文章编号：2095-2163(2019)05-0250-04 中图分类号：TP309.2 文献标志码：A

一种基于K匿名技术在轨迹隐私保护方法中的改进

张家磊，钟伯成，房保纲，丁佳蓉，贾媛媛
（上海工程技术大学 电子电气工程学院，上海 201620）

摘 要： 针对K匿名技术在轨迹隐私保护中存在搜索 $k-1$ 个匿名用户耗时过长、甚至搜索不到和易于被连续查询攻击所侵袭的现状，提出了一种改进的方法，方法利用时间截点对搜索时间进行约束，生成假名对用户发送查询请求时的真名进行保护，有效解决了上述问题。实验结果表明该方法的系统运行时间较短且隐私保护程度较高。

关键词： K匿名技术；轨迹隐私；时间截点；假名

An improvement of track privacy protection method based on K-anonymity technology

ZHANG Jialei, ZHONG Bocheng, FANG Baogang, DING Jiarong, JIA Yuanyuan
(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

【Abstract】 Aiming at the existence of K-anonymity technology in the trajectory privacy protection, the search for $k-1$ anonymous users is too long, even unsearchable and easy to be attacked by continuous query attacks. An improved method is proposed, which uses time interception pairs. The search time is constrained, and the pseudonym is generated to protect the real name when the user sends the query request. The experimental results show that the system has a shorter running time and a higher degree of privacy protection.

【Key words】 K-anonymity technology; trajectory privacy; time cut-off point; kana

0 引 言

随着智能设备的广泛普及，越来越多的用户开始使用基于位置的服务（Location Based Service，LBS）[1]，例如通过腾讯地图、高德地图等软件查询附近感兴趣的地点（酒店、超市、美食）。通过LBS服务，用户可以得到一条到达目的地路程最短的轨迹，大大方便了用户的出行需求。只是，一旦用户的轨迹信息被敌手所捕获，用户的家庭住址、工作地址、经常出入的场所等等个人隐私信息也都面临着泄露的风险，所以如何保护LBS服务中用户的轨迹信息已然成为学界的研究热点。

针对LBS服务中用户的轨迹隐私安全，目前的轨迹隐私保护技术主要有3种[2]：假轨迹法、泛化法、抑制法。这3种方法各有优缺点。其中，假轨迹法通过添加假轨迹来对用户的真实轨迹进行干扰，起着混淆敌手的作用，从而达到保护效果。这种方法计算简单，并且计算量相对较小，但是会降低数据的可用性。泛化法对轨迹上的点泛化为相应的匿名区进行保护，从而不能精确地定义用户位置[3-4]，这种方法采用的都是真实数据，但是其计算量太大。抑制法就是通过不对外发布轨迹上用户敏感位置或频繁访问位置的方式来保护用户的轨迹隐私安全，这种方法在实现上较为简单，但是容易造成用户数据的丢失。

近年来通过研究发现，基于泛化法的轨迹K匿名[5]技术在用户的隐私保护程度和数据的可用性上取得了一个较好的平衡，是目前轨迹隐私保护中较为常用的方法。其基本思想是：在真实用户周围寻找 $k-1$ 个其它用户的匿名框，并发送给LBS服务器进行查询，这样就能产生出混淆效果，使敌手无法辨别出真实用户，从而达到保护用户轨迹隐私的目的。但是传统K匿名技术在连续查询攻击下，敌手可以根据匿名区内一直存在的用户来推断出查询用户。例如，用户在3个不同时刻的匿名集分别为{S、A、B、C、D}、{S、A、B、E、F}、{S、G、H、L、J}，那么敌手就能通过对这3个连续的匿名集作分析得到真实的用户为S。而且在人群比较稀少的区域，$k-1$ 个匿名用户可能面临较大搜索难度，甚至搜索失败的情况。

作者简介：张家磊（1995-），男，硕士研究生，主要研究方向：网络安全、隐私保护；钟伯成（1964-），男，博士，教授，主要研究方向：计算机网络、网络拥塞控制。
通讯作者：张家磊 Email：835957313@qq.com

收稿日期：2019-07-27

针对上述问题，本文基于传统的 K 匿名技术做出了改进，提出了一种新的轨迹隐私保护方案。本文的主要贡献如下：

(1)提出一个时间截点 T，在一定时间截点 T 内如果搜索不到 $k-1$ 个匿名用户，则产生虚假位置来代替剩下难以搜索到的匿名用户，这样就大大节省了计算开销，也防止了在人群稀少的区域无法寻找到 $k-1$ 个匿名用户的情况。

(2)每次向 LBS 发送服务请求时，第三方匿名服务器生成不同假名代替真实用户发送给 LBS 服务器进行查询操作，有效地防止了连续查询攻击。

1　系统模型及相关定义

1.1　系统模型

由于轨迹隐私保护方案需要大量的计算资源，让资源有限的移动设备承担所有的计算开销显然不可能，因此本文采用第三方可信匿名服务器的系统模型，该系统模型包括用户、第三方可信匿名服务器和 LBS 服务器三个组成部分。系统结构如图 1 所示。由图 1 可知，用户是指携带具有无线定位、无线通信功能的移动智能设备，并向服务器连续发送请求的人们。第三方可信匿名服务器建立在可信的服务平台上，主要由历史数据库、匿名模块和结果处理模块组成。历史数据库保存用户的历史轨迹和请求概率；匿名模块对用户发送过来的位置和查询信息进行模糊化操作，根据用户的隐私保护需求将用户的真实位置混淆在匿名区域内一起发送给 LBS 服务器进行查询；数据处理模块负责对 LBS 发送过来的查询结果进行精确计算，将用户的查询结果筛选出来返回给用户。LBS 服务器能够利用数据库为用户搜索到用户感兴趣的点，例如，用户附近的酒店、电影院等。

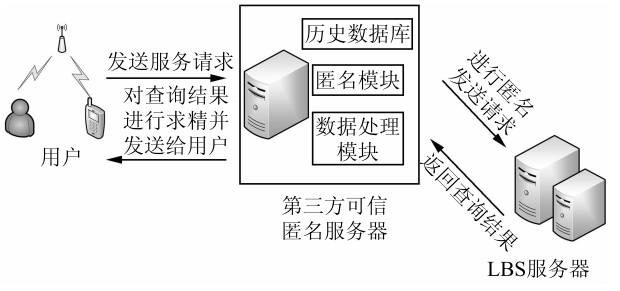

图 1　系统结构

Fig. 1　System structure

1.2　相关定义

定义 1　请求概率 q　假设把区域划分成 $n \times n$ 个网格，则每个网格中向 LBS 发送查询的请求概率为：

$$q_{(x,y)} = \frac{\text{网格}(x,y)\ \text{中发送查询请求的个数}}{\text{所有网格中发送查询请求的个数}}, \quad (1)$$

在此基础上，设定一个阈值 β，用来限制匿名用户与真实用户的请求概率相差过大，排除掉不合理的假位置点，例如湖泊、河流等不可能发送查询请求的位置点。

定义 2　距离偏差 D　这个参数表示假轨迹片段与用户真实轨迹片段在速度上的差异性。假设时刻 i 用户的位置为 P_i，上一个位置为 P_{i-1}，这两个位置在时刻 i 对应的假位置分别为 F_i 和 F_{i-1}，则距离偏差为：

$$D = |\ dist(P_i, P_{i-1}) - dist(F_i, F_{i-1})\ |. \quad (2)$$

定义 3　轨迹局部角度偏差 θ　这个参数表示假轨迹在某个时刻与用户真实轨迹的相似情况，即表示在某个时刻，假轨迹与用户真实轨迹之间的角度的偏差。假设时刻 i 用户的位置为 P_i，上一个位置为 P_{i-1}，与之对应的假轨迹片段上的位置点为 F_i 和 F_{i-1}，那么轨迹局部角度偏差 θ 为这两个向量之间夹角。

定义 4　时间截点 T　这个参数用来限制第三方服务器在搜索 $k-1$ 个匿名用户上花费过多的时间。

2　假轨迹生成算法

2.1　算法思路

算法开始时，用户首先设置隐私保护参数 (β, D, θ, T, k)，第三方可信匿名服务器根据用户的位置和隐私保护参数在用户的位置周围寻找 $k-1$ 个与查询用户的请求概率相差小于 β 匿名用户，如果在时间 T 内没有寻找到，则产生假位置来代替剩下难以寻找到的匿名用户，同时将生成一个假名代替查询用户向 LBS 服务器发送查询请求。在接下来的查询时刻，第三方可信匿名服务器继续按初始位置的规则生成假位置，只是不同时刻生成的假名要求不同，并且需要按照距离偏差、轨迹局部角度偏差两个约束，对匿名用户的位置或生成的假位置进行一定的约束，直到生成满足条件的 $k-1$ 条假轨迹。

2.2　算法步骤

输入：用户的位置，查询请求，隐私保护参数：β, D, θ, T, k

输出：假轨迹集

Step 1　在查询用户的初始位置搜索与查询用户请求概率差值小于 β 的 $k-1$ 个匿名用户，如果时

间 T 内仍搜索不到,则产生假位置代替剩下 n 个难以搜索到的匿名用户。

Step 2 第三方可信匿名服务随机生成一个假名代替查询用户向 LBS 发送这条查询信息。

Step 3 当用户开始移动、即用户不在初始位置时,将重复 Step 1 ~ Step 2 的操作过程。

Step 4 计算 2 个匿名区域内各位置点之间的欧式距离和与真实轨迹片段之间的角度偏差,把满足这 2 个约束条件的位置点连接起来生成假轨迹片段。

Step 5 重复 Step 3 ~ Step 4,直到在原来的轨迹片段上产生满足要求的 $k-1$ 条完整的假轨迹。

3 实验结果分析

本次实验中的数据集是由 Thomas Brinkhoff 移动对象生成器基于德国古登堡市的地图生成的。本实验所有的代码都是由 Java 编写的,实验环境为 Windows 10 操作系统,内存空间 8 GB,其处理器为英特尔 Core i7-7700HQ。编程环境为 MyEclipse10。实验中的参数设置为:$\beta=0.2$, $D=50$ m, $\theta=15°$, $T=0.5$ s。

3.1 连续查询攻击分析

敌手通过分析不同时刻匿名区域内重复出现的匿名用户,就能得到真实的查询用户,但是本方案在不同时刻利用假名来代替真实用户,使真实用户不会出现在各个时间点的匿名区域内,从而有效防止了连续查询攻击。例如,在传统的 K 匿名算法中用户在 3 个不同时刻的匿名集分别为{S、A、B、C、D}、{S、A、B、E、F}、{S、G、H、L、J},敌手可以对这 3 个匿名集做交集就能得到真实的查询用户 S,但是在本文提出的算法中,第三方匿名服务器在这 3 个时刻随机提供 3 个不同假名 V、M、Z 来代替真实的查询用户 S 向 LBS 服务器发送查询请求,从而有效避免了连续查询攻击。

3.2 运行时间分析

研究可得,不同算法下的系统运行时间的对比绘制结果如图 2 所示。由图 2 可以看出,随着匿名区域内数量 k 的增加,本文算法和传统的 K 匿名算法的系统运行时间都在增加。但是在 $k>15$ 以后,本文算法的系统运行时间明显优于传统的 K 匿名算法,这是因为当需要搜索的匿名用户数量 k 达到一定数量时,会扩大服务器的搜索范围,增加了计算开销,从而系统的运行时间也会相应地增加。

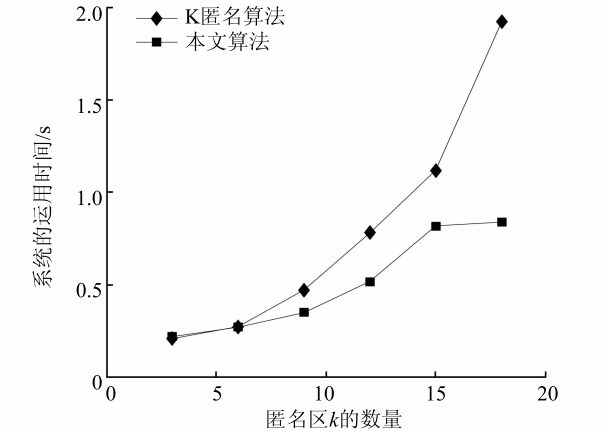

图 2 运用时间分析

Fig. 2 Run time analysis

3.3 隐私保护程度分析

对 2 种算法在隐私保护程度上进行了对比,对比结果如图 3 所示。由图 3 可以看出,本文算法在隐私保护程度上优于传统的 K 匿名方法,且保护程度较高。

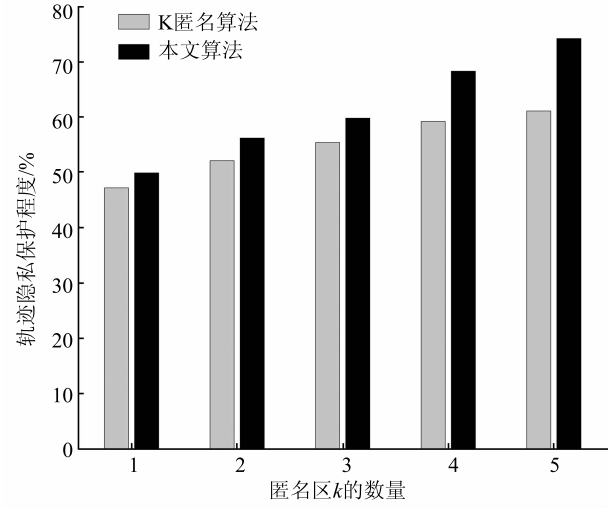

图 3 隐私保护程度分析

Fig. 3 Analysis of privacy protection degree

4 结束语

本文针对传统的 K 匿名技术在轨迹隐私保护中的不足,提出了一种改进方法,利用时间截点和假名有效解决了传统 K 匿名技术人群较少的区域计算开销过大和不能防止连续查询攻击的问题。实验表明,本文提出的方法安全程度更高,系统运行时间更少。

(下转第 256 页)