基于位置混淆的轨迹隐私保护方法

张少波[1]，刘琴[2]，王国军[3]

（1. 湖南科技大学计算机科学与工程学院，湖南 湘潭 411201；2. 湖南大学信息科学与工程学院，湖南 长沙 410082；
3. 广州大学计算机科学与教育软件学院，广东 广州 510006）

摘 要：在用户连续查询过程中，针对第三方匿名器结构中 K 匿名难以保证用户隐私的问题，提出一种基于位置混淆的轨迹隐私保护方法。首先通过位置预测机制和假位置选择机制获得(G−1)个查询混淆位置，然后将其与用户真实查询位置一起发送到不同匿名器形成匿名域后，再发送到 LBS 服务器进行查询，最后将获得的查询结果经不同匿名器返回给用户。该方法通过位置混淆来混淆用户的真实查询位置，使攻击者从单匿名器和 LBS 服务器不能推断出用户的真实轨迹，加强了对用户轨迹的隐私保护，也解决了单匿名器的性能瓶颈问题。安全分析表明了该方法的安全性，实验结果表明，该方法能减少用户与 LBS 服务器的交互次数以及单匿名器的开销。

关键词：轨迹隐私；位置混淆；位置预测；假位置；匿名器

中图分类号：TP393

文献标识码：A

doi: 10.11959/j.issn.1000−436x.2018119

Trajectory privacy protection method based on location obfuscation

ZHANG Shaobo[1], LIU Qin[2], WANG Guojun[3]

1. School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
2. College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China
3. School of Computer Science and Educational Software, Guangzhou University, Guangzhou 510006, China

Abstract: In the process of continuous queries, a method of trajectory privacy protection based on location obfuscation was proposed to solve the problem that K-anonymity was difficult to guarantee user privacy in third party architectrue. Firstly, the (G−1) query obfuscation locations through the location prediction was obtained and the dummy location selection mechanism, and then sent them together with the user's real query location to different anonymizers to form cloaking regions and sent them to the LBS server for queries, and the query results were returned to the user by different anonymizers. In this method, the user's real query location was confused by the location obfuscation, and the attacker couldn't deduce the user's trajectory from a single anonymizer or the LBS server. The method can enhance the privacy of the user's trajectory and can effectively solve the performance bottleneck in the single anonymizer structure. Security analysis shows the security of the proposed approach, and experiments show this method can reduce the number of interactions between the user and the LBS server and the overhead of the single anonymizer.

Key words: trajectory privacy, location obfuscation, location prediction, dummy location, anonymizer

收稿日期：2017−10−28；**修回日期**：2018−05−11

通信作者：王国军，csgiwang@gmail.com

基金项目：国家自然科学基金资助项目（No.61632009, No.61472451, No.61402161, No.61772194）；湖南省自然科学基金资助项目（No.2016JJ3046）；广东省自然科学基金资助项目（No.2017A030308006）；广东省高等教育高层次人才基金资助项目（No.2016ZJ01）

Foundation Items: The National Natural Science Foundation of China (No.61632009, No.61472451, No.61402161, No.61772194), The Natural Science Foundation of Hunan Province (No.2016JJ3046), The Natural Science Foundation of Guangdong Province (No.2017A030308006), The High Level Talents Program of Higher Education in Guangdong Province (No.2016ZJ01)

1 引言

近年来，随着无线通信技术、移动互联网和定位技术的迅速发展，基于位置服务（LBS, location based service）已受到人们的广泛关注[1-2]。用户使用智能手机或掌上电脑，可以从应用商店下载基于位置服务的软件，如 Twitter、Foursquare 和 Gowalla 等。通过使用这些 LBS 应用软件发送查询到 LBS 服务器，可以获得用户需要的兴趣点（POI, point of interest），如交通导航信息、基于位置的广告、最近提供用户最喜欢的菜肴的餐厅等[3-4]。然而，用户在享受 LBS 带来极大生活便利和娱乐的同时，他们需要将这些查询请求提交给不可信的位置服务提供商（LSP, location service provider）。在连续 LBS 查询中，LSP 根据收集的用户查询数据可以直接追踪到用户或推断出一些敏感的用户个人信息，如日常行为、家庭地址和社会关系等，这将严重导致用户个人隐私的泄露[5]。

为减少 LBS 中轨迹隐私泄露，国内外学者已提出一些轨迹隐私保护方法，它们主要采用 2 种基本结构[6]：基于点对点结构[7]和基于可信第三方（TTP, trusted third party）中心服务器结构[8]。在基于点对点结构中，Chow 等[9]首次提出用户协作的点对点匿名方法，但该方法在寻找用户的过程中会产生较大开销。为减少开销，Shokri 等[10]提出一种基于缓存的用户合作隐私保护方法，移动用户先在合作用户缓存中查找查询内容，当寻找失败时才通过协作的方式向 LSP 发出查询。Peng 等[11]也提出了一种基于用户合作的轨迹隐私保护方法，通过向周围多跳邻居搜寻有价值的信息构造匿名区域，并发布假查询阻止攻击者重构用户轨迹。总体而言，该结构中移动用户发送查询前需进行一定的匿名或变换处理，这将会对移动终端产生较大的计算开销，同时也不能避免恶意用户的攻击。

在基于 TTP 中心服务器结构中，引入一个可信匿名器作为移动用户和 LSP 之间的中间体，负责对用户位置进行匿名和查询结果的求精。当用户发出查询时，先将查询请求发送给匿名器进行泛化处理形成一个包括 K 个用户的匿名域[12]，然后将该匿名域发送给 LSP 进行查询，再将获得的查询候选结果集返回给匿名器，最后匿名器对候选结果集进行求精，得到精确结果并将其返回给用户。对于该结构，Gedik 等[13]最先提出了通过第三方可信服务器实现匿名功能的 TTP 结构，以达到保护用户位置隐私的目的。Hwang 等[14]提出一种移动轨迹查询点的时间混淆技术，该方法基于 TTP 结构并根据用户的隐私属性和周围条件形成匿名域，同时在查询点时间上进行混淆，攻击者不能重新构造用户的移动轨迹。Liao 等[15]设计了一种 K 匿名算法来保护用户的轨迹隐私。周长利等[16]基于虚假位置的思想，在 TTP 结构中提出一种在路网环境下的连续查询方法，该方法以路网交叉点作为锚点来代替真实用户的查询位置，以获取精确的 K 近邻查询结果。

然而基于 TTP 结构的方法也存在 2 个问题[17]：1)匿名器知道所有用户的精确位置和查询信息，如果它被攻破，这将会带来严重的安全威胁；2) 用户的查询请求和结果返回都必须经过匿名器，它承担着匿名、求精等繁重的计算任务，容易成为该结构中的性能瓶颈，同时也存在着中心点失效的风险。同时，在连续 LBS 查询过程中，基于 TTP 结构的 K 匿名技术也很难保证用户的轨迹隐私。当用户发出连续 LBS 查询时，匿名器将每个查询点都模糊成满足用户需求的匿名域，然而攻击者可根据匿名域顺序重构用户的轨迹，并且攻击者将这些匿名域包含的用户进行对比，也能识别出真实用户。

针对以上问题，本文提出一种在多匿名器框架下，引入查询位置混淆并结合匿名器选择机制和 K 匿名技术来解决用户轨迹在匿名器和 LSP 中的隐私保护问题，本文的主要贡献如下。

1) 提出一种基于位置混淆的轨迹隐私保护（TPLO, trajectory privacy protection based on location obfuscation）方法，该方法采用 $(G-1)$ 个查询混淆位置来混淆用户的真实查询位置，并结合匿名器选择机制和 K 匿名技术，以达到保护用户轨迹隐私的目的。

2) 提出一种基于位置预测和假位置选择机制的混淆位置选择方法，通过位置预测机制，减少用户信息暴露给 LBS 服务器的风险，同时使用假位置选择机制增加对用户真实位置的混淆度，以提高用户隐私。

3) 提出一种基于多匿名器进行位置匿名的框架，用户查询请求和结果信息通过多个匿名器进行处理和转发，能有效解决 TTP 结构中单匿名器单点失效风险和性能瓶颈问题。

2 系统模型和相关定义

2.1 系统模型

TPLO 模型如图 1 所示，其具体过程描述如下。

1) 用户首先在其轨迹上预测用户需要查询的 g 个预测位置，然后通过位置选择机制选择 $(G-g-1)$ 个假位置。2) 用户将 $(G-1)$ 个查询混淆位置与用户真实位置分别发送到随机选择的不同匿名器进行 K 匿名，然后将形成的 G 个匿名域发送到 LBS 服务器。3) LBS 服务器在数据库中分别查询 G 个匿名域内包含的 POI，然后将结果再经不同的匿名器返回给用户。4) 用户只接收自己轨迹上查询位置返回的结果，并将当前用户真实位置查询结果经过求精后，获得精确结果。

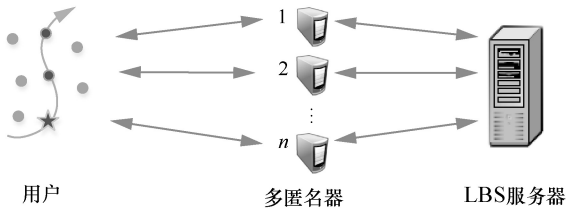


图 1　TPLO 模型

TPLO 方法的优点是攻击者不能从单个匿名器获得用户的真实轨迹，即使多个匿名器共谋，由于假位置混淆了用户的真实位置，攻击者也很难获得用户的轨迹。同时，通过混淆的用户位置并结合形成的匿名域发送到 LBS 服务器查询，LBS 服务器不能获得用户的真实轨迹。LBS 服务器查询的结果也同样经不同匿名器返回给用户。该方法中单个匿名器的失效并不影响系统的运行，能有效解决基于 TTP 结构中的单匿名器单点失效风险和性能瓶颈问题。根据系统中不同的角色和功能，系统主要由 3 类实体组成：用户、多匿名器和 LBS 服务器。

用户：携带具有全球定位、计算存储和无线通信功能智能终端的用户，他们可以通过多种方式（Wi-Fi 或 3G/4G 移动通信网络）接入移动网络，并将不同时刻的请求信息连续发送到 LBS 服务器进行查询，以获得预期的服务。本方案中用户能根据自身位置预测后续的几个查询位置，并能在其周围找到一些合适的假位置。

多匿名器：多匿名器是介于用户和 LSP 服务器之间的多个并行匿名器实体，它们具有对用户位置进行匿名，并转发用户查询请求和结果的作用。根据不同的网络环境，可以将多个匿名器部署在网络接入点或中间节点上，如基站和网关等。用户查询时分别发送 G 个不同的查询位置到不同匿名器进行 K 匿名，且查询结果也经不同匿名器返回给用户。

LBS 服务器：它是一个服务提供者，拥有大量与位置服务相关的服务和信息资源，能为用户提供各种数据服务。当 LBS 服务器收到用户发出的查询请求后，它在数据库搜索用户指定的 POI，并将满足用户需求的查询结果返回给用户。

2.2　位置预测

在 TPLO 方法中，采用基于轨迹模式的混合预测模型构建 g 个查询预测位置，相关定义如下所示。

定义 1　轨迹模式。轨迹模式是以 $R_{t_1}^{j_1} \wedge R_{t_2}^{j_2} \wedge \cdots \wedge R_{t_m}^{j_m} \xrightarrow{c} R_{t_n}^{j_n}$ 形成的特定关联规则，且时间满足 $t_1 < t_2 < \cdots < t_m < t_n$，其中，$R_{t_1}^{j_1} \wedge R_{t_2}^{j_2} \wedge \cdots \wedge R_{t_m}^{j_m}$ 为前提，$R_{t_n}^{j_n}$ 为结果，值信度 c 表示 $R_{t_1}^{j_1} \wedge R_{t_2}^{j_2} \wedge \cdots \wedge R_{t_m}^{j_m}$ 发生时 $R_{t_n}^{j_n}$ 发生的概率。

定义 2　远距离时间查询。远距离时间查询是一种时空预测性查询且满足 $t_q \geqslant t_c + d$，其中，t_q、t_c 分别表示查询时间和当前时间，d 是时间间隔阈值。

定义 3 轨迹模式树（TPT, trajectory pattern tree）。TPT 是签名树的一个变体，它具有与树相似的结构，但有不同的叶子节点。每个叶子节点包含 $\langle pk, c, p \rangle$，其中，pk 为轨迹模式的模式键值，c 为相应的值信度，p 为区域键值指针，表示模式结果。模式键值由前提键值和结果键值 2 个部分组成，为构建和搜索轨迹模式树，模式键值主要有以下操作。

1) $Union(pk_1, pk_2, \cdots, pk_n)$：对于给定的键值集 $pk_1, pk_2, \cdots, pk_n$，返回一个新的模式键值 $pk_1 \,|\, pk_2 \,|\, \cdots \,|\, pk_n$。

2) $Size(pk)$：对于给定的键值 pk，返回 pk 二进制编码中“1”的数目。

3) $Contain(pk_1, pk_2)$：对于给定的 2 个键值 pk_1 和 pk_2，如果 $pk_1 \,\&\, pk_2 = pk_2$，则返回“true”。

4) $Difference(pk_1, pk_2)$：对于给定的 2 个键值 pk_1 和 pk_2，返回 $Size(pk_1 \oplus (pk_1 \,\&\, pk_2))$。

5) $Intersect(pk_1, pk_2)$：$ck_1(ck_2)$ 和 $rk_1(rk_2)$ 分别表示 $pk_1(pk_2)$ 的结果键值和前提键值，如果 $Size(ck_1 \,\&\, ck_2) > 0$ 且 $Size(rk_1 \,\&\, rk_2) > 0$，则返回“true”，否则返回“false”。

其中，“&”“|”和“⊕”分别表示二进制中的“与”“或”和“异或”操作。

2.3 安全模型

目前，在位置隐私保护的研究方面，比较典型的攻击模型主要有 2 种[18]：强攻击者攻击模型和弱攻击者攻击模型。

1) 强攻击者攻击模型

在强攻击者攻击模型中，攻击者能监视整个系统中特定用户的行为记录。攻击者通常不破坏协议流程，但它试图从自己获取的信息中分析得到用户的其他信息。TPLO 方法中的匿名器和 LSP 可能成为潜在的强攻击者。匿名器在用户和 LBS 服务器之间进行匿名和转发信息，可能会对用户行为进行分析而造成用户信息泄露。LSP 管理所有用户的 LBS 查询数据，且可能会因利益关系泄露 LBS 服务器中的敏感信息给第三方。

2) 弱攻击者攻击模型

在弱攻击者攻击模型中，攻击者具有很少的关于用户的背景知识，攻击者可以通过使用背景知识或其他一些攻击手段进行攻击，试图知道其他用户的更多个人敏感信息。通常攻击者通过侦听不安全的无线信道，试图窃听信息并推断出一些用户的敏感信息，如用户的敏感位置、真实身份和兴趣爱好等。TPLO 方法中，攻击者通过试图窃听用户与 LBS 服务器之间的通信信道，并分析传输过程中的数据进行攻击。

3 TPLO 方法实现

实现 TPLO 方法的过程主要分为 5 个步骤：用户查询请求、匿名器匿名、服务器查询、匿名器转发与用户求精结果，本节将分别对其进行介绍。TPLO 方法中的符号定义及描述如表 1 所示。

表 1 TPLO 方法中的符号定义及描述

符号	描述
E 、En	非对称和对称加密函数
k_i	用户生成的第 i 个随机密钥
Q	用户的查询内容
(x_i, y_i)	第 i 个查询位置坐标值
K	匿名度
G	用户每次查询包含的查询位置数目
g	用户轨迹上预测的查询位置数目
R	查询范围半径
CR_i	匿名域
Re_i	用户查询兴趣点的集合
$MSG_{A_j}^{U_i}$	用户第 i 个位置向第 j 个匿名器发送的查询信息
$MSG_S^{A_j}$	第 j 个匿名器转发查询信息给 LBS 服务器
$MSG_{A_j}^S$	LBS 服务器返回结果给第 j 个匿名器
$MSG_U^{A_j}$	第 j 个匿名器转发查询结果给用户

3.1 用户查询请求

3.1.1 混淆位置选择

TPLO 方法通过位置预测机制和假位置选择机制，选择 $(G-1)$ 个查询混淆位置来混淆用户的真实查询位置，以提高用户连续查询过程中的轨迹隐私。首先使用位置预测机制在轨迹上预测 g 个查询预测位置，然后再选择 $(G-g-1)$ 个查询假位置，最后构成 $(G-1)$ 个查询混淆位置，其中，$g\varphi \leqslant G$ 且 $2 \leqslant \varphi \leqslant G$。$\varphi$ 越大，需要生成的查询预测位置就越少，相应地，需要的查询假位置就越多，从而用户查询真实位置被混淆的程度就越大，但会增加用户后续查询点的系统开销，因此应根据系统需求来设置一个合适的 φ。

1) 位置预测机制

基于 LBS 的移动用户位置预测技术已广泛被使用在推荐系统、广告推送和智能交通服务等方面。在基于个人移动位置的预测中，通常需要考虑用户的当前移动速度、方向和查询频率等轨迹模式，然后与预定义的运动函数相结合，以预测用户的未来位置。在 TPLO 方法中，使用文献[19]提出的基于轨迹模式的混合预测模型来预测用户的未来查询位置。在预测过程中，对于给定对象的最近移动位置和查询时间，首先生成模式键值 q，然后比较 TPT 中 q 的前提键值 rkq 与模式键值 pk 的前提键值 rk，rkq 和 rk 之间的前提相似值是通过将在 rk 和 rkq 中“1”的所有权重相加进行测量，其前提相似值 S_r 可以表示为

$$S_r = \sum_{i=1}^{size(rk\&rkq)} w_i, 0 \leqslant S_r \leqslant 1 \tag{1}$$

其中，w_i 是前提键值 rk 中二进制第 i 位置“1”的权重。

对于远距离的预测查询，通过使用额外的参数时间松弛长度 t_ε 来对结果键约束进行调整。在任意前提相似性情况下，任何轨迹模式的结果时间偏移在时间间隔 $[t_q - t_\varepsilon, t_q + t_\varepsilon]$ 是一个合格的候选模式。通过基于查询时间 t_q 及时间偏移 t 之间的时间差来权衡候选结果键值，其结果相似值 S_c 可表示为

$$S_c = 1 - \frac{|t_q - t|}{t_\varepsilon + 1}, 0 \leqslant S_c \leqslant 1 \tag{2}$$

同时通过 S_r 和 S_c，其模式相似值为

$$S_p(pk, q) = (S_r + S_c)c \tag{3}$$

对于远距离的预测查询，进一步定义模式相似值 S_p 可表示为

$$S_p(pk, q) = \left(S_r \frac{d}{t_q - t_c} + S_c \right) c \tag{4}$$

其中，t_c 和 t_q 分别表示当前偏离时间和更远的查询时间，且 $0 < \frac{d}{t_q - t_c} \leqslant 1$。

算法 1　用户查询预测位置算法

输入　查询模式键值 q，时间松弛长度 t_ε，g

输出　g 个预测位置

1) $i = 1$；
2) 在时间间隔 $[t_q - it_\varepsilon, t_q + it_\varepsilon]$，通过搜索 TPT 获得候选轨迹模式集 C；
3) if $C \neq \phi$ then
4)　　用式(4)对 C 中的模式进行排序；
5)　　返回最高 g 个模式的中间模式；
6) else
7)　　$i = i + 1$；
8)　　if $t_q - i \times t_\varepsilon > t_c$ then
9)　　跳转到第 2)步；
10)　　else
11)　　调用运动函数；
12)　　end if
13) end if

因此，在远距离的预测查询中，根据用户轨迹模式以及运动函数，由算法 1 可以预测用户未来的 g 个预测位置。同时通过将用户的查询预测位置事先在 LBS 服务器查询，可以减少用户和 LBS 服务器之间的交互，并减少用户信息暴露给 LBS 服务器的风险。

2) 假位置选择机制

假如用户的真实位置、预测位置和选择的假位置都非常靠近且都位于医院、学校等特定的场所，就容易暴露用户的隐私，同时假位置不能选择在海洋、湖泊等一些不可能的区域，因此如何选择合适的假位置至关重要[20]。在 TPLO 方法中，首先根据用户的真实位置和预测位置，在其附近生成一些均匀分散的临时位置，然后在这些临时位置附近的实际路网上最终确定其假位置，以避免假位置生成在一些不可能的区域。

用户在真实位置 L_u 发送查询请求时，首先通过位置预测机制获取 g 个预测位置 $\{L_1, L_2, \cdots, L_g\}$，然后随机选择一个中心点 L_c，并以该中心点 L_c 为圆心构建一个半径为 r 的虚拟圆，且满足用户的真实位置 L_u 以及最后一个查询预测位置分别与中心点 L_c 两点之间的欧几里得距离 $D(L_u, L_c) = D(L_g, L_c) = r$，且 $r \geqslant r_{\min}$，$r_{\min}$ 为系统设置的阈值。同时将虚拟圆使用 $(G - g - 1)$ 条分隔线将其划分成 $(G - g - 1)$ 个扇形区域。在第一部分的扇形区域中，中心角 $\theta = \angle L_u L_c L_g$，且 $0° < \theta \leqslant 180°$，其余扇形区域的中心角都为 $\delta = \frac{360° - \theta}{G - g}$，同时在这个虚拟圆与分隔线相交处获得 $(G - g - 1)$ 个候选位置点 $\{L'_{g+1}, L'_{g+2}, \cdots, L'_{G-1}\}$。最后，分别在每个候选位置点附近的实际路网上选择一个合适的假位置，获得 $(G - g - 1)$ 个假位置 $\{L_{g+1}, L_{g+2}, \cdots, L_{G-1}\}$。如图 2 所示，粗线表示路网，实五角星和空五角星分别表示用户的查询真实位置和预测位置，空心圆和实心圆分别表示候选位置和假位置。

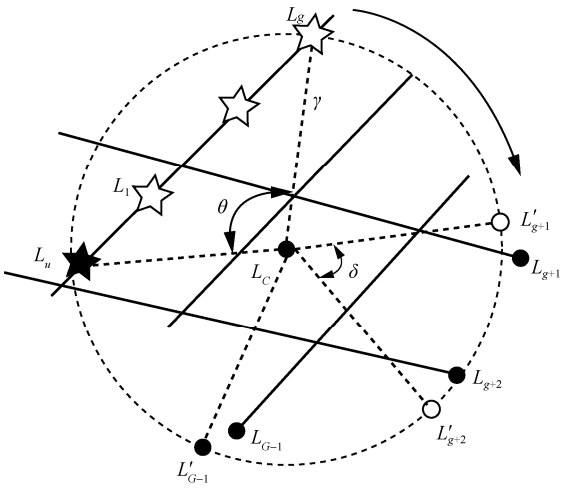


图 2　假位置选择

3.1.2　匿名器选择机制

在 TPLO 方法中，用户与 LBS 服务器之间共部署有 N 个匿名器 $(A_0, A_1, \cdots, A_{N-1})$，其编号分别为 $0, 1, \cdots, N-1$。用户查询时使用随机映射机制，将用户选择的 G 个查询位置分别分配到随机选定的

G 个不同匿名器进行处理，且 $N \geqslant G$。通过将 G 个查询位置的坐标值作为变量，构造一个散列函数并将其取模从而构造一个映射表，以获得映射到编号为 l 的匿名器 A_l，$l=0,1,\cdots,N-1$。

$$A_l=\text{Hash}(x_i+y_i) \bmod N,\ 1 \leqslant i \leqslant G, 0 \leqslant l < N \quad (5)$$

在以上过程中，如果存在不同位置都映射到编号相同的匿名器，就会产生冲突。为解决该问题，本文方案采用二次探测再散列的方法进行处理，对有冲突的匿名器编号，再通过式(6)进行计算。

$$A_l=(\text{Hash}(x_i+y_i)+P) \bmod N, 1 \leqslant t < N \quad (6)$$

其中，先取 P 值为 1，如果获得的匿名器编号出现冲突，依次在 P 值基础上增加 1，直到解决冲突为止。通过该方式可以将 G 个查询位置分别映射到不同的匿名器。

3.1.3 用户发起查询

用户形成 G 个查询请求消息，且每个查询请求消息包括使用对应匿名器公钥 PK_{A_j} 对用户的身份标识 ID_u、位置坐标 (x_i, y_i) 和匿名度 K 进行非对称加密形成的 $E_{PK_{A_j}}(ID_u,(x_i,y_i),K)$ 以及使用 LBS 服务器公钥 PK_S 对查询内容 Q、用户密钥 k_i 和查询半径 R 进行非对称加密形成的 $E_{PK_S}(Q,k_i,R)$。在映射表中，如果第 i 个查询位置被映射到第 j 个匿名器，则用户将形成基于第 i 个查询位置的查询请求，其请求消息为

$$MSG_{A_j}^{U_i}=\{E_{PK_{A_j}}(ID_u,(x_i,y_i),K), E_{PK_S}(Q,k_i,R)\},\ 1 \leqslant i \leqslant G, 0 \leqslant j < N \quad (7)$$

最后，用户根据映射表，将各查询位置分别形成查询请求信息，发送到不同的匿名器进行匿名。

3.2 匿名器匿名

用户将 G 个查询消息分别发送到不同的匿名器后，各匿名器首先对查询请求消息中的 $E_{PK_{A_j}}(ID_u,(x_i,y_i),K)$ 进行解密，获得用户的 ID_u、位置坐标 (x_i, y_i) 和匿名度 K。然后，匿名器根据位置 (x_i, y_i)、匿名度 K 选择其他 $K-1$ 个用户形成包含 K 个用户的匿名域 CR_j。在该匿名域中，攻击者能猜出用户的概率只有 $\frac{1}{K}$，因此，K 值越大，匿名度就越高。最后，各匿名器将其匿名器标识 ID_{A_j}、匿名域 CR_j 以及 $MSG_{A_j}^{U_i}$ 中其他信息组成新的查询请求消息 $MSG_S^{A_j}$ 发送到 LBS 服务器。

$$MSG_S^{A_j}=\{ID_{A_j}, CR_j, E_{PK_S}(Q,k_i,R)\} \quad (8)$$

3.3 服务器查询

LBS 服务器收到用户请求后，先验证匿名器 ID_{A_j} 的合法性。只有当 ID_{A_j} 合法时，LBS 服务器才为该匿名器提供查询服务，否则停止服务。LBS 服务器首先使用自己的私钥 SK_S 解密 $MSG_S^{A_j}$ 中的 $E_{PK_S}(Q,k_i,R)$，获得查询内容 Q、用户密钥 k_i 和查询半径 R。然后根据各匿名域 CR_j、查询内容 Q 和查询半径 R，在 LBS 数据库中查询用户需要的 POI，以获得用户需要查询的兴趣点集 Re_i $(1 \leqslant i \leqslant G)$，并对它们分别使用对称加密算法和密钥 k_i 进行加密得到 $En_{k_i}(Re_i)$。最后 LBS 服务器将加密的查询结果集 $En_{k_i}(Re_i)$ 返回给对应的匿名器，其消息为

$$MSG_{A_j}^{S}=\{ID_{A_j}, En_{k_i}(Re_i)\},\ 1 \leqslant i \leqslant G, 0 \leqslant j < N \quad (9)$$

3.4 匿名器转发与用户求精结果

G 个匿名器收到 LBS 服务器的转发请求消息 $MSG_{A_j}^{S}$ 后，分别将其转发给用户。匿名器转发给用户的消息为

$$MSG_U^{A_j}=\{En_k(Re_i)\},1 \leqslant i \leqslant G \quad (10)$$

用户只接收自己真实位置以及轨迹上查询预测位置返回的加密结果集，并使用密钥 k_i 解密 $En_{k_i}(Re_i)$，获得每个 POI 的精确位置 (x_j, y_j)，然后用户计算包含在自己查询范围内的 POI，获得精确查询结果。同时用户将预测的 g 个位置查询结果缓存在用户端，供后续的查询点使用，以减少用户与 LBS 服务器的交互，并提高用户隐私。

4 安全性分析

本节主要分析 TPLO 方法分别抵制强攻击者和弱攻击者的攻击，并将匿名器和 LSP 当作强攻击者，窃听者当作弱攻击者。具体分析如下。

4.1 抵制单匿名器的攻击

挑战。多个匿名器在用户和 LBS 服务器之间负责对用户位置进行匿名，并对查询请求、查询结果等信息进行转发。单匿名器作为强攻击者试图从用户这些数据中推断出一些敏感信息，从而揭露用户的运动轨迹。如果单匿名器可以确定地知道用户所对应的轨迹，那么单匿名器将赢得这个游戏。

定理 1 TPLO 方法能抵制单匿名器的推断攻击。

证明　TPLO 方法中，用户首先通过位置预测机制和假位置选择机制选择 $G-1$ 个混淆位置，并与用户真实位置一起发出 G 个位置查询请求，其查询请求消息包括用户的身份标识 ID_u、查询位置坐标 (x_i, y_i)、匿名度 K 以及加密的 $E_{PK_S}(Q, k_i, R)$。从该查询请求信息中，单个匿名器只能获得用户身份标识，它不能与用户的真实查询位置进行关联。同时，G 个查询位置同时发送到不同的匿名器处理，其中只有一个匿名器处理的是用户真实位置，因而用户与真实位置关联的概率只有 $\frac{1}{G}$。由于 $G-1$ 查询混淆位置会混淆用户的真实位置，从而即使多个匿名器共谋，攻击者也很难确定用户的真实轨迹。

当查询结果返回给用户时，G 个查询结果 Re_i $(1 \leqslant i \leqslant G)$ 都分别使用密钥 k_i 进行了加密 $En_{k_i}(Re_i)$。因此，随机选择的单个匿名器没有用户密钥 k_i，就无法解密查询结果获取用户结果信息。

从以上分析可知，单匿名器无法推断出用户的真实轨迹。

4.2　抵制 LSP 的攻击

挑战。LSP 管理所有用户的查询数据，LSP 作为强攻击者试图从用户这些查询数据中推断出一些关于用户的敏感信息，从而揭露用户的精确位置。如果 LSP 可以成功地猜测出指定用户的查询内容或所对应用户的精确位置，那么 LSP 将赢得这个游戏。

定理 2　TPLO 方法能抵制 LSP 的推断攻击。

证明　在 TPLO 方法中，当用户在查询预测位置发出查询请求时，用户可直接从缓存获取查询结果。在该过程中，用户与 LSP 没有进行交互，LSP 就无法获取用户的任何信息。

如果用户在其他位置发出查询请求，他将通过映射机制选择第 j 个匿名器转发用户的查询请求给 LSP，其查询请求消息为 $MSG_S^{A_j}$，它包括匿名器标识 ID_{A_j}、匿名域 CR_j、查询内容 Q、用户密钥 k_i 和查询半径 R。从这些信息中，LSP 无法获取准确的用户位置，而且 LBS 服务器同时获得 G 个匿名域 CR_j，它不能确定用户的真实位置存在于哪个匿名域中。即使 LSP 知道用户位于某个 CR_j 内，但该区域至少包括有 K 个用户，因此 LSP 能猜到是某个用户的概率最多只有 $\frac{1}{K}$。在 LBS 服务器根据 Q、CR_j 和 R 搜索查询结果的过程中，LSP 也只知道查询内容 Q，它不能将其与特定的用户进行关联。

从上述分析可知，LSP 无法准确地确定用户的位置，也不能从 LBS 服务器中的用户查询数据正确地猜测出与查询内容对应的用户。

4.3　抵制窃听者的攻击

挑战。弱攻击者通过侦听不安全的无线信道，试图从这些数据中能推断出一些用户的敏感信息，从而揭露出指定用户的轨迹或查询内容。如果弱攻击者可以成功地猜测出指定用户的查询内容或所对应的用户轨迹，那么弱攻击者将赢得这个游戏。

定理 3　TPLO 方法能抵制侦听者的攻击。

证明　在 TPLO 方法中，用户第 i 个查询位置发送给第 j 个匿名器的消息为 $MSG_{A_j}^{U_i}$，它包括信息 $E_{PK_{A_j}}(ID_u, (x_i, y_i), K)$ 和 $E_{PK_S}(Q, k_i, R)$，它们都分别使用了第 j 个匿名器的公钥 PK_{A_j} 和 LBS 服务器的公钥 PK_S 进行加密。从这些信息中，弱攻击者没有该匿名器和 LBS 服务器的私钥 SK_{A_j} 和 SK_S，就不能解密信息，从而得不到用户的任何信息。

在匿名器转发用户查询请求给 LBS 服务器的请求消息 $MSG_S^{A_j}$ 中，弱攻击者只能获取匿名器标识 ID_{A_j}、匿名域 CR_j 和加密的 $E_{PK_S}(Q, k_i, R)$。类似地，弱攻击者不能解密 $E_{PK_S}(Q, k_i, R)$，它只能获得匿名器标识 ID_{A_j}、匿名域 CR_j。从这些信息中，弱攻击者无法确定特定用户的位置。

在查询结果返回给用户的结果消息 $En_{k_i}(Re_i)$ 中，G 个查询结果 Re_i $(1 \leqslant i \leqslant G)$ 都分别使用密钥 k_i 进行了加密。因此，弱攻击者没有用户密钥 k_i 就无法解密这些查询结果，从而得不到任何有用的信息。

从以上分析可知，窃听者不能猜测出指定用户的查询内容或所对应的用户轨迹。

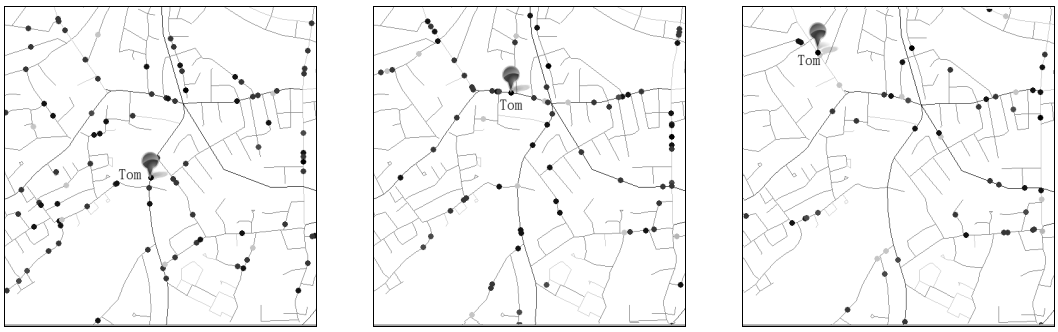
5　实验及结果分析

本节主要通过实验验证用户连续查询时与 LBS 服务器的交互情况、在相关参数变化下对 TPLO 方法性能的影响及在匿名器的开销上与 TTP 结构的 Gedik 方法[13]和 Hwang 方法[14]进行仿真实验对比。

实验采用 Brinkhoff 移动对象生成器[21]，输入德国奥尔登堡市交通网络图并生成 10 000 个移动

(a) 随机生成 10 000 个移动用户　(b)放大 4 倍后的移动用户分布　(c)放大 8 倍后的移动用户分布

图 3　德国奥尔登堡市交通网络图上生成的移动用户分布



(a) t_1 时刻的位置　(b) t_2 时刻的位置　(c) t_3 时刻的位置

图 4　移动对象 Tom 的运动轨迹

对象。德国奥尔登堡市交通网络图上生成的移动用户分布如图 3 所示。图 3(b)和图 3(c)是在图 3(a)基础上分别放大 4 倍和 8 倍后的移动用户分布。移动用户集数据随机分布，实验随机选取移动对象 Tom 的移动轨迹作为实验对象。移动对象 Tom 的运动轨迹如图 4 所示。实验参数设置如表 2 所示。实验的硬件环境为：Intel(R) Core(TM) i5-4590 CPU @3.30 GHz，4.00 GB 内存，操作系统为 Microsoft Windows 7，采用 MyEclipse 开发平台，以 Java 编程语言实现。

表 2　　实验参数

参数	取值
用户数 *m*/个	10 000
兴趣点数 *POI*/个	10 000
匿名器数 *N*/个	100
匿名度 *K*/个	10~100
每次查询位置数 *G*/个	10~100
查询半径 *R*/km	0.5~1.5

5.1　用户与 LBS 服务器的交互情况

当 G=20 、φ 分别取不同值时，分析用户在运动轨迹上连续查询发出 10 次查询请求过程中，取不同 φ 值时与 LBS 服务器交互次数的对比情况。由表 3 数据可知，φ 值越小，用户与 LBS 服务器交互的次数就越少，暴露给 LBS 服务器的信息就越少，因而用户隐私度就越大。

5.2　参数变化对 TPLO 性能的影响

当 R=1 、φ=2 、连续查询 50 个用户真实位置时，通过改变用户每次查询选择的查询位置数目来分析对 TPLO 性能的影响，实验结果如图 5 所示，由图 5 可知，在时间和通信开销上，它们都随着 G 值的增大而增大，同时 K 值越大，其开销也越大。因为用户每次查询时选择的查询位置数 G 越多，就需要更多的匿名器形成匿名域和转发查询结果，同时 LBS 服务器也需要为更多的匿名域查询结果，从而查询所需的时间和通信开销都随之增加。同时 K 值越大，形成的匿名域就越大，相应需要的查询范围就越大，查询所需的时间和通信开销就会越多。因此，G 或 K 值越大，用户查询所需的时间和通信开销就越多。

当 R=1 、G=50 时，通过改变 φ 和匿名度 K 值分析对 TPLO 性能的影响，实验结果如图 6 所示，由图 6 可知，在时间和通信开销上，查询的时间和通信开销都随着 φ 值的增大而增大，同时 K 值越

表 3　　**用户与 LBS 服务器交互次数对比**

连续查询次数	φ=2	φ=3	φ=4	φ=5	φ=6	φ=7	φ=8	φ=9	φ=10
1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	2	2	2	2
4	1	1	1	1	2	2	2	2	2
5	1	1	1	2	2	3	3	3	3
6	1	1	2	2	2	3	3	3	3
7	1	2	2	2	3	4	4	4	4
8	1	2	2	2	3	4	4	4	4
9	1	2	2	3	3	5	5	5	5
10	1	2	2	3	4	5	5	5	5

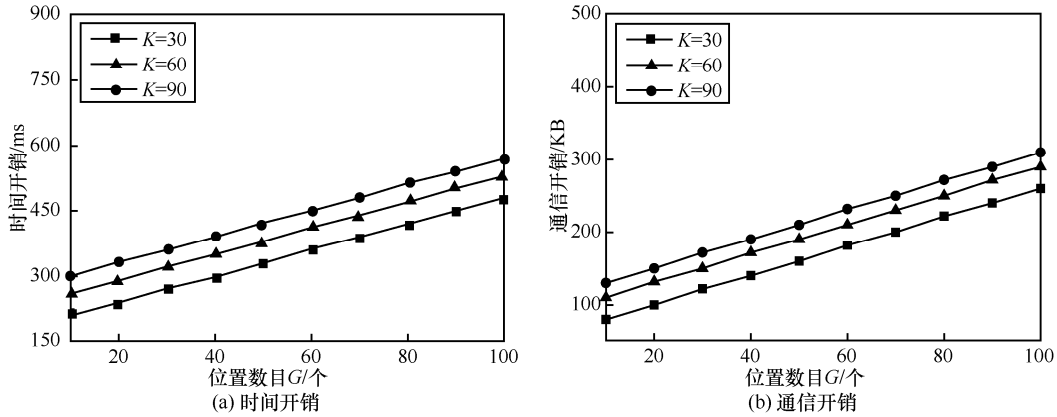


图 5　查询位置数目及匿名度变化对性能的影响

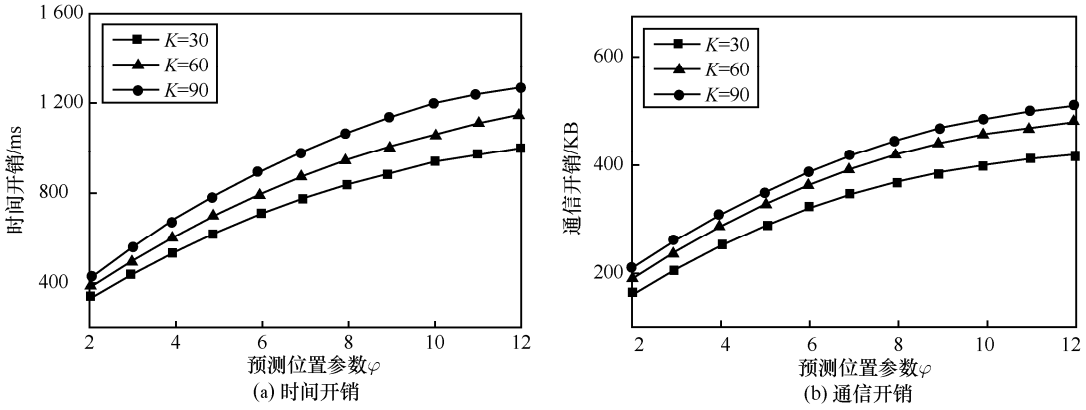


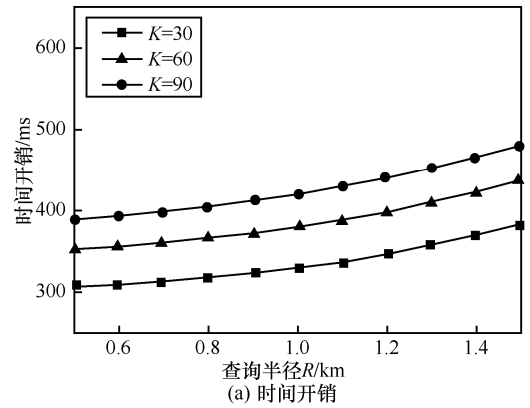
图 6　φ 及匿名度变化对性能的影响

大，时间和通信开销就越大。因为 φ 值越大，预测位置越少，而假位置就越多，所以需要更多的处理时间和通信开销。

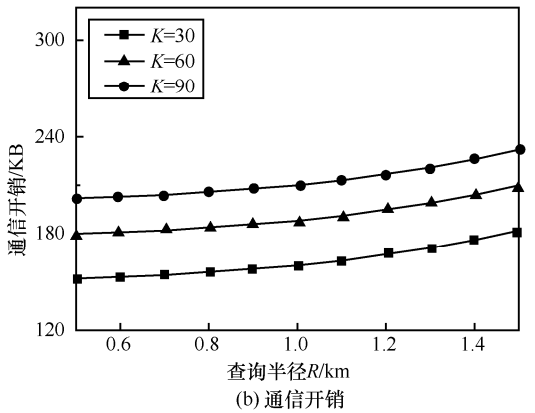
当 φ=2 、 G=50 时，通过改变查询半径 R 和匿名度 K 值分析对 TPLO 性能的影响，实验结果如图 7 所示，由图 7 可知，在时间和通信开销上，查询的时间和通信开销都随着 R 值的增大而增大，同时 K 值越大，时间和通信开销就越大。因为 R 值越大，用户需要查询的范围就越大，相应地会查询到更多的 POI，所以需要更多的处理时间和通信开销。

5.3　匿名器性能对比

本节主要从单个匿名器的平均计算时间和通信开销将本文方法与 TTP 结构的 Gedik、Hwang 方法进行仿真实验比较。图 8 所示为 R=1、φ=2 以及

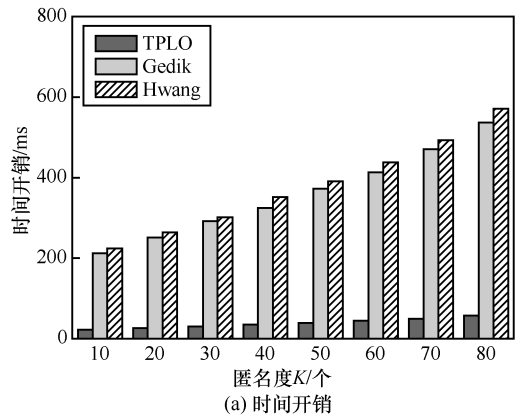


(a) 时间开销

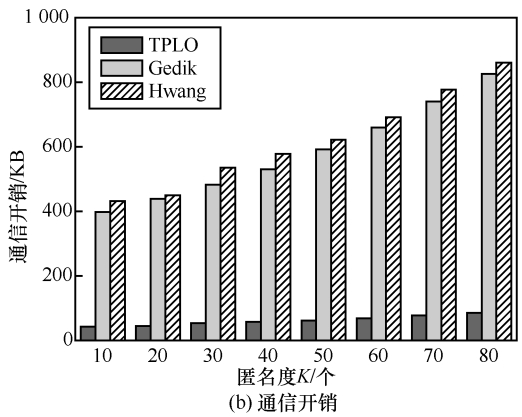


(b) 通信开销

图 7　查询半径及匿名度变化对性能的影响



(a) 时间开销



(b) 通信开销

图 8　匿名器的性能对比

$G=20$ 时，通过改变匿名度 K，TPLO 与 Gedik、Hwang 方法对匿名器性能的影响。

由图 8 可知，在匿名器的时间和通信开销上，随着 K 值增大，TPLO 相对于 TTP 结构的 Gedik、Hwang 有较大优势。因为 TPLO 方法是从 N 个匿名器随机选择 G 个匿名器一起处理用户的查询，而 TTP 中仅由一个匿名器处理，所以在单个匿名器的平均时间和通信开销上，TPLO 方法相对于 TTP 结构的 Gedik、Hwang 方法有很大优势。

6　结束语

在用户连续查询过程中，针对 TTP 结构中 K 匿名存在的用户隐私风险，本文提出了一种基于位置混淆的轨迹隐私保护方法，该方法通过在用户和 LSP 之间部署多个匿名器，采用基于位置预测和假位置选择机制的位置混淆技术，并结合匿名器选择机制和 K 匿名技术来保护用户轨迹在匿名器和 LSP 中的隐私，同时也解决了 TTP 结构中的单点失效风险和性能瓶颈问题。但在使用基于轨迹模式的混合预测模型进行查询位置预测过程中，用户查询位置预测的准确率会随着预测位置数目的增加而降低。因此，在下一步的研究工作中将考虑怎样选择一个合适的预测位置数目，在保证用户隐私的前提下，以进一步提高用户的服务质量。

参考文献：

[1] YI X, PAULET R, BERTINO E, et al. Practical approximate *k* nearest neighbor queries with location and query privacy[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(6): 1546-1559.

[2] PRIMAULT V, BOUTET A, MOKHTAR S B, et al. Adaptive location privacy with ALP[C]//The 35th Symposium on Reliable Distributed Systems (SRDS). 2016: 269-278.

[3] 雷凯跃, 李兴华, 刘海,等. 轨迹发布中基于时空关联性的假轨迹隐私保护方案[J]. 通信学报, 2016, 37(12):156-164.
LEI K Y, LI X H, LIU H, et al. Dummy trajectory privacy protection scheme for trajectory publishing based on the spatiotemporal correlation[J]. Journal on Communications, 2016, 37(12): 156-164.

[4] GRISSA M, YAVUZ A, HAMDAOUI B. Preserving the location privacy of secondary users in cooperative spectrum sensing[J]. IEEE Transactions on Information Forensics & Security, 2017, 12(2): 418-431.

[5] 张学军, 桂小林, 伍忠东. 位置服务隐私保护研究综述[J]. 软件学报, 2015, 26(9): 2373-2395.
ZHANG X J, GUI X L, WU Z D. Privacy preservation for location-based services: a survey[J]. Journal of Software, 2015, 26(9): 2373-2395.

[6] 万盛, 李凤华, 牛犇, 等. 位置隐私保护技术研究进展[J]. 通信学报, 2016, 37(12): 1-18.
WAN S, LI F H, NIU B, et al. Research progress on location priva-

cy-preserving techniques[J]. Journal on Communications, 2016, 37(12): 1-18.

[7] ARDAGNA C A, CREMONINI M, VIMERCATI S D C D, et al. An obfuscation-based approach for protecting location privacy[J]. IEEE Transactions on Dependable & Secure Computing, 2010, 8(1):13-27.

[8] 霍峥, 孟小峰, 黄毅. PrivateChechIn:一种移动社交网络中的轨迹隐私保护方法与进展[J]. 计算机学报, 2013, 36(4): 716-726.
HUO Z, MENG X F, HUANG Y. PrivateChechIn: trajectory privacy-preserving for chech-in services in MSNS[J]. Chinese Journal of Computers, 2013, 36(4): 716-726.

[9] CHOW C Y, MOKBEL M F, LIU X. A peer-to-peer spatial cloaking algorithm for anonymous location-based service[C] //The 14th annual ACM International Symposium on Advances in Geographic Information Systems. 2006:171-178.

[10] SHOKRI R, THEODORAKOPOULOS G, PAPADIMITRATOS P, et al. Hiding in the mobile crowd: location privacy through collaboration[J]. IEEE Transactions on Dependable and Secure Computing, 2014, 11(3): 266-279.

[11] PENG T, LIU Q, MENG D, et al. Collaborative trajectory privacy preserving scheme in location-based services[J]. Information Sciences, 2017, 387:165-179.

[12] ZHANG Y, TONG W, ZHONG S. On designing satisfaction- ratio-aware truthful incentive mechanisms for k-anonymity location privacy[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(11): 2528-2541.

[13] GEDIK B, LIU L. Protecting location privacy with personalized k-anonymous: architectrue and algorithms[J]. IEEE Transaction on Mobile Computing, 2008, 7(1):1-18.

[14] HWANG R H, HSUEH Y L, CHUNG H W. A novel time-obfuscated algorithm for trajectory privacy protection[J]. IEEE Transactions on Services Computing, 2014, 7(2): 126-139.

[15] LIAO D, LI H, SUN G, et al. Protecting user trajectory in location-based services[C]//IEEE Global Communications Conference (GLOBECOM). 2015: 1-6.

[16] 周长利, 马春光, 杨松涛. 路网环境下保护 LBS 位置隐私的连续 KNN 查询方法[J].计算机研究与发展, 2015, 52(11): 2628-2644.
ZHOU C L, MA C G, YANG S T. Location privacy-preserving method for LBS continuous KNN query in road networks[J]. Journal of Computer Research and Development, 2015, 52(11): 2628-2644.

[17] PENG T, LIU Q, WANG G J. Enhanced location privacy preserving scheme in location-based services[J]. IEEE Systems Journal, 2017, 11(1): 219-230.

[18] GAO S, MA J F, SHI W S, et al. TrPF: a trajectory privacy-preserving framework for participatory sensing[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(6): 874-887.

[19] JEUNG H Y, SHEN H T, LIU Q, et al. A hybrid prediction model for moving objects[C]//The 24th International Conference on Data Engineering. 2008:70-79.

[20] NIU B, ZHANG Z Y, LI X Q, et al. Privacy-area aware dummy generation algorithms for location-based services[C]//The International Conference on Communications. 2014: 957-962.

[21] BRINKHOFF T. Generating traffic data[J]. Bulletin of the Technical Committee Data Engineering, 2003, 26(2): 19-25.

[作者简介]

张少波（1979−），男，湖南邵东人，博士，湖南科技大学讲师，主要研究方向为移动社交网络、隐私保护、云计算安全、大数据安全和隐私等。

刘琴（1982−），女，湖南长沙人，博士，湖南大学副教授，主要研究方向为云计算安全、大数据安全与隐私保护等。

王国军（1970−），男，湖南长沙人，广州大学博士生导师、广东省珠江学者特聘教授，主要研究方向为信息安全、可信计算、大数据安全和隐私保护等。