# 基于位置服务中时空关联的隐私保护方案

李维皓，丁晟，孟佳洁，李晖
（西安电子科技大学网络与信息安全学院，陕西 西安 710071）

**摘 要**：基于位置服务（LBS, location-based service）在为人们的生活带来便捷的同时，对用户的隐私信息带来了不可忽略的威胁。利用时空关联性，提出基于伪位置生成的隐私保护方案，在保护用户位置隐私的同时，考虑到空间和时间之间的关联性，选取与伪位置关联的伪查询内容。时空关联的隐私保护方案包含 2 个算法，地图分割算法和伪内容生成算法。地图分割算法通过维诺多边形将地图划分为离散的位置单元，保证每一个离散的位置单元互不相邻，伪内容生成算法利用用户在下一时刻将要前往的位置作为前一时刻的查询内容，从而很好地避免了攻击者根据时间和空间的关联性来推测用户的真实信息。最后，通过实验证明所提方案的有效性和安全性。

**关键词**：位置服务；隐私保护；位置隐私；社交网络
**中图分类号**：TN929.5
**文献标识码**：A
**doi:** 10.11959/j.issn.1000-436x.2018084

# Spatio-temporal aware privacy-preserving scheme in LBS

LI Weihao, DING Sheng, MENG Jiajie, LI Hui
School of Cyber Engineering, Xidian University, Xi'an 710071, China

**Abstract:** Location-based service (LBS) brings a lot of conveniences in people's daily life, but the conveniences are accompanied with the leaking of privacy. A dummy-based location-preserving scheme was proposed, which took the correlation between spatial issues and temporal issues into account. Two algorithms were included in this scheme, map dividing algorithm and dummy contents determining algorithm. The map dividing algorithm divided the map into discrete location cells by Voronoi Diagram to ensure these discrete location cells were not adjacent to each other. The dummy contents determining algorithm replaced the query content in previous moment by the intending location in next moment, which efficiently avoided the adversary inferring mobile users' sensitive information according to the correlation between spatial issues and temporal issues. The simulation experiments show that the proposed scheme is effective and efficient.

**Key words:** location-based service, privacy-preserving, location privacy, social network

## 1 引言

移动智能终端的普遍应用为用户提供了便利的生活方式，其中基于位置服务的应用更是受到广大用户的喜爱，例如，美团外卖、大众点评、微博等应用软件。用户可以在智能终端安装各种基于位置服务的应用程序，这为用户的日常生活带来了无往不利的便捷。然而，用户在享受 LBS 带来的便利时，自身的隐私信息也在使用各类应用的过程中泄露给服务提供商，恶意用户通过获取发送请求的内容分析出用户的隐私信息，例如，ID 地址、工作、生活习惯甚至健康状况。为了保证用户在享有服务的同时，不泄露用户个人的隐私信息，保护移动用户的位置信息已经成为当下众多学者们研究的课

**收稿日期**：2017-11-08；**修回日期**：2018-03-03
**通信作者**：丁晟，shawnding.xdu@gmail.com
**基金项目**：国家重点研修基金资助项目（No.2017YFB0802201，No.2017YFB0802203）；国家自然科学基金资助项目（No.61672411, No.U1401251）
**Foundation Items:** The National Key Research and Development Program of China (No.2017YFB0802201, No.2017YFB0802203), The National Natural Science Foundation of China (No.61672411, No.U1401251)

题之一。

现有的隐私保护方案主要分为 2 类，基于可信匿名中心的隐私保护方案[1]和基于移动终端的隐私保护方案[2~4]。利用可信匿名中心来完成用户和服务提供商之间的交互，减少了移动终端在计算和存储上的压力，然而可信匿名中心成为系统性能和安全性的瓶颈，从而无法避免单点泄露的问题。当下移动终端不断智能化，在处理能力和存储能力都有着跨越式的提高，能够高效、准确地完成基本的处理、计算和存储任务，实现相应的功能，相比之下，没有可信匿名中心的参与，基于可信匿名终端的方案有效地避免了单点泄露问题，在隐私保护方面有着更明显的优势。根据隐私保护方案技术的不同，主要分为位置偏移和模糊技术[5,6]、空间匿名技术[7,8]、群组协作技术[9,10]以及基于密码学技术的隐私保护方案[11, 12]。其中位置偏移和模糊技术的安全性较低、执行简单，群组协作技术需要较高的假设前提，而密码学技术较高的安全性需要依靠强大的计算处理能力，并且十分耗时。综合考虑，空间匿名技术避免了较高的计算量和较低的安全性等弊端，能够在用户的智能终端实现较高的隐私保护，包含位置隐私和查询隐私。相较于位置隐私，查询隐私是通过用户发送的查询内容来分析、推断，进而得到用户潜在的隐私信息，查询内容背后潜在的用户隐私则被视为查询隐私。例如，用户查询附近的医院，那么该用户很有可能是一名病患。在移动社交网络中，用户的查询隐私和位置隐私都是现有隐私保护方案需要考虑、设计来实现保护的方面。

在隐私保护方案中，背景信息的存在成为攻击者推断用户真实信息的一项必不可少的条件，然而有些方案并未考虑背景信息的存在[13]，或不全面地考虑背景信息[7,8]，单一考虑地图中的历史查询概率，而忽略了在每个特定时间点的概率分布情况。从而为攻击者提供了可乘之机，通过分析、推断背景信息和发布的请求信息之间的关联，得到隐藏在请求信息后的用户隐私。因此，在设计隐私保护方案时，背景信息的存在是不容忽略的。

基于伪位置生成的隐私保护方案，即时空关联的隐私保护方案，利用空间匿名技术来保护用户的位置信息，轻量级的算法能够成功并高效地运行在移动智能终端，从而很好地避免了可信匿名中心的存在和单点泄露问题。此方案首先利用维诺多边形将地图分割为若干个多边形，进而通过用户的移动模型来预测用户在下一个时间段可能出现的位置，并通过该位置的 POI（point of interest）来决定前一时刻的请求内容。本文创新点共分为以下 3 个方面。

1) 区别于现有方案，不仅考虑了背景信息的存在，同时分别考虑了在不同的时间点上每一个位置的查询概率。

2) 考虑时空关联性，避免了攻击者通过分析在特定时间点发送请求的可能性以及移动用户在特定位置发送请求的合理性来推断用户的隐私信息。

3) 通过分析计算数据集 Geolife 得到用户在相邻时间段可能前进的位置，预测用户在下一个时间段可能出现的位置，从而决定前一个时间段的查询内容。这一创新点成功地考虑了查询内容和位置之间的可行性，提高了伪位置对攻击者的混淆程度。

## 2　相关工作

本节根据是否有可信匿名中心的参与，分析、介绍现有的隐私保护方案中使用的技术以及存在的问题。

### 2.1　基于可信匿名中心方案

目前，在基于位置服务的隐私保护方案中，大多数的方案都需要依靠可信匿名中心的参与来完成隐私保护的全过程。*k*-匿名技术是最为广泛应用的一项技术，其中将用户真实的位置信息和 $k-1$ 个虚假的用户信息发送给移动服务提供商，在暴露给服务提供商的位置信息中保证用户的真实位置信息和其他的虚假位置信息不可区分，从而实现对位置信息的保护[1,14]。*l*-多样性[15]作为 *k*-匿名的一种扩展，要求隐匿区域内至少包含一种隐私信息。可信匿名中心介于移动用户和服务提供商之间，移动用户将信息发送到可信匿名中心进行匿名处理，然后将匿名处理后的信息发送给服务提供商来换取服务信息。Beresford 等[16]提出了 Mixzone 的概念，在 Mixzone 中的用户不需要更新自己的位置信息，从而不需要和服务提供商进行交互，避免了多次交互泄露的隐私信息。Palanisamy 等[17]利用 Mixzone 的概念，在用户离开 Mixzone 时对其进行假名更换，攻击者不能通过分析进入 Mixzone 和离开该区域的关联性来获取用户的真实信息。Meyerowitz 等[18]通过缓存用户请求的服务信息来避免和服务提供商的多次交互，通过减少交互次数保护了用户的隐私信息。Xu 等[19]提出了一种基于感知的隐私保护模型，利用信息熵来衡量位置区域的受欢迎程度，

并且利用四叉树来分离请求信息和特定用户，针对不同的分类提供个性化的隐私保护。

### 2.2 非基于匿名中心方案

随着科技的进步，当下的移动智能终端具有强大的计算、处理能力和大容量的存储空间。近几年来，基于移动终端的隐私保护方案也被不断提出[3,20,21]。相较基于匿名中心的隐私方案，此类方案的架构中不需要任何可信第三方的参与，完全避免了可信第三方潜在的安全和性能隐患。CAP[2]利用四叉树对地图进行细粒度的划分，并利用 Hilbert 曲线对地图进行遍历。该方案考虑了道路的多样性，但并未将移动用户的运动模式考虑在内。SMILE[3]利用 *k*-匿名技术来度量用户的隐私程度，在基于遇见的位置服务中，通过获取用户位置信息前缀的散列值，从而避免了将用户的个人信息泄露给服务提供商。Fawaz 等[22]提出了一个新型的隐私保护框架，能够针对每一个应用程序来实现细粒度的隐私保护，并且不需要依靠任何的可信第三方。

## 3 准备工作

为了方便后文的阅读和理解，本节主要介绍文中所需要用到的一些基本概念、研究动机和基本解决方案。

### 3.1 基本概念

1) 查询内容

移动用户发送服务请求给服务提供商，其中主要包含用户的身份信息、位置信息、查询内容和时间戳。这不仅保护了用户的位置信息，同时避免了查询内容所导致的隐私泄露问题，即查询隐私[23]。在现有的隐私保护方案中，将用户的隐私主要分为位置隐私和查询隐私。例如，用户请求附近 1 km 以内的加油站，其中加油站被视为查询内容，查询内容所代表的隐私称为查询隐私。攻击者通过分析用户发出的查询内容，推断出用户的隐私信息，相较位置隐私，查询隐私同样需要保护。

2) 位置信息

移动用户在发送请求服务时，根据服务需求，需要将用户当前所处的位置信息发送给相关的服务提供商，以便获取相关的服务信息。本文提到的位置信息是指在请求服务时，用户所处的位置坐标。用户利用智能终端中的 GPS 定位模块获取当前的位置坐标，隐私保护方案旨在保护用户的位置信息不被泄露，从而避免非法用户推测出用户的真实敏感信息。

3) POI

POI 指的是在某一个位置点区别于坐标信息来辨别该位置的点。例如，Alice 在麦当劳请求周围公交车站，那么麦当劳就是该位置的 POI，而该位置的坐标则是位置信息，公交车站是查询内容。

从某种程度上来讲，POI 和查询内容在语义上具有重合性，例如，上述的例子中，麦当劳是该位置的 POI，公交车站则是查询内容。然而从公交车站的位置坐标来说，公交车站是该位置的 POI。综上所述，POI 是位置坐标上基础建设（或其他能够区别辨识的设施）的语义内容，能够作为用户进行查询的关键字，成为查询内容。

4) 背景信息

随着科技发展信息的海量增加，任何一个具有计算、处理和存储功能的个体都能够获取到地图中用户所处区域的历史查询数据，其中包含了某个地点曾经以及当前发生的请求记录，该地的查询概率以及该地点所位于的商圈和 POI（Google 地图）。现有的隐私保护方案单一地考虑了背景信息中历史数据的集合，未关注到某个特定时间点的查询概率以及与该时间点相邻时间点的查询概率。本文所涉及的背景信息是通过 Google 地图的 API 来获取区域内每个位置的查询概率，并将背景信息存储于智能终端中。

### 3.2 研究动机

现有的隐私保护方案忽略了背景信息的存在，从而给用户的隐私带来的潜在威胁，即使某些方案考虑到背景信息也可能会被恶意用户获取来推断用户的隐私，但是通常都只单一地考虑某一个位置的历史查询概率，而忽略了在不同时间点之间、时间和空间之间存在的关联性。通过一个简单的例子来解释研究动机，如图 1 所示。选取地图的一部分，其中，[0, 20]×[0, 20]是用户的请求区域，虚线表示地图中的道路情况，12:00am，用户 Alice 位于位置(5, 5)。基于背景信息所能提供的道路信息，与该区域连通的道路有且只有 2 条，它们分别是[0,10]×[0,10]→[10,20]×[0,10]和[0,10]×[0,10]→[0,10]×[10,20]，其中第一条路通往餐馆（如肯德基、麦当劳、星巴克），第二条路通往酒吧。首先，在伪位置生成的隐私保护方案中，如果选取的伪位置是一间酒吧，那么在时间戳为 12:00am 的时候发送请求给服务提供商，攻击者则可以根据在该时间酒吧不营业的原

因将该伪位置过滤掉，因此，需要考虑时间和位置所代表的 POI 之间的可行性。其次，为了保护 Alice 的位置信息，将 Alice 的隐身空间设置为[0,10]×[0,10]，基于用户的心理，在用户想要请求相应的服务时，一定是在下一个时间点或将来某个时间段要去的位置，那么在该位置发送的查询内容一定是在时间允许情况下可达的，即用户极有可能请求附近存在的 POI。在图 1 右侧，用户将要去的地方一定是他的请求区域[0, 20]×[0, 20]中，而非区域以外的位置，那么用户的查询内容应该存在于该查询区域。同理，如果选取的伪位置周围的可达范围内只有餐馆和酒吧 2 种 POI，那么如果发送的查询内容为医院，则会被攻击者判定该位置为虚假位置，即需要考虑到查询内容和所选取位置之间的关联性。

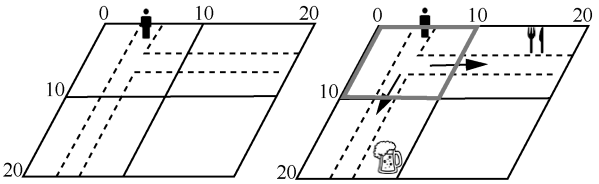



图 1　研究动机示例

现有的基于伪位置生成的隐私保护方案总是根据概率分布来选取相应的伪位置，而并未考虑到在不同的时间点所选取到的伪位置并不具有足够的说服力，即午餐时间，在酒吧发送请求的可能性远小于在办公室请求周边餐馆的可能性。同时，用户请求身边并不存在的 POI 也将会被攻击者判定为虚假的位置，即午餐时间，在城南请求城北的公园的可能性很小。

### 3.3　基本解决方案

根据以上提出的研究动机，提出了基于伪位置生成的时空关联的隐私保护方案，能够更好地保护用户的位置隐私和查询隐私。该方案包含了 2 个算法，算法 1 是地图分割算法，不同于现有方案中单一考虑地图中的查询概率，而忽略了在特定时间点不同的位置具有不同的查询概率的情况，算法 1 首先从 Google 地图中的 popular time 中获取到在特定的时间段某一个具体位置的概率情况，然后通过在特定时段发送请求的可能性对地图进行第一层的过滤，得到了若干个位置点。接着，基于第一层过滤后的位置点，利用维诺图将地图分割为若干个不规则的维诺多边形，每一个维诺多边形中有且只有一个离散的位置存在（维诺图的性质）。通过选取与用户所在的维诺多边形共边的维诺多边形，保证了待选的离散位置单元能够尽可能远，且与用户的位置单元不相邻。从而过滤掉和用户所在的维诺多边形公共边的位置单元，并将算法 1 最后得到的位置单元集作为算法 2 的输入。

算法 2 为伪内容生成算法，其中包含选取伪位置和伪查询内容。为了避免攻击者分析用户在相邻时间段的位置可达性，通过数据集 Geolife 得到转移矩阵来推测每一个位置单元在相邻的 2 个时间点之间的移动概率，通过选取概率高的位置单元所处的 POI 为前一时刻想要请求的查询内容，作为伪查询内容，结合算法 1 所得到的位置单元集合，寻找伪查询位置相临近的位置作为伪位置单元。最终，用户将自身的真实信息和伪位置以及伪查询内容一起发送给服务提供商来换取服务。

### 3.4　移动模型

利用二维坐标空间来表示地图信息，将地图划分为 $M$ 个大小相同的单元，即 $Map=\{\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_M\}$，其中，$M\in N^+,\boldsymbol{C}_i\in Map$。如图 2 所示，单元 $C_4$ 的向量可表示为

$$\boldsymbol{C}_4=\begin{bmatrix}0&0&0&1&0&\cdots&0\end{bmatrix}$$

$X=[4,6]^{\mathrm{T}}$，其中 $X[1]=4,\ X[2]=6$。

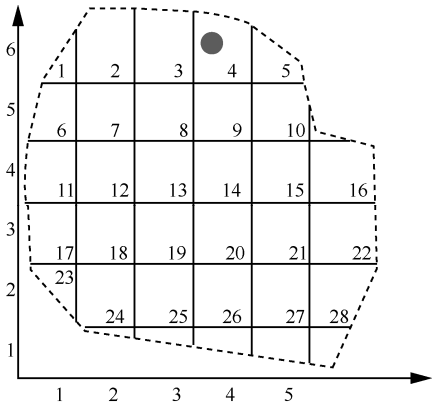



图 2　二维坐标空间

根据用户的移动模型，利用隐马尔可夫模型[24,25]来描述用户在时间不同时位置的关联性。用户将处理过的位置信息暴露给服务提供商，攻击者通过观测发送出的信息对用户的真实信息进行推断。

在时间点 $t$，利用向量 $\boldsymbol{P}^t$ 来表示用户位置的概率分布，$P_i^t=\Pr(\boldsymbol{C}_i)=\Pr(X[1],X[2])$，其中，$P_i^t$ 表示向量 $\boldsymbol{P}^t$ 中的第 $i$ 个元素，$\boldsymbol{C}_i\in Map$。通过转移矩

阵 $\boldsymbol{MA}$ 来定义用户从一个位置移动到另一个位置概率，在矩阵 $\boldsymbol{MA}$ 中存在元素 $ma_{ij}$，其中，$i$ 表示矩阵中的第 $i$ 行，$j$ 表示矩阵中的第 $j$ 列。

已知 $t-1$ 时刻用户的位置向量 $\boldsymbol{P}^{t-1}$，从而通过转移矩阵可得知 $t$ 时刻用户的位置向量 $\boldsymbol{P}^t$，$\boldsymbol{P}^t=\boldsymbol{P}^{t-1}\boldsymbol{MA}$。假设当前地图上存在 5 名移动用户，如图 2 所示，分别位于地图中的 $\{\boldsymbol{C}_2,\boldsymbol{C}_4,\boldsymbol{C}_6,\boldsymbol{C}_8,\boldsymbol{C}_9\}$，则可以得到以下的概率向量。

$$\boldsymbol{P}^t=[0\quad 0.2\quad 0\quad 0.2\quad 0\quad 0.2\quad 0\quad 0.2\quad 0.2\quad \cdots\quad 0]$$

为了更好、更准确地获取转移矩阵 $\boldsymbol{MA}$，有研究者对用户在不同时间点的具体位置信息进行了采集[26,27]，利用数据集 Geolife[22]中的位置信息，计算出转移矩阵 $\boldsymbol{MA}$。在数据集 Geolife 中，研究者收集了 182 名用户在北京三环内、每隔 1~60 s 所处的位置信息，其中每一个位置信息包含了经、纬度和当前的时间戳，历时 3 年。通过该数据库所提供的位置信息 $\boldsymbol{P}^{t-1}$ 和 $\boldsymbol{P}^t$，通过 $\boldsymbol{P}^t=\boldsymbol{P}^{t-1}\boldsymbol{MA}$ 可进一步推算出矩阵，即转移矩阵 $\boldsymbol{MA}$ 默认为已知的。

根据隐马尔可夫模型，用户的每一个状态取决于前面的有限状态，即用户在 $t$ 时刻的位置取决于 $t-1$ 时刻的位置信息。利用这一特性，通过转移矩阵获知用户下一时刻可能出现的位置，并将该位置的 POI 作为当前时刻的查询内容。其中，转移矩阵 $\boldsymbol{MA}$ 的获取是通过现有数据库的训练学习后得到的矩阵，能够利用该矩阵推测得到用户下一时刻的位置信息。

### 3.5 隐私度量

现有很多隐私保护方案采用不同的数学工具来衡量位置隐私的程度[4,7,8]，其中信息熵是最为广泛运用的一种隐私度量。利用信息熵作为隐私度量，具体定义如下。

**定义 1** 已知位置 $j$，则信息熵为

$$E=-\sum_{i=1}^{N}Pr_{i|j}^{t}\mathrm{lb}Pr_{i|j}^{t}$$

其中，$Pr_{i|j}^{t}$ 指的是在时刻 $t$，用户在位置 $j$ 处请求 $\mathrm{POI}_i$ 时的条件概率。

因为信息熵的计算并不需要过高的计算能力，现有的智能终端能够承载信息熵的计算，从而不需要和可信第三方或其他的用户终端进行交互，避免了额外的风险。该隐私度量选取的位置概率不同于现有方案单一地考虑用户在传统的背景信息下的查询概率，而是针对某一个特定的时间戳，位置 $j$ 请求查询内容为 $i$，其中 $i$ 表示某一种特定的 POI。

## 4 时空关联隐私保护方案

本节详细介绍了时空关联隐私保护方案所包含的算法，即地图分割算法和伪内容生成算法，系统结构如图 3 所示。图 3 中，用户通过移动终端向 GPS 获取位置信息，作为算法的输入，然后在移动终端中对地图进行分割，通过维诺多边形筛选出不相邻的位置区域，在伪内容生成算法中通过概率和转移矩阵构造伪位置请求，并通过移动终端发送给 LBS 服务提供商。

在用户使用 LBS 的相关服务时，在不同的需求和客观条件下，为用户选取最佳的匿名数目 $k$，例如，在 Wi-Fi 条件允许的条件下，可以选取较大的

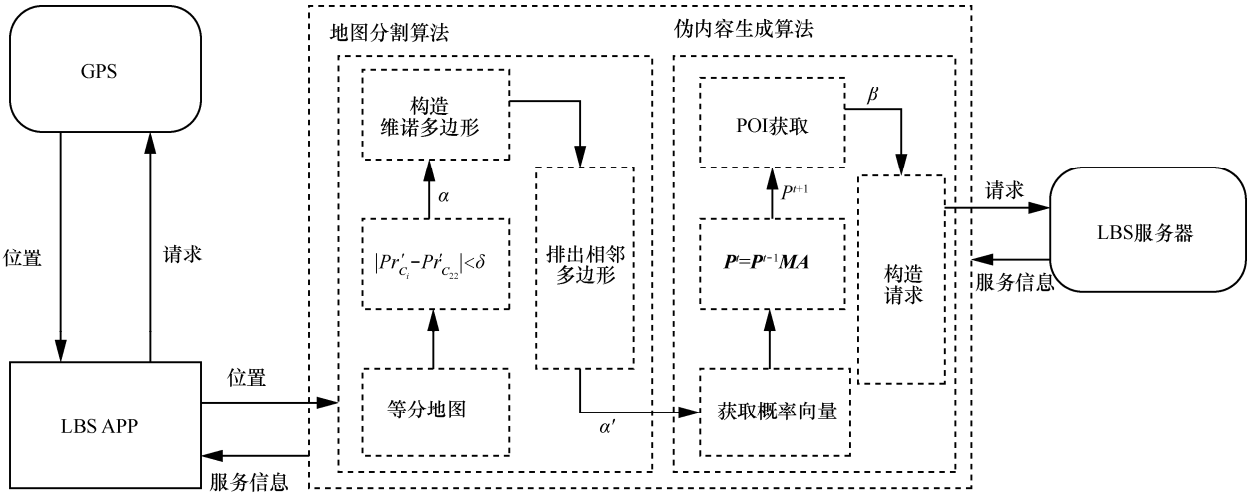
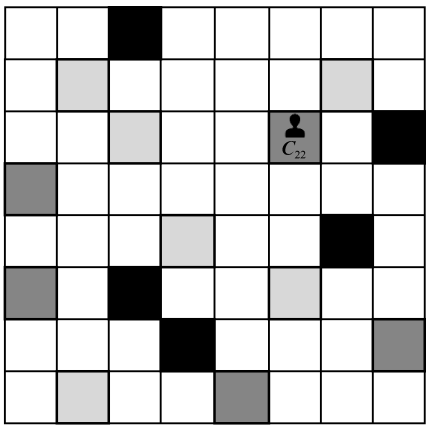



图 3 系统结构

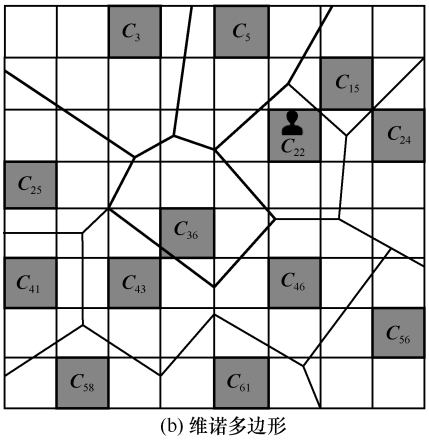
$k$ 值，而在移动数据流量时则选取较小的 $k$ 值。

### 4.1　地图分割算法

在该算法中，首先需要将地图划分为 $M$ 个位置单元，用户位于其中某一个位置单元。在 $t$ 时刻，用户请求身边 1 km 以内的 POI，如图 4(a)所示，将地图等分为 8×8 个区域，即 $M=64$，每一个区域代表一个位置单元，即 $Map=\{\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_{64}\}$。在图 4(a)中，每个单元中不同的灰度代表在 $t$ 时刻该位置单元的查询概率，用 $Pr_{\boldsymbol{C}_i}^t$ 表示。用户 Alice 位于单元 $\boldsymbol{C}_{22}$ 中，该单元在 $t$ 时刻的查询概率为 $Pr_{\boldsymbol{C}_{22}}^t$。在 $t$ 时刻 $\boldsymbol{C}_i$ 的查询概率 $Pr_{\boldsymbol{C}_i}^t$ 满足 $|Pr_{\boldsymbol{C}_i}^t - Pr_{\boldsymbol{C}_{22}}^t|<\delta$，其中 $\delta$ 为阈值，并且满足 $\delta>0$。将选取出来的位置单元存储在集合 $\alpha$ 中，该集合中的元素包含了符合要求的位置单元，即 $\alpha=\{\boldsymbol{C}_3,\boldsymbol{C}_5,\boldsymbol{C}_{15},\boldsymbol{C}_{22},\boldsymbol{C}_{24},\boldsymbol{C}_{25},\boldsymbol{C}_{36},\boldsymbol{C}_{41},\boldsymbol{C}_{43},\boldsymbol{C}_{46},\boldsymbol{C}_{56},\boldsymbol{C}_{58},\boldsymbol{C}_{61}\}$。在图 4(b)中，选取出来的位置单元用度较高的位置单元表示。




(a) 地图划分




(b) 维诺多边形

图 4　地图分割

维诺多边形对地图划分的算法采取 plane sweep，因该算法成熟且不是重点，在此并不赘述，该算法的时间复杂度分析见第 5 节。

根据集合 $\alpha$ 中的位置单元构造维诺多边形，其中每 2 个相邻的位置单元连线的中垂线构成了 $|\alpha|$ 个维诺多边形，并且每一个多边形中有且仅有唯一一个离散的位置单元。根据维诺多边形的性质，已知一个维诺多边形 $V_i$，与 $V_i$ 共享一条边的维诺多边形则是距离 $V_i$ 最近的维诺多边形，那么该多边形内的点则是距离 $V_i$ 内的离散点最近的点。在图 4(b)中，用户 Alice 所位于的维诺多边形具有 5 条边，那么共用这 5 条边的维诺多边形其中的位置单元则是和 Alice 邻近的区域。从集合 $\alpha$ 中过滤掉这 5 个相邻的多边形，剩余的位置单元构成集合 $\alpha'$，即 $|\alpha|-|\alpha'|=5$。

**算法 1**　地图分割算法

**输入**　位置信息，地图信息

**输出**　$\alpha'$

1) 等分地图；
2) 选取满足 $|Pr_{\boldsymbol{C}_i}^t - Pr_{\boldsymbol{C}_{22}}^t|<\delta$ 的区域；
3) 存储于集合 $\alpha$ 中；
4) 构造维诺多边形；
5) 获知真实用户所在的维诺多边形；
6) 删除具有相邻边的维诺多边形；
7) 筛选过的维诺多边形存储于 $\alpha$ 中；
8) $\alpha'$ 作为下一算法的输入。

### 4.2　伪内容生成算法

在该算法中，伪内容主要包含了伪查询内容和伪位置选取，在地图分割算法中得到的集合 $\alpha'$ 的基础上，伪内容算法需要选取合适的伪位置和伪查询内容。根据 3.4 节，集合 $\alpha'$ 中有元素 $\{\boldsymbol{C}_3,\boldsymbol{C}_{22},\boldsymbol{C}_{25},\boldsymbol{C}_{41},\boldsymbol{C}_{43},\boldsymbol{C}_{56},\boldsymbol{C}_{58},\boldsymbol{C}_{61}\}$，则得到概率向量 $\boldsymbol{P}^t=|0\ 0\ 0\ 0.125\ 0\ \cdots\ 0.125\ \cdots\ 0|$，其中分别在集合 $\alpha'$ 中有元素所在的位置定值为 $\frac{1}{\alpha'}$，即 $\frac{1}{8}$。已知转移矩阵 $\boldsymbol{MA}$ 和概率向量 $\boldsymbol{P}^t$，则通过 $\boldsymbol{P}^{t+1}=\boldsymbol{P}^t\boldsymbol{MA}$ 得到 $t+1$ 时刻的概率向量。通过得到的 $t+1$ 时刻的概率向量 $\boldsymbol{P}^{t+1}$ 可以得知集合 $\alpha'$ 中的位置单元可能去的位置，这些位置所对应的 POI 相当于在 $t$ 时刻想要请求的查询内容。

在该算法的例子中，假设 $k=4$，需要挑 $k-1$ 个位置单元，即 3 个位置单元作为伪位置发送给服务提供商。根据概率向量 $\boldsymbol{P}^{t+1}$ 中每个位置单元的访问可能性，从大到小降序选取 3 个位置单元存储于集合 $\beta=\{\boldsymbol{C}_a,\boldsymbol{C}_b,\boldsymbol{C}_c\}$。与地图对照，得到集合 $\beta$ 中每

一个位置单元所对应的 POI 以及和集合 $\alpha'$ 最邻近的位置单元分别构成伪查询内容 $\mathrm{POI}_a$、$\mathrm{POI}_b$、$\mathrm{POI}_c$ 和伪位置 $\boldsymbol{C}'_a$、$\boldsymbol{C}'_b$、$\boldsymbol{C}'_c$。

最后，用户 Alice 将自身的位置 $\boldsymbol{C}_i$ 和查询内容 $\mathrm{POI}_i$ 构成二元组 $[\boldsymbol{C}_i,\mathrm{POI}_i]$，同样得到伪位置请求 $[\boldsymbol{C}_a,\mathrm{POI}_a]$、$[\boldsymbol{C}_b,\mathrm{POI}_b]$、$[\boldsymbol{C}_c,\mathrm{POI}_c]$。最后通过移动智能终端将这 4 个二元组和相关信息构造成请求信息，继而发送给服务提供商来获取相关的服务信息。

**算法 2** 伪内容生成算法

**输入** $\alpha'$，$\boldsymbol{MA}$，$k$

**输出** $k$ 个查询请求

1) 已知用户的位置；
2) 当前时间为 $t-1$；
3) 通过 $\boldsymbol{P}^t=\boldsymbol{P}^{t-1}\boldsymbol{MA}$ 得到 $\boldsymbol{P}^t$；
4) 选取 $k$–1 个概率向量；
5) 存储于集合 $\beta$ 中；
6) 降序排列 $\beta$ 中元素；
7) 对应得到相应的 POI；
8) 构造请求；
9) 发送 $k$ 个请求给 LBS 服务器。

## 5 实验验证

本节分别从性能验证以及隐私验证对所提方案的性能和安全性进行测试和验证。实验利用 Matlab 在 PC 机（2.9 GHz 英特尔 i7 CPU，8 GB 内存）上进行模拟仿真并对方案的性能进行实验验证。

选取北京三环内 8 km×8 km 的地图，划分为 160×160 个位置单元，每个位置单元的大小为 50×50 的矩形。实验中参数 $k$ 指的是 $k$-匿名中发送给服务提供商的请求数量，选取范围为[3,20]。参数 $\delta$ 是地图分割算法中的阈值。

### 5.1 性能验证

1) 位置单元的数量与集合 $\alpha$ /执行时间的关系

在地图分割算法中，需要对地图中的位置单元进行过滤，从而所划分的位置单元越多执行算法所消耗的时间也就越多；随着地图中位置单元的数量增加，在构造维诺多边形时需要计算的离散点也就逐渐增加，故而最终得到的集合 $\alpha$ 的大小也随着地图划分的粒度逐渐增加，具体结果如图 5 所示，在该实验中选取参数 $k=5$。在地图划分算法中，避免用户参与过多复杂的过程，为了实现最简便的操作、最合理的隐私保护力度，地图的粒度由系统根据用户所需要的 $k$ 值的大小自动地匹配出相应的粒度。其中用户所需要的 $k$ 值和用户所处的 Wi-Fi 环境相关，由系统自动选取，不需要用户参与，大大简化了操作，方便了用户。

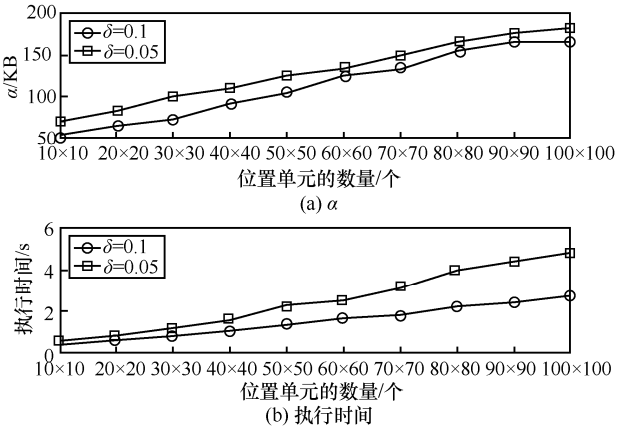



图 5 随位置单元的数量不同，$\alpha$ 与执行时间的变化情况

2) 参数 $\delta$ 与通信开销/执行时间的关系

方案利用的是伪位置生成算法，发送给服务提供商的请求信息的条数取决于参数 $k$ 的值，随着参数 $\delta$ 的变化，所发送给服务提供商的信息条数是不变的，通信开销不随 $\delta$ 的变化而变化。在算法的执行时间中，参数 $\delta$ 的值越大，则对应集合 $\alpha$ 中的元素就越多，从而在伪位置生成算法中的计算元素则越多，故而随着 $\delta$ 的增加，算法的执行时间不断增加，具体变化如图 6 所示。

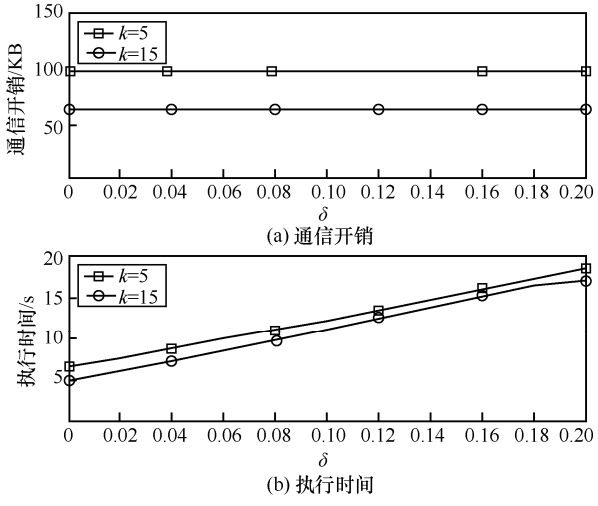



图 6 随 $\delta$ 不同，通信开销和执行时间的变化情况

3) 时间与 $\alpha$ /执行时间的关系

考虑时间和空间的关系，在现实生活中，从 0:00 开始一直到 24:00，移动用户的运动模式不同，从

而在不同的时间段有明显的概率分布，根据数据集中不同时间对应的峰值和低谷决定数据集 $\alpha$ 的大小，如图 7(a)所示。然而对于执行时间则没有明确的变化，因为地图的位置单元数目一定，$\delta$ 越大则执行时间会增加，而在不同时间点的执行时间并不会有明显的变化，如图 7 所示。

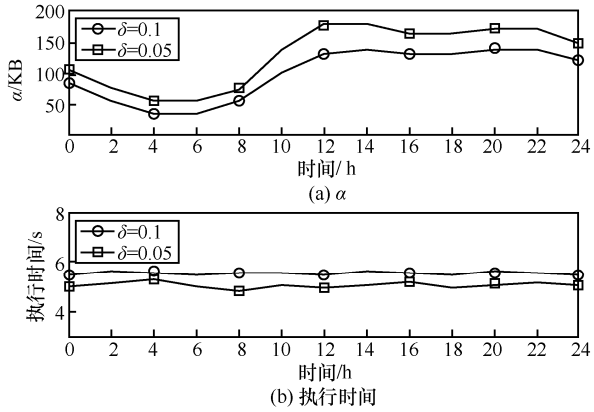



图 7　随时间不同，$\alpha$ 和执行时间的变化情况

4) 构造维诺多边形的算法复杂度

构造维诺多边形的方法分为定义法（intersect of halfplanes）、增量（incremental）算法、分治法、plane sweep 算法。采用的是 plane sweep 的方法，该方法在构造维多诺多边形可以归约为 $n$ 个实数的排序问题，则实现该算法的时间计算复杂度为 $O(n\log n)$，同时该算法在现有的 4 种算法中具有最优的完成效果。

### 5.2　隐私验证

通过比较所提隐私保护方案和现有的伪位置生成算法隐私保护方案根据参数 $k$ 的变化引起泄露概率的变化，为了保证可比性，选取同时保护位置隐私和查询隐私相类似的隐私保护方案，故而比较了 Dummy-Q[28]和 TTcloak[29]2 个方案，其中最优方案是在理论上的最优值。该方案发送给服务提供商的信息中每一个位置都会有不同的查询内容，图 8 比较了泄露概率随着参数 $k$ 的变化趋势，随着参数 $k$ 的增加泄露概率逐渐降低，因为暴露越多的请求，导致攻击者越难分析出真实的信息。图 9 比较了最优方案和随机方案的隐私度量，因为隐私度量涉及参数 $Pr_{i|j}^{t}$，并未有其他方案涉及该查询概率，故这 2 个方案的隐私度量不会随 $k$ 的变化而变化。在 TTcloak[29]方案中，为了选取相互不相邻的伪位置，将隐身区域划分为 4 等份，进而在每个等分的区域选取伪位置。当 $k$ 的取值过大时，在等分线附近容易出现伪位置过于接近的情况。该方案利用维诺多边形对地图进行划分，利用维诺多边形的性质避免了邻近位置的选取，不会出现选取的位置处于邻近区域的情况。

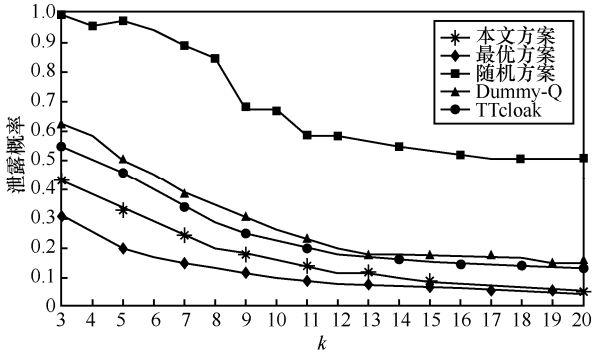



图 8　泄露概率随参数 $k$ 的变化趋势

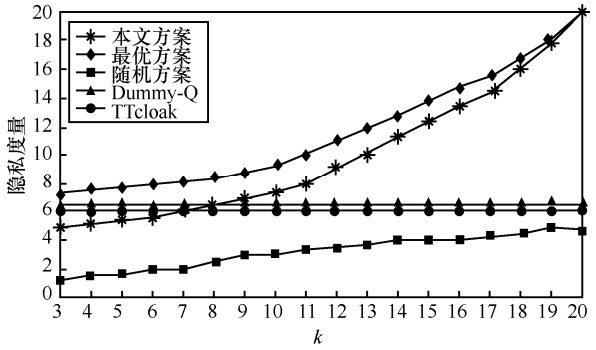



图 9　隐私度量随参数 $k$ 的变化趋势

## 6　结束语

在基于位置服务中，提出了一种时空关联的隐私保护方案，包含了 2 个算法，即地图分割算法和伪内容生成算法。地图分割算法利用维诺图将地图划分为多个维诺多边形，并且每一个多边形中有且仅有一个离散点，伪内容生成算法利用移动模型预测用户在相邻时间点中将要去的位置来选择前一个时间的查询内容，从而有效地避免了时空关联可能会导致的隐私泄露问题，同时保护了用户的位置隐私和查询隐私。最后，本文通过详细的性能和安全性实验验证了所提方案的有效性和安全性。

### 参考文献：

[1] GEDIK B, LIU L. Protecting location privacy with personalized *k*-anonymity: architecture and algorithms[J]. IEEE Transactions on Mobile Computing, 2008, 7(1): 1-18.

[2] PINGLY A, YU W, ZHANG N, FU X, et al. Cap: a context-aware privacy protection system for location-based services[C]// ICDCS.

2009: 49-57.
[3] LI X, ZHANG C, JUNG T, QIAN J, et al. Graph-based privacy-preserving data publication[C]//INFOCOM.2016.
[4] CHEN Z, HU X, JU X, et al. Lisa: location information scrambler for privacy protection on smartphones[C]//CNS. 2013.
[5] ANDRES M, BORDENABE N, CHATZIKOKOLAKIS K, et al. Geo-indistinguishability: differential privacy for location-based systems[C]// CCS. 2013.
[6] PERAZZO P, DINI G. A uniformity-based approach to location privacy[J]. Computer Communications, 2015, 64(1): 21-32.
[7] NIU B, LI Q, ZHU X, et al. Achieving *k*-anonymity in privacy-aware location-based services[C]// INFOCOM.2014.
[8] NIU B, LI Q, ZHU X, et al. Enhancing privacy through caching in location-based services[C]// INFOCOM. 2015.
[9] SHOKRI R, THEODORAKOPOULOS G, PAPADIMITRATIS 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.
[10] 黄毅, 霍峥, 孟小峰. CoPrivacy: 一种用户协作无匿名区域的位置隐私保护方法[J]. 计算机学报, 2011, 34(10): 1976-1985.
HUANG Y, HUO Z, MENG X F. CoPrivacy: a collaborative location privacy-preserving method without cloaking region[J].Chinese Journal of Comuputers, 2011, 34(10): 1976-1985.
[11] 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, PP (99): 1-14.
[12] PAULET R, KAOSAR M, YI X, et al. Privacy-preserving and content-protecting location based queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(5): 1200-1210.
[13] CHOW C, MOKBEL M, LIU X. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments[J]. GeoInformatica, 2011, 15(2): 351-380.
[14] MOKEL M, CHOW C, AREF W. The new casper: query processing for location services without compromising privacy[J]. The IEEE VLDB Journal 2006.
[15] ZHANG X, XIA Y, BAE H. A novel location privacy preservation method for moving object[J]. International Journal of Security and Its Applications, 2015, 9(2): 1-12.
[16] BERESFORD A, STAJANO F. Location privacy in pervasive computing[J]. IEEE Pervasive Computing, 2003, 2(1):46-55.
[17] 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.
[18] MEYEROWITZ J, CHOUDHURY R. Hiding stars with fireworks: location privacy through camouflage[C]//MobiCom.2009.
[19] XU T, CAI T. Feeling-based location privacy protection for location-based services[C]//CCS. 2009.
[20] MONTAZERI Z, HOUMANSADR A, PISHRO H. Achieving perfect location privacy in wireless devices using anonymization[J]. IEEE Transactions on Information Forensics and Security, 2017,PP (99): 1.
[21] WANG X, PANDE A, ZHU J, et al. Stamp: enabling privacy-preserving location proofs for mobile users[J]. IEEE/ACM Transactions on Networking, 2016,24(6):3276-3289.
[22] FAWAZ K, SHIN K. Location privacy protection for smartphone users[C]//CCS. 2014.
[23] SHOKRI R, THEODORAKOPOULOS G, TRONCOSO C, et al. Protecting location privacy: optimal strategy against localization attacks[C]//CCS. 2012.
[24] GOTE M, NATH S, GEHRKE J. Maskit: privately releasing user context streams for personalized mobile applications[C]//SIGMOD. 2012.
[25] SHOKRI R, THEODORAKOPOULOS G, BOUDEC J, et al. Hubaux. Quantifying location privacy[C]//SP. 2011.
[26] ZHENG Y, XIE X, MA W. Geolife: a collaborative social networking service among user, location and trajectory[C]//Data Eng. Bull.2010.
[27] CHO E, MYERS S, LESKOVEC J. Friendship and mobility: user movement in location-based social networks[C]//KDD. 2011.
[28] PINGLEY A, ZHANG N, FU X, et al. Protection of query privacy for continuous location based services[C]// INFOCOM.2011.
[29] NIU B, ZHU X, LI W, et al. A personalized two-tier cloaking scheme for privacy aware location-based services[C]//ICNC.2015.

[作者简介]

**李维皓**（1990-），女，辽宁沈阳人，西安电子科技大学博士生，主要研究方向为社交网络中的隐私保护。

**丁晟**（1990-），男，陕西西安人，西安电子科技大学博士生，主要研究方向为数据安全与隐私保护。

**孟佳洁**（1985-），女，河北邯郸人，西安电子科技大学博士生，主要研究方向为流量监测与隐私保护。

**李晖**（1968-），男，河南灵宝人，博士，西安电子科技大学教授、博士生导师，主要研究方向为密码学、无线网络安全、云计算安全、信息论与编码理论。