[14]之和,并使用一个扩大的 CR A'与 B 的交集,来缩小或定位用户的隐私位置.

本文解决了 TCS 系统客户端的安全问题.首先通过形式化用户的移动轨迹,建立对手知识攻击模型,证明当攻击者获取用户的移动方式,以及用户的本地存储的样本轨迹的情况下,攻击者可以通过基于移动方式的背景知识攻击来推理用户的隐私信息.其次针对这种攻击提出了一种多维轨迹匿名方法,通过时空匿名算法,防止基于位置和移动方式的背景知识攻击.最后通过仿真实验检验该算法的有效性.为了给数据收集者提供精确的移动位置数据,避免预先设置的 CRs 泄漏隐私信息,本文的解决方案将以在线的方式生成 CRs 及匿名轨迹.

2 移动攻击模型

假设攻击者具有用户移动方式的背景知识,可以根据外部数据库来推测用户的敏感位置.本部分定义了相应的移动攻击模型,并形式化描述用户的隐私需求,最后给出用户确保隐私安全必须满足的安全条件.

2.1 匿名位置收集系统架构

隐私位置数据收集系统主要包含两个部分:基于中心服务器的服务端,以及基于在每一个移动手机上的客户端.客户端能通过移动连接协议如 SMS, MMS, GPRS 等与服务端相连,而每一个客户端间也能通过 P2P 网络互相连接.系统结构如图 1 所示.

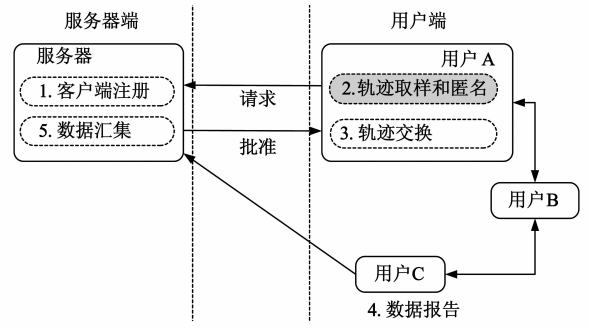

图1 匿名轨迹收集系统

用户通过注册确定轨迹采样周期和上传周期,并为采集的轨迹样本生成相应的假轨迹,以及相应的副本(真轨迹为奇数、假轨迹为偶数).多个用户通过 P2P 网络构造一个匿名组,组内的用户之间交换彼此的轨迹样本.所有组内用户在各自的上传周期结束之前上传本地的轨迹样本到服务器端.服务器根据轨迹副本的奇偶特征,筛选出真实的匿名轨迹.由于该系统采用了在线的轨迹收集方式,并且基于多用户协作的隐私保护模型,构造了 $k*k$ 的轨迹匿名矩阵,有效的消除了服务器不可信的隐私安全威胁.然而,将采集的轨迹样本以 k 匿名的形式存储在本地的同时,也为用户端的隐私安全带来了威胁.本文的工作主要集中在系统的轨迹采用和匿名部分,提出一种多维的轨迹匿名方法,改进了客户端的安全性.

2.2 形式化描述

设移动用户 u 的轨迹形式为: $T=\{(p_1,t_1),(p_2,t_2),\cdots,(p_n,t_n)\}$,其中, $p_i=(x_i,y_i)$ 是用户 u 在时间 t_i 时的二维位置坐标点.在给定的时间 t 时一个位置 snapshot 被定义为一个二元组 $S=(p,t)$. S 表示一系列位置 snapshots 集与用户轨迹相联合, $S=\{S_i\}_{1\leqslant i\leqslant n}$,如 $S_1=\{(p_1,t_1),(p_2,t_2),\cdots,(p_k,t_k)\}$和 $S_2=\{(p_{k+1},t_{k+1}),(p_{k+2},t_{k+2}),\cdots,(p_n,t_n)\}$为轨迹 S 的子片段.假设这些 snapshots 在时间上是连续的,即 $|t_{i+1}-t_i|=|t_i-t_{i-1}|$,用户以一定的时间间隔周期性的采取本地的位置数据样本.

从隐私保护的角度考虑,这些 snapshots 在他们原

来的形式中是没有被泄漏的.在每个时刻 t_i,用户根据相应的隐私设置在每一个 snapshots 周围生成 $k-1$ 个假点,从而构造一个匿名的隐藏区域 CR_i.用户本地位置样本的隐藏位置 snapshots 的子集标示为 $S'=\{(CR_i,t'_i)\}$,其中 $i\in[T_s,T_s+\lambda]$,T_s 为采样起始时间,λ 为采样周期长度.

2.3 攻击模型

本文假设用户已经具有一定的隐私保护能力,即攻击者无法在隐藏区域 CR 中获取敏感的位置信息[15].攻击者在只具有用户移动方式的背景知识的情况下,对用户的人身安全造成伤害.隐私保护的目的是防止用户精确位置信息泄漏.

设 CR A 和 CR B 分别为用户 u 在时间 t_1 和 t_2 的生成的位置点 $p_i,q_j,1\leqslant i,j\leqslant k$ 所构成的隐藏区域,k 为用户设定的隐私保护度,根据采样的连续性,不失一般性认为 $t_1<t_2,|t_1-t_2|=\lambda$,其中 $t_1,t_2\in[T_s,T_s+\lambda]$,$\lambda$ 为采样周期的长度.攻击者根据用户的移动方式推理出最大移动速度 v.

设用户 u 在隐藏区域 A 中的任一位置点为 $p_i,i\in k$.如图 2 所示.如果存在一个 $q_j,j\in k$,q_j 为用户在区域 B 中可能的位置点.即当攻击者利用获取到的最大的估计移动速度 v,使得 $d(p_i,q_j)>v\cdot\lambda$,则认为攻击者成功推理出用户的隐私位置.即:

$$\exists\, p_i\in A,\forall\, q_j\in B,d(p_i,q_j)>v\cdot\lambda$$

如图 2 所示,设用户的隐私保护度为 $k=5$,如果在隐藏区域 B 中存在点 q_1,通过递归的方法,逐步计算 $d(q_1,p_i)$,可以发现 p_5 为不可能的位置点,则削减了用户的隐私度,使得用户在 t_1 时刻的隐私匿名度为 $k'=4$.进一步计算 $d(q_1,p_i)$,发现 p_1,p_4 为不可能点,从而成功的推理出用户的可能位置为 p_2,p_3.

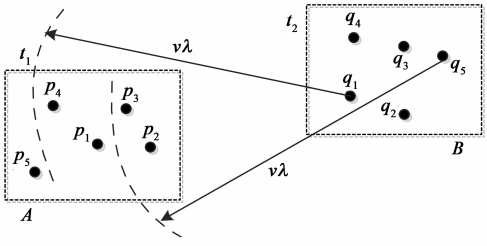

图2　移动速度攻击模型

因此,为了保护用户的隐私,则在生成用户的隐藏区域 CR 时,必须要考虑用户的移动速度的背景知识攻击,以保证隐私的安全性.

3　隐私保护算法

用户在采集真实轨迹样本的同时,按照相应的算法轨迹来生成 $k-1$ 条假轨迹,实现轨迹匿名的隐私保护.表 1 为假轨迹生成过程中所涉及变量的说明.

表 1　假轨迹生成过程相关变量声明表

变量	声明
k	轨迹匿名度
L	差异性
R	取值范围
v_i,v'_i	真位置点 i 及其对应假点的速度
(X_i,Y_i)	轨迹中的第 i 位置点($1\leqslant i\leqslant K$);
(X'_i,Y'_i)	由点 i 生成的第一个假点($1\leqslant i\leqslant K$);
o	辅助变量,取值为$\sqrt{L}$

本文以一个 $2D$ 平面来描述假轨迹的生成过程,所有轨迹点的生成均在该坐标轴平面中.例如 $k=3$,即由一条真轨迹生成两条假轨迹,算法设计如下:

(1)初始化.根据用户的采样周期 λ 将用户的轨迹片段 T 分割为多个连续的位置点.T 的定义为:$T=\{id,(x_1,y_1,t_1),(x_2,y_2,t_2),\cdots\cdots,(x_n,y_n,t_n)\}$其中 id 是该轨迹的标识符,$t_i(1\leqslant i\leqslant n)$为采样时间,$(x_i,y_i)$表示用户在 t_i 时刻的位置,称为采样位置或者采样点.则 $|t_{i+1}-t_i|=\lambda$,并且两点之间的距离 $d((x_i,y_i),(x_{i+1},y_{i+1}))=\lambda\cdot v_i$.

(2)空间匿名.通过设置最小隐私保护区域,从而使得所生成的假轨迹同时满足差异性和初始点的相似性.

取任意两个轨迹 T,T' 之间的最小边界矩阵为 AREA(MBR($\{T,T'\}$)),当最小边界矩阵满足 AREA(MBR($\{T,T'\}$))$>L$时,即只要这两条轨迹上同一时间相应两个位置点形成的最小隐藏区域(边界矩阵)面积大于 L,就达到了空间隐私保护的要求[16].因此,若$|X_1-X'_1|>o$ 与$|Y_1-Y'_1|>o$ 同时成立,则有:AREA(MBR(X_1,Y_1),(X'_1,Y'_1)) = $|X_1-X'_1|\,|Y_1-Y'_1|>L$ 其中,$[X_1-o,X_1+o]$与$[Y_1-o,Y_1+o]$为点(X'_1,Y'_1)的禁止取值区域.则首个假轨迹点(X'_1,Y'_1)的取值范围为:

$$\begin{aligned}&X'^2_1+Y'^2_1\leqslant R^2(R>0),\qquad \text{且}\\ &X'_1>X_1-o,\quad X'_1<X_1+o\\ &Y'_1>Y_1-o,\quad Y'_1<Y_1+o\end{aligned}\tag{1}$$

其中,R 为以(X_1,Y_1)为圆心的圆半径.

(3)时间匿名.当在真实的位置点周围生成相应的假位置点时,使得连续两个隐私区域内的位置点之间的最小距离小于隐私安全阈值.并且在移动速度上保持一定的相似性.

设(X'_2,Y'_2)为第一个假轨迹点(X_2,Y_2)的第一个后续点.后续位置点的取值范围除了需要满足上面的隐私禁止区域以外,还需要在移动速度上保持一致性,由参数 a,b 所决定.根据速度一致性原则:

$$a\left|X'_1-X_1\right|\leqslant\left|X'_2-X_2\right|\leqslant b\left|X'_1-X_1\right| \quad (2)$$

$$a\left|Y'_1-Y_1\right|\leqslant\left|Y'_2-Y_2\right|\leqslant b\left|Y'_1-Y_1\right| \quad (3)$$

式(2)、式(3)所设定的是第一条假轨迹与真轨迹之间的速度关系,同样对其他的假轨迹有效.

在第一个假位置点(X'_2,Y'_2)及其后续假位置点生成过程中,假设是在两个相邻的时间节点的情况下,在此单位时间内,设 Max_v,Min_v 分别为该取值的临界最大、最小值,即:

$$\text{Max\_}v=2R+b\cdot v\cdot d=2k\sqrt{L}+b\cdot v\cdot d \quad (4)$$

$$\text{Min\_}v=2\sqrt{L}+a\cdot v\cdot d \quad (5)$$

当满足式(4)、式(5)时,则认为达到了隐私保护的要求.其中 v 为用户最大可能移动速度,$b\cdot v$ 表示假轨迹能达到的理论最高速度,$a\cdot v$ 表示假轨迹能达到的理论最小速度,d 则表示真轨迹两点之间的距离.即当限定条件 $v\cdot\left|T_a-T_b\right|$ 在[Min_v,Max_v]的范围内时,所生成的轨迹点能够有效的防止移动攻击.

4 仿真实验

为了测试本文所提出的假轨迹生成算法的有效性.本次实验所处的硬件环境为:Inter(R) Core(TM)i5-2450M CPU,500G 硬盘,4G 内存,win7 64 位系统.实验数据来源为武汉大学自主研发的一套基于位置的网络信息自适应推送系统,其轨迹数据库共包含了约 1.4 万条位置记录.本文采集轨迹数据库中某段真实轨迹,经过多维轨迹匿名算法处理后得到的假轨迹.如图 3 所示.在以下实验过程中,使用一个常见的地图映射工具(如 Universal Transverse Mercator, UTM)将精确的几何位置(经度、纬度)通过椭圆曲线映射到笛卡尔坐标系统当中,将几何位置(精度、纬度)转换成坐标点.

(*a*) 真轨迹片段

(*b*) 假轨迹片段

图3

4.1 空间匿名

为了验证假轨迹生成算法的隐私保护效果,对轨迹的覆盖率和拥挤度进行了实验分析.通过将轨迹所在的区域划分成相同大小的网格.定义轨迹所在区域为 AS,区域内的每个网格为 Grid.并使用覆盖率和拥挤度来评价所生成的假轨迹的效果.覆盖率越高,攻击者通过轨迹获得用户隐私的可能性就越小.并且在同一个 Grid 中,拥挤度越高,安全性越好.

(1)覆盖率测试

实验包括一条真轨迹和数目不定的假轨迹,每条轨迹包括了 8 个位置点,在面积固定的 AS 范围内,分别对 Grid 为 4 * 4、6 * 6、8 * 8 并且假轨迹数目由 1 ~ 10 变化时的覆盖率进行了研究.结果如图 4 所示.

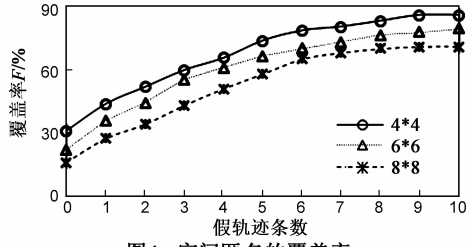

图4 空间匿名的覆盖率

实验结果显示,当 Grid 大小固定时,假轨迹的数目越多,覆盖率就越高,并且增加速度刚开始较快,随着假轨迹的增多而慢慢减缓.当假轨迹的数目固定时,Grid 面积越小,覆盖率越低.这说明本文所提出的假轨迹生成算法确实是符合真轨迹隐匿的需要,并不会随着轨迹数目的增多覆盖率一直增加,它始终是环绕在真轨迹附近,不会越出真轨迹相应的 AS;其次,假轨迹也并没有完全的参照真轨迹的行走路线,它和真轨迹保持着一定的差异性.

(2)拥挤度测试

在之前对拥挤度的描述中,真轨迹在某个 Grid 内的拥挤度能够说明真轨迹在该 Grid 内的匿名程度.对于真轨迹位置点所在的每一个 Grid 内,如果都存在至少 k-1 个假轨迹位置点,那么对于这条真轨迹来说,其总体匿名度不低于 k.拥挤度测试的目的在于,测试一条真轨迹至少需要多少条假轨迹同时进行混淆,才能

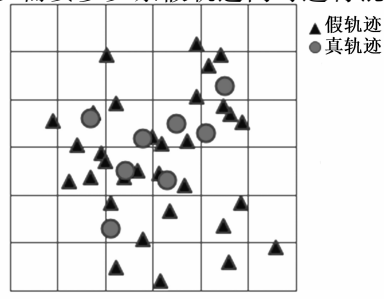

图5 空间匿名的拥挤度

够使得真轨迹所在的每一个区域都有其他假轨迹节点存在,保证达到匿名效果.实验在采用 1 条真轨迹、4 条假轨迹的情况下,拥挤度如图 5 所示.

如图所示,真轨迹位置点所在的 8 个 Grid 都同时存在至少 2 个假位置点.因此,在该情况下,用户的位置匿名度不低于 3,与设定的隐私保护度 $k=3$ 相等,达到了隐私保护要求.

4.2 时空匿名

(1)安全距离分析

为了验证时间匿名对抗移动速度攻击的有效性,本文分别在不同的移动速度下,通过计算两个相邻隐私区域中任意两点之间的最大距离,来分析其有效性.

实验环境保持不变,将 $v\cdot|T_a-T_b|$ 的值分别取 18,15,12. a, b 分别为 0.5 和 1.5.在取每一个值的情况下,分别取 20 次数据,取出下一个时间节点产生的假轨迹点到前一时间节点的所有位置点的最大距离 $D\_\max$ 与 $v\cdot|T_a-T_b|$ 的对应关系如表 2 所示.

表 2　相邻时间的位置点距离

$v\cdot\|T_a-T_b\|$	18	15	12
$D\_\max$	17.5688	14.0827	11.023

通过实验发现所产生的假轨迹点完全符合要求,可以有效的防止移动模型的攻击.

(2)覆盖率测试

为了讨论防止移动攻击的限制对产生的假轨迹的覆盖率的影响,对 Grid 的取值为 6 * 6,隐私匿名度取 $k=3$.在这两个条件固定的情况下,将记录 $v\cdot|T_a-T_b|$ 取值在等于 12,15,18,21,24,27 的几种情况下的覆盖率,实验结果如图 6 所示.

如图 6 所示,覆盖率 $F\%$ 随着速度限制的加大而有轻微的加大的现象,而当速度限制值逐渐增大时,覆盖率趋于稳定.其中原因在于当速度限制得比较紧的时候,点产生的位置则会相对集中些,所以对覆盖率造成了轻微的影响.而从总体上来说,随着速度限制的变化对覆盖率的变化没有太大的影响,覆盖率总体上保持一个稳定的数值的,在 6 * 6 的范围内覆盖率的变化在 20% 以下,满足了隐私保护的要求.

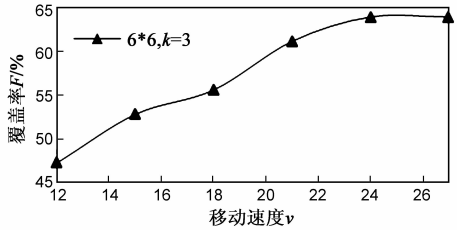

图6　时空匿名的覆盖率

(3)拥挤度测试

实验环境不变,加入移动限制的条件,分别对上述 6 种情况都做了拥挤度的评定,任意选出其中 $v=18$ 的情况进行说明.如图 7 所示.

从图 7 中可以清楚的发现,在每个拥有真轨迹点的 Gird 中,都有至少 2 个假轨迹点的存在,真轨迹的匿名度达到 3,与设定的隐私匿名度 $k=3$ 相等,满足隐私保护的需求.

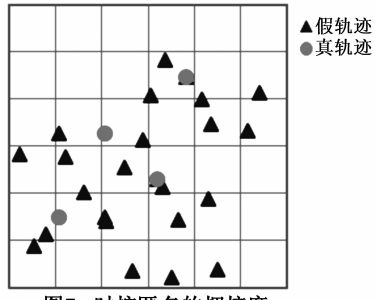

图7　时控匿名的拥挤度

5　相关工作对比

在线的轨迹收集隐私保护系统 TCS 是目前最为先进的轨迹隐私保护系统,本文选择该系统作为分析的对象,分别从轨迹收集过程中的隐私安全性以及算法复杂度方面进行了对比分析.由于本文的主要工作体现在用户端在采集轨迹样本后的匿名过程,因此,在安全性对比方面主要进行了安全距离和覆盖率方面的实验对比,而在拥挤度方面,通过 4.2 已经证明本文的方法满足了匿名要求.

(1)安全距离对比

为了验证时间匿名对抗移动速度攻击的有效性,本文分别运行 TCS 系统和本文提出的方法,计算两个相邻隐私区域中任意两点之间的最大距离.采用上一章中的实验环境,将 $v\cdot|T_a-T_b|$ 的值取 18, a, b 分别为 0.5 和 1.5,分别运行 20 次数据,取出下一个时间节点产生的假轨迹点到前一时间节点的所有位置点的最大距离,实验结果如图 8 所示.

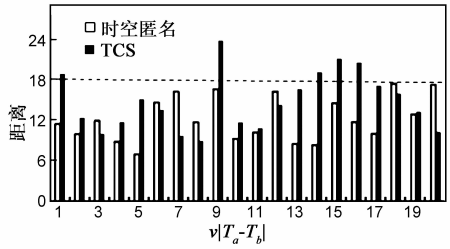

图8　相邻时间的位置点距离对比

通过实验发现,在不同的速度模式下分别计算 20 次,本文的方法均满足安全距离的要求,而 TCS 方法总体约有 25% 大于 18,不满足安全距离要求,实验说明本

文的方法更能够防止移动模型的攻击.

(2)覆盖率对比

覆盖率体现在一定区域内隐私保护的程度,覆盖率越高则隐私保护度越好.对 Grid 的取值为 $6*6$, $v\cdot|T_a-T_b|$ 取值为 18,隐私匿名度 k 取值为 0,1,…,10,分别运行 TCS 方法和本文提出的时空匿名方法的覆盖率进行了对比实验,如图 9 所示.

如图 9 所示,在不同匿名度 k 情况下,时空匿名算法的覆盖率 $F\%$ 相对于 TCS 方法下的覆盖率平均提升了约 22.3%,提高了隐私保护度.

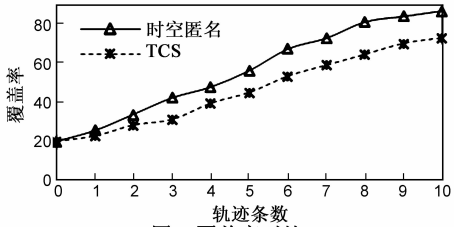

图9 覆盖率对比

(3)复杂度对比

设 n 为一条移动轨迹中位置点的个数,k 为生成的假轨迹的条数.在采用 TCS 方法生成假轨迹时,每一个移动轨迹点的周围都会生成对应的 k 个假轨迹点,其时间复杂度为 $O(kn)$.本文提出的时空轨迹匿名算法,采用了时空区域的限制范围内随机取点的方法,与 TCS 方法在生成假轨迹过程中的不同之处在于取点范围上有不同,因此本文提出的时空匿名算法的时间复杂度为 $O(kn)$.

通过以上的对比实验可以发现,本文提出的时空匿名方法在安全性优于 TCS 方法,满足了隐私安全的需求,有效的抵抗了基于位置和移动方式的背景知识攻击,在性能上具有较小的开销,具有较好的性能.

6 总结

本文基于轨迹数据收集系统,提出了一种多维轨迹匿名的个人隐私保护方法.该方法通过对用户位置数据的时间和空间匿名,防止攻击者利用对外部数据的推理而威胁用户的隐私安全.通过仿真实验检验,本文的方法能够有效的防止基于位置和移动方式的背景知识攻击.将该方法应用到匿名轨迹数据收集系统当中,可以进一步增强用户的隐私安全性.

参考文献

[1] 刘经南,郭迟,彭瑞卿.移动互联网时代的位置服务[J].中国计算机学会通讯,2011,7(12):40-50.
Jingnan Liu, Chi Guo, Ruiqing Peng. LBS in mobile internet [J]. Communication of the CCF, 2011, 7(12): 40 - 50. (in Chinese)

[2] Bonchi F, Lakshmanan L V S, Wang H (Wendy). Trajectory anonymity in publishing personal mobility data [J]. ACM Sigkdd Explorations Newsletter, 2011, 13(1): 30 - 42.

[3] Freudiger J, Shokri R, Hubaux J P. Evaluating the privacy risk of location-based services[A]. Financial Cryptography and Data Security[C]. Berlin: Springer, 2012. 31 - 46.

[4] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking[A]. 1st International Conference on Mobile Systems, Applications and Services [C]. New York: ACM, 2003. 31 - 42.

[5] Damiani M L, Bertino E, Silvestri C. Protecting location privacy against spatial inferences [A]. 2nd SIGSPATIAL ACM GIS 2009 International Workshop on Security and Privacy in GIS and LBS SPRINGL 09[C]. New York: ACM, 2009. 32 - 41.

[6] Kalnis P, Ghinita G, Mouratidis K, et al. Preventing location-based identity inference in anonymous spatial queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19 (12): 1719 - 1733.

[7] Jianguo H, Yiqi D. Achieving controllable privacy protection in position service for VANETs[J]. Chinese Journal of Electronics, 2011, 20(3): 395 - 400.

[8] Machanavajjhala A, Kifer D, Gehrke J, et al. L-diversity: Privacy beyond k-anonymity[J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1): 3 - 55.

[9] 杨高明,杨静,张健沛.聚类的(α,k)-匿名数据发布[J].电子学报,2011,39(8):1941 - 1946.
Yang Gaomin, Yang Jing, Zhang Jianpei. Achieveing (α, k)-anonymity via clustering in data publishing[J]. Acta Electronica Sinica, 2011, 39(8): 1941 - 1946. (in Chinese)

[10] Yarovoy R, Bonchi F, Lakshmanan L V S, et al. Anonymizing moving objects: how to hide a MOB in a crowd? [A]. The 12th International Conference on Extending Database Technology: Advances in Database Technology[C]. Saint-Petersburg: ACM, 2009. 72 - 83.

[11] Gidófalvi G, Huang X, Pedersen T B. Privacy-preserving data mining on moving object trajectories[A]. 2007 International Conference on Mobile Data Management[C]. Washington: ACM, 2007. 60 - 68.

[12] Nergiz M E, Atzori M, Saygin Y. Towards trajectory anonymization: A generalization-based approach[J]. ACM Transactions on Data Privacy, 2009, 2(1): 47 - 75.

[13] Gidófalvi G, Huang X, Pedersen T B. Privacy Preserving trajectory collection[A]. The 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems[C]. New York: ACM, 2008. 1 - 4.

[14] De B M, Cheong O, Van K M. Computational Geometry: Algorithms and Applications[M]. CA, USA: Springer, 2008.

[15] Gedik B, Liu L. Location privacy in mobile systems: A per-

sonalized anonymization model[A]. the 25th IEEE International Conference on Distributed Computing Systems[C]. Columbus, OH, USA: IEEE, 2005. 620 – 629.

[16] Ghinita G, Kalnis P, Khoshgozaran A, et al. Private queries in location based services: anonymizers are not necessary[A]. The 2008 ACM SIGMOD International Conference on Management of Data[C]. New York, NY, USA: ACM, 2008. 121 – 132.

作者简介

王丽娜　女. 1964 年 10 月出生,辽宁营口人. 教授、博士生导师. 1989 年于东北大学获硕士学位. 2001 年于东北大学获得博士学位. 现为武汉大学计算机学院副院长,信息安全教育部重点实验室主任,主要从事隐私保护、多媒体安全等方面的研究工作.
E-mail: lnawang@163.com

彭瑞卿　男. 1985 年 2 月出生,湖北广水人. 2010 年于三峡大学获硕士学位. 2010 年进入武汉大学计算机学院信息安全专业. 现为博士研究生,从事位置隐私保护、网络安全等方面的研究工作.
E-mail: rqp1985@gmail.com

赵雨辰　男. 1988 年 10 月出生. 辽宁鞍山人. 2011 年于华中科技大学获得学士学位. 2011 年进入武汉大学计算机学院信息安全专业. 现为硕士研究生,主要从事位置隐私保护方面的研究工作.

陈　栋　男. 1992 年 10 月出生. 河南信阳人. 2012 年于湖南科技大学获得学士学位. 2012 年进入武汉大学计算机学院信息安全专业. 现为硕士研究生,主要从事位置隐私保护方面的研究工作.