基于连续查询的用户轨迹 k-匿名隐私保护算法

作者	胡德敏，郑霞
机构	上海理工大学 光电信息与计算机工程学院；上海理工大学 计算机软件技术研究所

发表期刊	《计算机应用研究》
预排期卷	2017 年第 34 卷第 11 期
访问地址	http://www.arocmag.com/article/02-2017-11-001.html
发布日期	2016-11-28 14:47:25
引用格式	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[J/OL]. [2016-11-28]. http://www.arocmag.com/article/02-2017-11-001.html.

摘要	随着移动服务和移动网络的持续发展，基于 LBS 的连续查询服务被广泛应用。基于单点的 K-匿名位置隐私保护算法已经不能满足连续查询下用户位置隐私需求。针对用户轨迹隐私保护提出新的保护方法，该方法采用不可信第三方中心匿名器，用户获取自己的真实位置后首先在客户端进行模糊处理，然后提交给第三方匿名器,第三方匿名器根据用户的隐私需求结合用户某时刻的模糊位置信息生成虚假用户，然后根据历史数据生成虚假轨迹。为了进一步提高虚假轨迹与用户真实轨迹的相似性，该算法提出了虚假轨迹生成的两个约束条件：虚假轨迹距用户真实轨迹…

关键词	连续查询, k-匿名，基于位置服务，用户轨迹隐私保护，位置隐私保护
中图分类号	TP309.2
基金项目	国家自然科学基金项目（61170277）；国家自然科学基金项目（61472256）；上海市教委科研创新重点项目（12zz137）；上海市一流学科建设项目（S1201YLXK）

基于连续查询的用户轨迹 k-匿名隐私保护算法 *

胡德敏 [a, b]，郑 霞 [a]

(上海理工大学 a. 光电信息与计算机工程学院；b. 计算机软件技术研究所，上海 200093)

摘 要：随着移动服务和移动网络的持续发展，基于 LBS 的连续查询服务被广泛应用。基于单点的 K-匿名位置隐私保护算法已经不能满足连续查询下用户位置隐私需求。针对用户轨迹隐私保护提出新的保护方法，该方法采用不可信第三方中心匿名器，用户获取自己的真实位置后首先在客户端进行模糊处理，然后提交给第三方匿名器,第三方匿名器根据用户的隐私需求结合用户某时刻的模糊位置信息生成虚假用户，然后根据历史数据生成虚假轨迹。为了进一步提高虚假轨迹与用户真实轨迹的相似性，该算法提出了虚假轨迹生成的两个约束条件：虚假轨迹距用户真实轨迹的距离约束和相似性约束。经大量实验证明，该算法与不同时刻 K-匿名算法相比，不仅可以满足连续查询的用户轨迹隐私保护而且可以满足基于快照的 LBS 用户位置隐私保护。

关键词：连续查询；k-匿名；基于位置服务；用户轨迹隐私保护；位置隐私保护

中图分类号：TP309.2

K-anonymous privacy protection algorithm for user trajectory protection based on continuous query

Hu Demin[a, b], Zheng Xi[a]

(a. School of Optical-Electrical & Computer Engineering, b. Institute of Computer Technology, University of Shanghai for Science & Technology, Shanghai 200093, China)

Abstract: With continued advances in mobile service and mobile Internet,Based on the continuous query service of LBS is widely used. in our daily life. For continuous query, a new protection method is proposed for the user trajectory privacy protection. This method without trusted third party server, Users get real information about their location and blurred location information on the client side, and then submitted to the third party server. The third-party service in the user location to generate true false user location according to the user's personalized privacy requirements,Then according to the historical data generated into a false track. In order to further improve the similarity between the false track and the user's true trajectory, the proposed algorithm proposes two constraints: the distance constraint of the false track and the real trajectory of the user and similarity constraints. Compared with the traditional K-anonymous algorithm, the proposed algorithm not only can satisfy the user's trajectory privacy preserving of continuous query, but also can satisfy the user's location privacy protection based on snapshot LBS.

Key Words: Continuous query; k-anonymous;Location-base service;user trajectory protection;location privacy protection

0 引言

随着全球定位技术和无线通信的快速发展不仅给基于位置服务创造了更大的发展空间，也给基于位置的移动应用程序创造了新的机遇，尤其是嵌入基于位置服务功能的应用增长迅速（如：聊天工具、手机地图、移动手游等应用），给人们日常生活带来了极大的便捷。例如：一个移动端用户在陌生的环境下寻找附近的地铁站、娱乐场所、住房等等。但随着用户的安全意识越来越高，用户在享受基于位置服务的同时，也担心自己提交给 LBS 服务器的位置信息被泄露或被不法分子获取,造成不必要的损失。

根据查询方式不同 LBS 分为快照 LBS 和连续 LBS。快照 LBS 是指用户只需向 LBS 服务器提供一次位置信息，LBS 服务器根据用户的位置和搜索内容提供基于位置服务；连续 LBS 是指用户按照一定频率将自己的位置信息周期性发送给 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

基金项目：国家自然科学基金项目（61170277）；国家自然科学基金项目（61472256）；上海市教委科研创新重点项目（12zz137）；上海市一流学科建设项目（S1201YLXK）

作者简介：胡德敏（1963-），男，上海人，副教授，博士，主要研究方向为计算机网络、分布式计算、云计算；郑霞（1990-），女，山东莱西人，硕士，主要研究方向为位置隐私保护.

将最新的结果返回给用户。相比快照 LBS 位置隐私保护，连续 LBS 位置隐私保护更具有挑战性。

1　相关工作

关于连续 LBS 位置隐私保护的问题已经引起很多用户和研究者的关注，但针对快照 LBS 的经典 K-匿名方法不能直接应用于连续 LBS 用户轨迹隐私保护，因为攻击者可以根据用户每个时刻的 K-匿名区，推断出用户的真实位置。例如，用户在三个不同时刻 K 匿名区域用集合表示分别是{A、B、C、D、E}{A、B、G、F、H、I}{A、C、D、H}三个集合，攻击者很容易根据用户的匿名轨迹推断出发送 LBS 请求的用户是 A。

目前，对于连续 LBS 的用户轨迹隐私保护已经引起很多学者的关注。学者 Huang 等提出了 Silent Period 方法[1,2]保护用户轨迹隐私，该方法将用户的活动区域划分为混合区和应用区，当用户所在位置属于混合区时，不会向 LBS 服务器发送请求信息，当移动用户离开或进入混合区时会再次更换自己的假名，由此保护用户的轨迹信息，但由于混合区之间没有通信会导致服务质量降低。学者 Shi Minyi 在文献[3]中将用户活动区域划分敏感等级并设置假名的生命周期，当假名的有效期结束或者用户进入或离开敏感区域时都将更换假名并借用脚印辅助形成匿名区保护用户轨迹隐私；palanisamy 等人发现[4]在欧式空间里攻击者可以用移动轨迹获取用户的位置信息，甚至很容易根据路网环境的限制获取用户轨迹信息。为了保护路网环境下的用户位置隐私，研究人员 Wang 在文献[5]提出了路段 L-多样性，用户位置隐私保护在满足用户的 K 匿名的同时，匿名区至少包含 L 条不同的路段。

目前大多数位置隐私保护方法只是针对快照 LBS 或者是连续 LBS 的用户位置隐私保护方法。为了同时满足快照 LBS 和连续 LBS 用户位置隐私保护，本文基于连续 LBS 查询提出基于轨迹位置隐私保护算法。该方法采用不可信第三方中心匿名器，用户的真实位置首先在客户端进行模糊处理提交给第三方匿名器，第三方匿名器根据用户的个性化隐私需求生成虚假用户在某时刻的 K 匿名框，并结合两个约束条件然后生成虚假轨迹。本文提出的用户轨迹隐私保护方法不仅适合连续 LBS 查询而且适合快照 LBS 查询。

2　基于虚假轨迹的 *k*-匿名方法

2.1　相关定义

定义 1　用户隐私需求 UQL。

用户隐私需求 UQL={K,lmin,lmax}

其中 K 表示用户匿名区至少包含 K 个移动用户;Lmin 表示匿名区的最小半径；Lmax 表示匿名区的最大半径，该属性为了保护由于匿名区过大，导致 LBS 服务器的负载过大。用户隐私模型中所有的隐私需求都是由用户请求服务时提前设定的。

定义 2　用户请求内容。

Content={ID,first,T,GID,content,UQL,Speed}

其中 ID 代表用户的唯一标识符;first 代表用户是否是第一次发送请求;T 表示用户发出该请求的时间;GID 代表用户所在单元格;content 表示用户请求 LBS 服务器的内容;UQL 代表用户的隐私需求;Speed 代表查询 LBS 用户的移动速度。

定义 3　轨迹偏移度 TD。

$$TD=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$$

轨迹偏移度是指在 T 时刻假轨迹上的点距离用户真轨迹的点的距离。假轨迹距用户真实轨迹的距离计算方法目前常用的有：欧几里德、最长同序列距离、编辑距离、对数距离等等[6,7]。本算法采用最常用的欧几里德的计算方法。

2.2　系统结构

k-匿名算法采用第三方可信匿名器，第三方可信匿名器根据用户的真实位置搜索 K 个用户（其中一个是真实用户），构成匿名区发送给 LBS 服务器请求服务[8]。但在实际生活中，并不可能存在完全可靠的第三方匿名器，一旦第三方可信匿名器被攻击，用户的位置隐私就会被泄露[9]。本文提出的基于连续 LBS 用户轨迹隐私保护算法采用不可信第三方服务器体系结构，主要包括：移动用户，第三方匿名器(简称 CLSP)和 LBS 服务器（如图 1 所示），第三方匿名器包含了用户隐私处理模块、查询结果处理模块和存储用户历史轨迹三大模块。

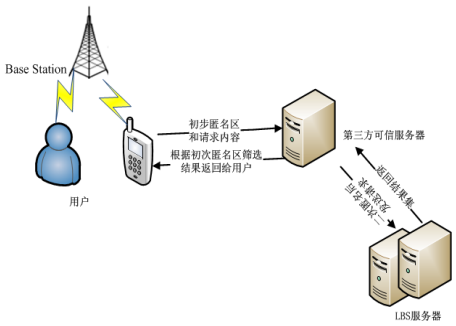

图 1 体系结构图

2.3　虚假轨迹生成的约束条件

用户端将初次的匿名区域（具体见下面用户端初步匿名算法）以及用户的位置隐私需求发送到第三方中心匿名器，中心匿名器基于用户端的位置隐私需求中的 K，Lmin，Lmax 生成某个时刻轨迹匿名区。为了进一步加强连续查询的位置隐私保护，虚拟轨迹某时刻的位置信息不能随机生成，必须遵循用户的真实轨迹信息。因此，虚假轨迹的约束条件如下：

1）虚拟轨迹距用户真实轨迹的距离(DC)

用户在请求连续 LBS 服务时，先将初次的匿名区域以及用户的隐私需求发送到中心匿名器，中心匿名器根据用户的隐私需求 UQL={K,lmin，lmax}中的 lmin，lmax，K 生成匿名区，匿名区分别以 lmin，lmax 为半径生成的圆环内随机生成 K 个用户，并以 K 个用户的所在的最小的正方形作为 K-匿名区这样既可以满足用户的隐私需求，而且不会因为随机生成的虚假位置过近或过远，导致 LBS 服务器的负载过大。

2）相似原则(DSC)

虚假轨迹的位置信息不仅要满足一定的距离约束，同时也要再方向以及速度上也与真实的轨迹信息具有一定的相似性。

2.4 用户端初步匿名算法

为了防止第三方匿名器被黑客攻击或者故意泄露用户隐私，该算法采用第三方服务器体系结构，此处假设第三方匿名器不可信。因此用户端向中心匿名器发送位置时，不能发送用户的真实位置，要在用户端首先进行模糊处理。该算法的主要思想如下：

a)用户通过定位技术获取自己所在的区域。

b)将用户所在区域进行网格划分，每个网格都有唯一的标志 GID，用户将自己所在的当前网格代替用户的真实位置发送给第三方匿名器请求服务。

c)用户以一定频率获取自己的位置信息，查看

自己当前位置是否还在该网格内。

d)如果用户获取的位置信息后，发现自己还在当前网格，不继续提交当前位置，当用户发现自己

2.5 基于轨迹的连续 LBS 位置隐私保护算法

用户将自己的初步匿名区、查询内容以及用户位置隐私需求提交给中心匿名器，如果该用户用户是第一次发起连续查询请求会将查询内容中的 first 属性设置为 1，否则设置成其他任意属性。中心匿名器根据用户请求内容中的 first 属性判断用户是否是第一次请求连续查询，如果该用户是基于连续 LBS 查询中的第一次请求则执行算法一，否则执行算法二。基于轨迹的连续 LBS 位置隐私保护算法的伪代码如下：

输入：用户隐私需求 UQL,查询内容 content，GID 用户初步匿名区

输出：根据用户的个性化隐私需求结合约束条件生成虚假轨迹。

1. 调用用户端初步匿名

2. Send Content={ID,first,T,GID,content,UQL} to LSP

// 用户请求内容、隐私需求、初步匿名区发送给中心匿名器

3. If (first==1) GO Algorithm 1//中心服务器根据 first 属性判断该用户是否是第一次请求连续服务，是执行算法一

Else GO Algorithm 2 //否则执行算法 2

算法 1 的中心思想：用户通过某种定位技术获取自己的当前位置，并根据用户提供的参数 first=1 判断出用户是第一次发送请求至中心匿名器。然后根据用户的隐私需求 UQL={K,lmin，lmax}，根据虚拟轨迹距用户真实轨迹的距离的约束条件生成匿名区，匿名区的范围为以用户所在单元格的中心点为圆心，分别以 lmin，lmax 为半径生成的圆环内随机生成 K 个用户，以这 K 个用户所在的最小的正方形作为 K-匿名区。这样攻击者也无法根据真实用户是圆心推断出用户的真实位置。（此处省略伪代码）例如：用户的个性化隐私需求为 lmin=5,lmax=8,K=4 时，生成匿名区的最后结果如图 2 所示。

中心匿名器根据用户的个性化隐私需求形成匿名区，并由中心匿名器中的存储用户历史轨迹模块储存该用户的信息，然后以该匿名区为用户的真实位置发送到 LBS 服务器，LBS 服务器根据匿名区域进行内容查询，然后将查询结果发送给中心匿名器，中心匿名器对查询结果进行过滤，将结果返回给用户。

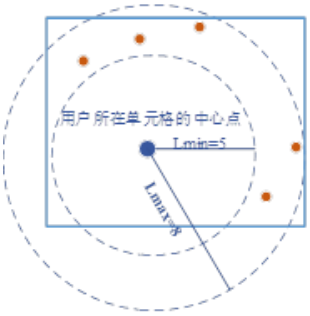

图 2 用户匿名区生成示例图

算法 2 的中心思想：当用户在时刻 t=1.....n 的时刻的匿名区中假用户的生成，此时不仅仅要考虑用户距离的约束条件，还要考虑方向和速度相似性的原则以及时间的有效性。算法如下：

1.If (first !=1) //该用户在 t1,t2.....tn 等时刻发送的查询请求。

2．{

3.Send Content={ID,first,T,GID,content,UQL} to CLSP

//将用户将内容及所在单元格发送给第三方服务器

4. Random(user) between in lmin and lmax

5. If(User satisfy DSC && K)

//判断虚假用户满足相似性原则 DSC,以及虚假用户是否等于 K

6. {TO step 8；}//满足相似性原则 DSC,及虚假用户为 K

7. Else{ TO step 4} //假用户不符合要求，继续生成新的虚假用户

8. FT=(Bi,Bi+1) //中心匿名器中根据历史记录模块，生成轨迹

9. TT=(Ai,Ai+1) //生成用户的真实轨迹

10. 如果 1<first<n+1 时，继续执行上述语句，更新虚假轨迹的信息

11. 根据上述的方法将用户连接成虚假轨迹，即可

3 实验结果与分析

本次实验采用 Thomas Brinkhoff 路网数据生成器[10]并分别在 oldenburg 城市的交通路网生成移动对象数据。本次实验所用到的算法及程序都是用 JAVA 编写。本课题提出的基于连续查询的用户轨迹隐私保护算法（简称 CPP）的有效性，对该算法的响应时间及隐私保护程度、精确度多方面进行验证。首先介绍了仿真实验的运行环境以及运行参数的设置。具体的实验参数如表 1 所示。

表 1 实验参数

参数名称	设置
K	[3,8]
Lmin	80 米
Lmax	120 米
用户速度 Speed	20m/s
方向相似性	15%

为了验证该算法的有效性，本实验首先从该算法的系统响应时间、位置隐私保护度、与不同时刻 K-匿名算法相比较。将

用户所在区域划分成 180*180 个单元格。将基于连续查询的用户轨迹保护方法从响应时间和用户轨迹隐私保护度两方面与不同时刻 K-匿名算法（以下简称 TK）相比较。比较结果如图 3 和图 4 所示。

为了验证 CPP 算法的有效性，首先从系统响应时间这方面与不同时刻 K-匿名算法（TK）进行相比较。本次实验分别为 6 个用户设置 K 为 3，4，5，6，7，8 不同的匿名隐私需求，此次实验每个用户分别发送 20 次连续查询请求。实验的最终比较结果如图 3 所示。

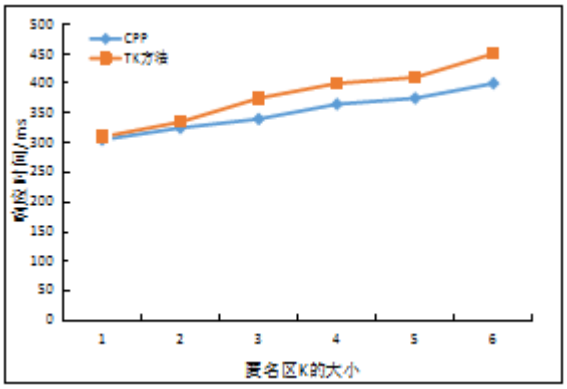

图 3 TC 算法与本文算法（CPP）系统响应时间对比

由图 3 可以看出，当用户的 Lmin 和 Lmax 以及用户速度相同时，随着 K 值的增大，系统的响应时间也随之增加。用户的响应时间是指通过匿名算法对用户的真实位置进行匿名，然后将匿名区以及请求内容发送给 LBS 服务器，直至服务器返回第一个检索点的这段时间。这是由于当 K 值越小时，用户寻找或者生成匿名区花费的时间越小，当 K 值达到一定值时，系统的响应时间明显增加，这是因为当 K 值越大的情况下，用户搜索其他 K-1 个用户的时间和匿名区的大小随之增加。由图 3 可以看出，本章提出的用户轨迹隐私保护算法的系统响应时间虽然也会随着 K 值的增加，系统的响应时间也会增加，但是与不同时刻 K-匿名算法（TK）相比，本文提出的用户轨迹隐私保护算法的系统响应时间较短。

无论是连续查询的位置隐私保护算法还是基于快照的位置隐私保护算法，用户的响应时间通与用户的位置隐私保护程度程是一对矛盾体。因此，在保证系统的响应时间的情况下，用户的位置隐私保护程度也是衡量一个算法的重要指标。图 4 是 CPP 算法与 TK 算法的用户轨迹隐私保护程度的对比。

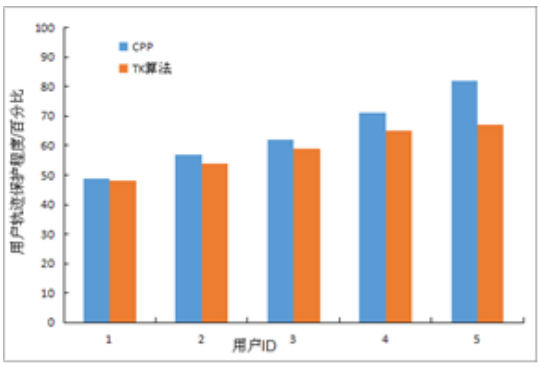

图 4 TC 算法与本文算法（CPP）隐私保护程度对比

由图 4 可以看出,CPP 算法与 TK 算法相比，当 K 比较小时，两个算法的用户轨迹的保护程度相差并不大，但是随着 K 值的增加，可以明显看出 CPP 算法的用户轨迹保护程度明显增加。这是因为当 K 的值达到一定数量时，虚假轨迹是根据真实轨迹的相似及虚假越少条件随机生成的，K 越大，与真实轨迹相似度高的虚假轨迹数量就会越多，因此 K 值越大，CPP 算法的用户轨迹保护程度越好。

连续查询的用户轨迹隐私保护算法（CPP）不仅适合连续 LBS 下的用户轨迹保护而且对于快 LBS 场景下用户位置隐私保护也同样适用。为了验证 CPP 算法在基于快照 LBS 效性，本次实验用户只发出一次请求，然后查看 CPP 算法针对单点的位置隐私保护程度与 SPP 算法相同的情况下，服务质量的对比。CPP 算法与 SPP 算法查询精度的对比如图 5 所示。

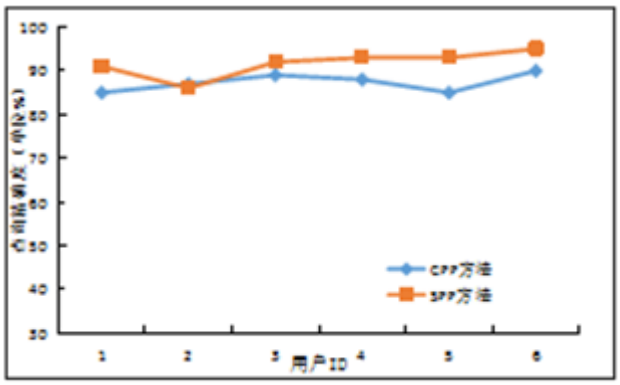

图 5 TC 算法与本文算法（CPP）查询精度对比

由图 5 可以看出,针对快照 LBS 的基于连续查询的用户轨迹位置隐私保护方法（CPP）与用户单点位置隐私保护算法（SPP）在相同的位置隐私保护程度下，两个算法的准确度都在维持在相对稳定并且比较高的百分。虽然位置 SPP 算法单点的查询率比 CPP 算法略高一点，但 CPP 算法在服务质量较高的前提下能有效的保护用户的轨迹隐私信息。

基于连续查询的用户轨迹位置隐私保护方法（CPP）为了进一步加强虚假轨迹的真实性，虚拟轨迹的某时刻的位置信息不能随机生成，必须遵循用户的真实的轨迹信息的相似性。因此提出了匿名区虚假轨迹用户的位置信息不仅要满足一定的距离约束，同时也要在方向以及速度上也与真实的轨迹信息具有一定的相似性。这样可使攻击者从所有的用户轨迹信息中分辨出用户的真实的轨迹。通过上述响应时间、用户位置隐私保护程度方面的对比，可以看出 CPP 算法与不同时刻 K-匿名算法（TK）相比不仅响应时间相对稳定，而且在保障用户的查询服务质量的同时可以有效的保护用户的轨迹隐私。

4　结束语

本文对连续 LBS 查询的用户轨迹信息提出了新的保护算法，为了进一步加强虚假用户的真实性，虚拟轨迹的某时刻的位置信息不能随机生成，必须遵循用户的真实的轨迹信息的相似性。因此提出了匿名区虚假用户要满足一定的约束条件。虚假轨迹的某时刻的位置信息不仅要满足一定的距离约束，同时

也要再方向以及速度上也与真实的轨迹具有一定的相似性。经过实验验证该算法不仅可以满足连续查询的用户轨迹保护而且可以满足快照 LBS 用户位置隐私保护。位置的语义隐私保护是本课题继续研究的方向

参考文献

[1] Huang L, Matsuura K, Yamane H, *et al*. Enhancing wireless location privacy using silent period[C]// Prpc of IEEE Wireless Communications & Networking Conference. 2005: 1187-1192.

[2] Huang L, Yamane H, Matsuura K, *et al*. Silent cascade: enhancing location privacy without communication qos degradation[C]// Prpc of the 3rd International Conference Security in Pervasive Computing. 2006: 165-180.

[3] Zhong G, Goldberg I, Hengartner U. Louis, lester and pierre: three protocols for location privacy[M]// Privacy Enhancing Technologies. Springer Berlin Heidelberg, 2007: 62-76.

[4] Lu Qin, Jeffrey Xu Yu, Bolin Ding, *et al*. Monitoring aggregate kNN objects in road networks[C]// Proc of International Conference on Scientific and Statistical Database Management. Springer-Verlag, 2008: 168-186.

[5] Wang T, Liu L. Privacy-Aware Mobile Services over Road Networks. [J]. Proceedings of the Vldb Endowment, 2009, 2(1): 1042-1053.

[6] Reem D. An Algorithm for Computing Voronoi Diagrams of General Generators in General Normed Spaces[C]// Proc of the 6th International Symposium on Voronoi Diagrams. 2009: 144-152.

[7] Reem D. The geometric stability of voronoi diagrams with respect to small changes of the sites[C]// Proc of the 27th annual symposium on Computational geometry. 2011: 254-263.

[8] Che Y Z, Chiew K, Hong X Y, *et al*. EDA: an enhanced dual-active algorithm for location privacy preservation inmobile P2P networks[J]. Journal of Zhejiang University Science C, 2013, 14(5): 356-373.

[9] Che Y, Yang Q, Hong X. A dual-active spatial cloaking algorithm for location privacy preserving in mobile peer-to-peer networks[C]// Proc of Wireless Communications and Networking Conference. 2012: 2098-2102.

[10] Kalnis P, Ghinita G, Mouratidis K, *et al*. Preventing Location-Based Identity Inference in Anonymous Spatial Queries[J]. IEEE Trans on Knowledge & Data Engineering, 2008, 19(12): 1719-1733.