DOI：10.7652/xjtuxb201612013

用户为中心的差分扰动位置隐私保护方法

张学军[1,2,3]，桂小林[1]，蒋精华[1,4]
（1.西安交通大学电子与信息工程学院，710049，西安；2.兰州交通大学电子与信息工程学院，730070，兰州；
3.兰州交通大学光电技术与智能控制教育部重点实验室，730070，兰州；4.香港城市大学计算机系，999077，香港）

摘要：基于隐形区的位置混淆技术是实现位置隐私广泛研究的技术，但该技术需要可信第三方且无法防止基于背景信息的推理攻击，容易泄露位置隐私。针对这一难题，提出了以用户为中心的差分扰动位置隐私保护方法，不需要可信第三方，同时增强了用户位置隐私。该方法采用修改的 Hilbert 曲线映射技术将地图中用户的每个位置投影到一维空间，通过组合 k 匿名和差分隐私技术随机产生扰动，并将扰动位置作为用户真实位置提交给服务商。为了解决移动设备资源受限问题，采用基于四分树的方法将用户的上下文存储和转换为比特流，由此获得了有效的时空复杂度和很高的检索准确率。安全分析表明，该方法能有效保护用户位置隐私；实验评估表明，与采用标准 Hilbert 曲线映射的方法相比检索准确率平均提高了 15.4%。所提方法在隐私保护和服务精度之间取得了较好的权衡，对隐私保护系统设计具有一定的理论和实际意义。

关键词：位置服务；差分隐私；隐私保护；背景信息；位置稳私
中图分类号：TP309 **文献标志码**：A **文章编号**：0253-987X(2016)12-0079-08

A User-Centric Location Privacy-Preserving Method with Differential Perturbation for Location-Based Services

ZHANG Xuejun[1,2,3], GUI Xiaolin[1], JIANG Jinghua[1,4]
(1. School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
3. Key Laboratory of Opto-Technology and Intelligent Control Ministry of Education, Lanzhou Jiaotong University, Lanzhou 730070, China; 4. Department of Computer Science, City University of Hong Kong, Hong Kong 999077, China)

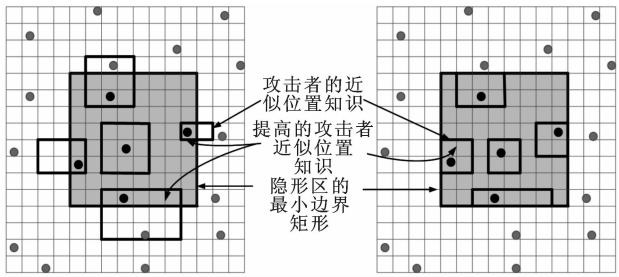
Abstract: A user-centric location privacy-preserving method with differential perturbations (Ulp2mDP) is proposed to solve the problem that the location obfuscation technique using cloaking region requires a trusted third part (TTP) and cannot sufficiently prevent inference attacks based on background information, and hence is easy to leak location privacy. The method can enhance the user's location privacy without requiring a TTP. The Ulp2mDP uses a modified Hilbert curve to project each 2-D geographical location of user into a 1-D space, and then randomly generates a reasonable perturbed location by combining the k anonymity with differential privacy technique. The perturbed value is then submitted as the user's real location to the service provider. In order to address the resource limitation of mobile devices, a quad-tree based scheme is used to transform and to store the user's context information as bit stream,

收稿日期：2016-08-08。 作者简介：张学军(1977—)，男，博士生；桂小林(通信作者)，男，教授，博士生导师。 基金项目：国家自然科学基金资助项目(61472316)；陕西省科技计划资助项目(2016ZDJC-05，2013SZS16-Z01)；兰州交通大学光电技术与智能控制教育部重点实验室开放课题资助项目(KFKT2016-7)；甘肃省自然科学基金资助项目(2016GS07203)。
网络出版时间：2016-10-19 网络出版地址：http://www.cnki.net/kcms/detail/61.1069.T.20161019.1622.006.html

which are highly efficient with respect to time and space complexities, hence to achieves high precision of retrieval. Security analysis shows that the Ulp2mDP can effectively protect user's location privacy. Experimental evaluation and a comparison with the approach using standard Hilbert curve show that the average retrieval accuracy of the Ulp2mDP increases by 15.4%. It is concluded that the Ulp2mDP provides a tradeoff between privacy preserving and service accuracy, and has a certain theoretical and practical significance for the design of privacy-preserving systems.

Keywords: location-based service; differential privacy; privacy preserving; background information; location privacy

随着智能移动设备和无线通信技术的快速发展,位置服务(location-based service,LBS)给社会和个人提供了巨大的便利,但也引发了严重的隐私关注。为了让用户放心使用 LBS,需要实施隐私保护,防止用户隐私泄露,然而实施隐私保护必然会降低服务质量。因此,如何保护用户隐私而又不妨碍服务的可用性是一个重要的挑战[1-2]。为了面对这个挑战,该领域的研究者提出了众多解决方案,其中基于隐形区的位置混淆方法[3]是研究最广泛的技术,它使用 k 匿名指标并依赖可信第三方(trusted thirty part,TTP)。为了实现 k 匿名,位置混淆方法将包含用户真实位置的查询通过 TTP 提交给位置服务商(location-based service provider,LSP),TTP 将用户真实位置扩大为地理上包含其他至少 $k-1$ 个用户的隐形区。因此,不可信的 LSP 很难把用户的真实位置和其他 $k-1$ 个用户的位置区分开。虽然使用 k 匿名方法可以增强用户的位置隐私,但仍存在一些缺点。首先,该方法严重依赖 TTP,TTP 容易成为系统瓶颈,而且 TTP 知道所有用户 LBS 查询的全部知识,易遭受单点故障。如果攻击者俘获了 TTP,则所有用户的隐私都将泄露。新近的研究[4]试图通过使用网格标识匹配和保序对称加密技术解决这一问题,但是需要改变客户端、TTP 和服务端的系统模型且客户端计算开销很大。其次,大部分已有的方法[3-4]都假定攻击者没有关于用户的背景信息,如近似位置知识、与位置相关的查询频率、用户属性等。实际中,因为一些攻击者可能拥有这样的一些背景信息,所以这些方法不能很好地保护位置隐私。例如,图 1a 中拥有近似位置知识的攻击者利用隐形区提高了多个关于用户近似位置的精确度。因此,隐形区能给攻击者提供额外知识,会导致位置隐私泄露。理想情况下,如图 1b 所示,仅当隐形区能确保包含每个用户对应的近似位置时,由隐形区造成的隐私泄露问题才能被消除。不幸的是,很难判断攻击者拥有知识的范围和程度。最后,用户分布稀疏时,很难在一个合理的隐形区中发现足够的用户。为此,隐形区会出现不必要的扩大,最坏情况下用户的服务将会被拒绝。为了避免 TTP,一些基于移动设备的隐私保护方法,如 CloakP2P[5]、MobileHide[6]等,被引入到 LBS 隐私保护系统中,但是它们仍需使用隐形区,且要求用户相互通信以获取额外信息。



(a)隐私攻击结果 (b)隐私理想保护结果

图 1 使用隐形区的位置隐私攻击与理想保护结果

哑元技术[7]不使用隐形区,但当攻击者具有背景信息时,它不能保证位置隐私。哑位置选择算法[8]和缓存感知的哑位置选择算法[9]利用虚假信息和缓存等技术来存储历史查询信息,以便重复利用,提高了隐私保护强度,然而它们会带来很高的计算和通信开销。上下文感知位置隐私保护方法[10]在解决时空复杂度上非常有效,但是当攻击者知道扰动算法和一些背景信息(如包含用户真实位置的位置点集合)时,它就不能很好地保护位置隐私。局部差分扰动技术[11]可以抵御背景信息攻击,但仍需要 TTP。差分隐私技术[12]不使用 TTP 且无需考虑攻击者的背景信息,能针对攻击者的先验知识进行有效保护。目前,大部分差分隐私技术主要应用于离线数据的隐私保护,针对 LBS 场景的隐私保护方法较少见且面临许多新的挑战,需要继续研究和完善。因此,如何设计一个适合移动设备、鲁棒性很强的

LBS 隐私保护系统仍然是一个很大的挑战性问题。

针对这一挑战，本文提出了一个能抵御背景信息攻击且无需任何 TTP 的差分扰动位置隐私保护方法(Ulp2mDP)。Ulp2mDP 方法仅运行在移动设备上，能够防止具有背景信息的攻击者获取用户的位置隐私。Ulp2mDP 方法首先使用修改的 Hilbert 曲线(modified hilbert curve，MHC)将地图中用户的每个二维地理位置投影到一维空间，然后利用空间 k 匿名互惠框架[13]选择 k 个位置，接着按期望的隐私水平通过向 k 个位置中添加可控的拉普拉斯噪声随机产生扰动，并用扰动位置代替用户的真实位置来请求 LBS。

Ulp2mDP 方法旨在确保隐私保护和 LBS 精度，由于其采用用户为中心的模式，所以面临移动设备资源受限的挑战。为解决这一问题，使用基于四分树的模式将用户位置的上下文转换和存储为比特流，由此可获得高压缩率、支持有效检索。由于 MHC 映射基于真实地图离线计算，仅需很小的存储空间和检索成本，在解决时间和空间复杂度上很有效[10]。

1　Ulp2mDP 方法的设计思想

1.1　系统模型与背景信息

Ulp2mDP 方法的系统模型主要由 LBS 用户和 LSP 两部分构成，如图 2 所示。LBS 用户拥有位置感知无线设备，能通过 WiFi、GPRS 或 3G 等无线协议连接移动互联网。LBS 用户将使用位置扰动算法生成的扰动查询提交给 LSP。LSP 不可信，假定为攻击者，它会响应用户的扰动查询并返回查询结果。LSP 通过观察接收的扰动查询，获取关于用户的所有背景信息。此外，攻击者还知道扰动算法和系统使用的噪声分布。基于这些背景信息，LSP 通过执行推理攻击尽力推断用户的真实位置。

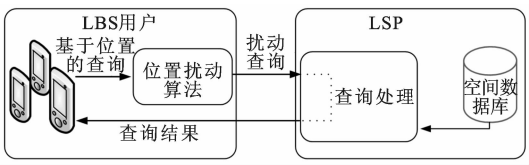


图 2　Ulp2mDP 方法的系统模型

本文将攻击者的背景信息限定为关于用户的近似位置知识(一些包含用户真实位置的位置点集合)，它能够通过多种方式获得，例如使用蜂窝塔的设备通信日志、违规停车的公共记录、随意谈话中的社会工程方法等。除非有法律规定，否则近似位置知识能够很容易地从蜂窝塔和无线访问点的广播信息中直接推算出来[11]。

1.2　设计思想与动机

Ulp2mD 方法的设计源于下面的观察及其可能的影响。具体地，携带智能手机的用户在某个地理区域移动时查询 LBS，以定位他们感兴趣的周围兴趣点(例如饭店或地铁站等)。假定 LSP 诚实且好奇，即它能准确回答用户的查询，但希望通过分析用户访问的位置收集用户的隐私信息。

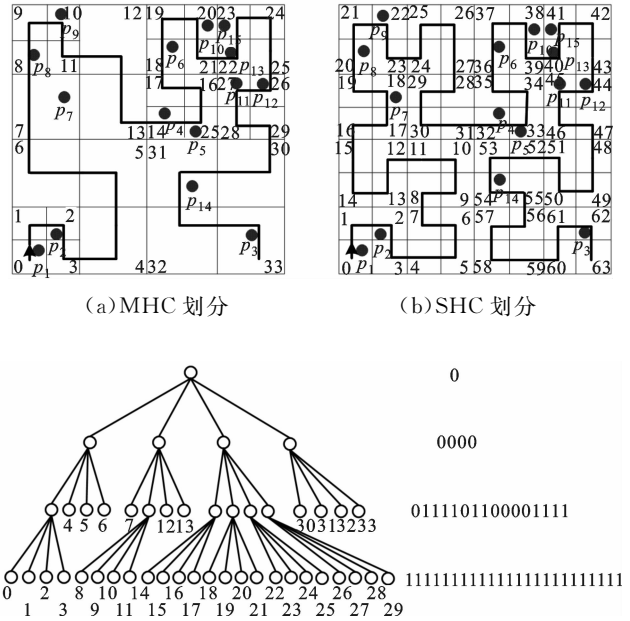
在 LBS 应用(例如 Foursquare、WeChat 等)中，用户总是倾向于从对他们有意义的地方(例如家、办公室等)提交查询。在这些地方，用户通常更有可能执行某些活动而没有太多的运动[14]。我们称这些地方为用户的兴趣点，并定义其为地理空间中的真实兴趣点，假定用户从他们的兴趣点请求 LBS。本文主要关注保护用户请求 LBS 的兴趣点。事实上，对兴趣点的推理会造成非常严重的位置隐私侵害[15]。例如，推理兴趣点可能会让好奇的 LSP 发现用户的家庭住址和工作地点，甚至推断出用户的宗教和政治倾向。假定每个兴趣点 ψ_i 都近似的表示为 (x_i, y_i, ζ_i)，其中 (x_i, y_i) 表示兴趣点的位置坐标，ζ_i 代表兴趣点位置坐标 (x_i, y_i) 的语义属性，例如它的语义位置。

有效保护位置隐私的一种简单方法是基于经纬度上预先确定的噪声分布来随机扰动用户位置。但是，这种方法对 LBS 不充分，因为有了预先确定的噪声分布，隐私保护水平和 LBS 精度很大程度上依赖用户所在位置的上下文。例如，直觉上，为获得同样的隐私水平和 LBS 精度，在郊区扰动位置偏离用户的位置要比在市区更远，因为市区的路网和兴趣点密度分布要比郊区密集很多。

位置隐私保护的一个关键挑战是如何实现上下文感知的隐私保护。已有的 k 匿名技术通过 TTP 上用户分布的全局知识实现上下文感知，例如隐形区在人口稀少的郊区会自动变得更大。没有了 TTP，就必须从其他来源获得上下文。为此，本文设计了基于兴趣点密度分布的 MHC 映射技术来获得上下文。MHC 将地理区域中的每个二维兴趣点投影到一维空间，使一维空间的每个数据点都有同质的上下文(如相同的密度)，而且原来相邻的兴趣点在投影后仍然保持邻近。图 3a 为基于兴趣点密度分布的 MHC 映射技术对地理区域进行分割的一个实例，图中的圆点 $P_i(i=1,\cdots,15)$ 为兴趣点，每个原子单元中的数字代表该原子单元的 Hilbert

值。通常,兴趣点密集的区域会导致更细的粒度划分,这样一维空间中的每个数据点都有相同的密度。相应地,图 3b 为使用标准 Hilbert 曲线(standard hilbert curve,SHC)映射技术对空间进行分割的一个实例,由于 SHC 使用了相同的粒度分割区域,所以它不能反映兴趣点密度分布特点。

Hilbert 曲线的最优数据聚类和距离保持特性使得一维空间中相邻的两个点更有可能在原来空间也相邻。因此,给定一个特定的点,能很容易地发现其周围邻近的点。根据这一属性,用户请求 LBS 时,首先使用 MHC 投影其兴趣点到一维空间,然后利用 k 匿名互惠框架[13]选择包含自己兴趣点在内的 k 个兴趣点,并向 k 个兴趣点中添加仔细选择的拉普拉斯噪声产生扰动,保证由 k 个兴趣点分别计算得到同样扰动的概率相似(达到 e^{ε})。这样,具有近似位置知识的攻击者很难推断出用户的真实兴趣点。



(a)MHC 划分 (b)SHC 划分

(c)MHC 划分区域的四分树表示 (d)四分树的序列化存储

图 3 MHC 和 SHC 映射技术

1.3 添加拉普拉斯噪声扰动

形式上让 $l_1,\cdots,l_k$ 表示 k 个兴趣点(k 为兴趣点数),其中的一个是用户的真实兴趣点 $l_r=(x_r,y_r)$。$l_p=(x_p,y_p)$是 l_r 对应的扰动兴趣点。对于这 k 个兴趣点中的任意两个兴趣点 l_i 和 l_j,它们生成扰动兴趣点 $l_p=(x_p,y_p)$的概率应满足下式

$$P(x_p \mid x_i) \leqslant e^{\varepsilon}P(x_p \mid x_j) \tag{1}$$

$$P(y_p \mid y_i) \leqslant e^{\varepsilon}P(y_p \mid y_j) \tag{2}$$

式中:$\varepsilon \geqslant 0$;$i,j \in \{1,\cdots,k\}$;$P(x_p \mid x_i)$为 x_i 对应 x_p 的概率;$P(y_p \mid y_i)$为 y_i 对应 y_p 的概率。这一属性可以通过使用拉普拉斯噪声扰动 $l_r=(x_r,y_r)$获得。研究表明,分别向每个坐标点添加噪声提供的隐私保护比分别向每个位置点添加噪声更强[16]。因此,采用分别向每个坐标添加尺度参数为 b 的拉普拉斯分布,使得独立扰动 $l_i=(x_i,y_i)$的坐标 x_i 和 y_i 满足

$$P(x_p \mid x_i) = \frac{1}{2b}e^{-\frac{|x_i-x_p|}{b}} \tag{3}$$

$$P(y_p \mid y_i) = \frac{1}{2b}e^{-\frac{|y_i-y_p|}{b}} \tag{4}$$

每个坐标添加的噪声数量为 $-b\mathrm{sign}(r_{nd})\ln(1-2|r_{nd}|)$,$r_{nd}$ 是$[-1/2,1/2]$中的一个均匀随机值。基于下面的观察,将 b 设为$(\max_n x_n-\min_n x_n)/\varepsilon$ 产生 x_p、$(\max_n y_n-\min_n y_n)/\varepsilon$ 产生 y_p,其中 $\max_n x_n$ 表示 $x_1,\cdots,x_n$ 中的最大值;$\min_n x_n$ 表示 $x_1,\cdots,x_n$ 中的最小值。由此,可获得扰动兴趣点 $l_p=(x_p,y_p)$。

不失一般性,对于兴趣点 l_i、l_j、l_p,由三角不等式可得

$$|l_j-l_p| \leqslant |l_j-l_i|+|l_i-l_p| \tag{5}$$

将式(5)重新整理,两边除以 b 并将其提升为底数为 e 的幂指数,然后再在两边乘以 $1/2b$ 可得

$$\frac{1}{2b}e^{-\frac{|l_i-l_p|}{b}} \leqslant \frac{1}{2b}e^{-\frac{|l_j-l_p|}{b}}e^{-\frac{|l_i-l_j|}{b}} \tag{6}$$

由式(3)和式(4),可将式(6)修改为

$$P(l_p \mid l_i) \leqslant P(l_p \mid l_j)e^{\frac{|l_j-l_i|}{b}} \tag{7}$$

对每个坐标 x_i 和 y_i,式(7)相应地可表示为

$$P(x_p \mid x_i) \leqslant e^{\frac{|x_j-x_i|}{b}}P(x_p \mid x_j) \tag{8}$$

$$P(y_p \mid y_i) \leqslant e^{\frac{|y_j-y_i|}{b}}P(y_p \mid y_j) \tag{9}$$

分别设定式(8)和式(9)的指数边界,可得

$$P(x_p \mid x_i) \leqslant e^{\frac{|\max_n x_n-\min_n x_n|}{b}}P(x_p \mid x_j) \tag{10}$$

$$P(y_p \mid y_i) \leqslant e^{\frac{|\max_n y_n-\min_n y_n|}{b}}P(y_p \mid y_j) \tag{11}$$

因此,只要将 b 设为 k 个兴趣点中任意两个兴趣点各自坐标分量的最大距离,则这 k 个兴趣点集合中任意两个兴趣点产生特定扰动的概率总是被限定在常数因子 e^{ε} 内,即任意两个兴趣点产生扰动的概率相似。

2 Ulp2mDP 实现方法

2.1 基于兴趣点密度分布的 MHC 映射方法

MHC 映射方法的实现需要参考上下文信息,

如兴趣点或路网密度。一般而言，路网密度和兴趣点密度强相关，路网密度越大的区域兴趣点密度也越大。因此，MHC 映射技术可以很容易地适应于路网密度。下面给出 MHC 映射技术的构建方法。

令 R 是用户移动的地理区域的边界，Ψ 是 R 中所有可能的兴趣点的集合。不失一般性，认为用户移动区域 R 是一个矩形区域。当且仅当原区域中的兴趣点数大于预定义的阈值 σ 时，可将原区域递归地划分为 4 个相等大小的子区域。图 3a 是这种划分的一个例子，它根据给定的阈值 $\sigma=1$ 对区域进行递归划分，直到区域中的兴趣点数不超过给定的阈值 σ 为止。从图中可以看到，MHC 划分覆盖了具有细粒度的高密度兴趣点区域，每个原子单元，即不能进一步划分的单元，包含大约 σ 或更少的兴趣点。如图 3c 所示，这种划分很容易表示为一棵四分树。特别地，树中的每个节点仅有两种可能：叶子节点或包含 4 个孩子的节点。为了有效地存储树，可以构建一棵广度优先搜索树，并为每个节点存储 1 bit的信息，表明该节点是否为叶子节点，按此可以将用户移动空间转化为序列化的地图存储文件。图 3d 给出了图 3c 中四分树表示地图的序列化存储的一个例子。假定兴趣点集合的规模为 n，则生成叶子节点的个数为 n/σ。一棵具有 n/σ 个叶子节点的四分树最多只有 $4n/3\sigma$ 个节点，那么序列化文件需要的空间最多是 $4n/3\sigma$ 个 bit，因此文件的存储开销是 $O(n)$。由于某个地理区域内的兴趣点数在一定时期内不会发生太大变化，所以 MHC 的构建可以通过离线方式实现。为了给每个原子单元分配一维空间中的 Hilbert 值，需要对四分树进行深度优先遍历，按照每个叶子节点的遍历先后次序分配一维空间中兴趣点的 Hilbert 值。如图 3c 所示，叶子节点下面的数字代表每个叶子节点被访问的先后次序，即为该叶子节点对应原子单元的 Hilbert 值，也是该原子单元中兴趣点的 Hilbert 值。例如，在图 3a 中，包含兴趣点 p_3、p_4 的原子单元的 Hilbert 值为 33 和 14，相应的兴趣点 p_3、p_4 的 Hilbert 值也为 33 和 14。计算各个兴趣点 Hilbert 值的时间复杂度是 $O(n)$。

2.2　位置差分扰动

由 1.3 节分析可知，位置差分扰动算法的思想为：首先在 MHC 映射的基础上使用空间 k 匿名互惠框架[13]生成 k 个兴趣点，然后使用差分隐私技术向这 k 个兴趣点中添加拉普拉斯噪声，以确保这 k 个兴趣点中任意两个兴趣点产生同一扰动的概率总是被限定在常数因子 e^{ε} 内，从而使得具有近似位置知识的攻击者很难推理出用户的真实兴趣点。显然，k 个兴趣点中任何一个兴趣点都可用于产生扰动，本文选择所有其他 $k-1$ 个兴趣点平均距离最小的兴趣点作为用户的扰动。产生扰动兴趣点的时间复杂度是 $O(\log_4 n)$。

3　Ulp2mDP 安全性分析

有两种类型的攻击者：主动攻击者和被动攻击者。被动攻击者经常基于获取的侦听信息进行修改、重放或注入攻击。实际中，这种类型的攻击可以通过使用一些成熟的加密工具加以避免，如公钥基础设施。因此，我们主要关注如何避免来自主动攻击者的共谋攻击和推理攻击，这两种攻击会引发严重的隐私泄密问题。主动攻击者能俘获 LSP，而且能获得移动用户的所有信息，通过执行共谋攻击和推理攻击获得特定用户的敏感信息。

定理 1　Ulp2mDP 方法是抗共谋攻击的。

证明　共谋攻击经常发生在一组用户之间。然而，在 Ulp2mDP 方法中因为不存在和任何其他用户的交互，所以与用户共谋对其他用户没有任何影响。因此，Ulp2mDP 方法是抗共谋攻击的。这种类型攻击者的最好情况是能通过俘获 LSP 得到所有信息，这时他就变成了一个主动攻击者执行推理攻击。

定理 2　Ulp2mDP 方法是抗推理攻击的。

证明　在 Ulp2mDP 方法中，为了享受 LBS，用户需要向攻击者提交 LBS 查询。理想情况下，由于扰动，攻击者不能识别扰动位置和用户之间的任何直接连接。然而，攻击者知道整个地图上的兴趣点密度、关于用户的近似位置知识以及噪声分布。基于这些信息，攻击者能够执行推理性攻击来获得查询用户的真实位置。具体来讲，攻击者知道所有的兴趣点集合 Ψ 等可能的位置点集合 $l_1,\cdots,l_r,\cdots,l_k$，位置扰动算法以及噪声分布 $P(l_i)$。因为在知道噪声分布的情况下，攻击者近似知识中的某些位置是极不可能产生扰动的，所以攻击者就要利用新了解的信息分布来提高从这些可能的位置点集合 $l_1,\cdots,l_r,\cdots,l_k$ 中成功猜测真实位置 l_r 的概率。在本文算法中，推理攻击可以通过使用拉普拉斯分布来避免。由于使用了尺度参数 $b\geqslant 0$ 的拉普拉斯分布，从位置点集合 $l_1,\cdots,l_r,\cdots,l_k$ 中计算出同样扰动位置 l_p 的概率被限定在一个很小的常数因子 e^{ε} 范围内。添加到某个位置点上的拉普拉斯噪声的数量取决于这 k 个位置点中任意两个位置的各自坐标分量的最大

距离，只要尺度参数 b 使用了这个最大距离，扰动 $l_p=(x_p,y_p)$ 就会满足概率比 e^{ε}，而且被选中的 k 个位置保持了互惠性。不管哪一个兴趣点是查询点，通过扰动算法都会得到同样的匿名集。因此，给定扰动 $l_p=(x_p,y_p)$，不管哪个兴趣点被用于执行扰动，k 个兴趣点产生扰动的概率都是相同的(在常数因子 e^{ε} 内)，也就是说攻击者不能使用这些信息提高他成功猜测用户真实位置的概率。

4 实验评测与结果分析

4.1 实验环境与方法

算法使用 Java 语言实现，并在 Intel Core i7-4790 3.6 GHz 处理器、4 GB 内存、Windows OS 计算机上完成，实验结果为执行 100 次相应算法的平均值。实验中的兴趣点集合 Ψ 取自相应的北美道路网络兴趣点集合数据集 NA(http://www.cs.utah.edu/~lifeifei/SpatialDataset.htm)。

4.2 MHC 和 SHC 映射技术的参数选择

MHC 映射技术会依据曲线的遍历顺序给每个原子单元分配唯一的 Hilbert 值，包含在原子单元内的兴趣点的索引就是该原子单元的 Hilbert 值。如果原子单元内包含多个兴趣点，则其索引值都相同。使用重叠因子 λ 量化兴趣点索引值相同的情况

$$\lambda=\frac{1}{M}\sum_{i=0}^{H}n_i \tag{12}$$

式中：M 为包含兴趣点的原子单元的数量；H 为填充曲线索引值的上限；n_i 是索引值等于 i 的兴趣点数量。当设置相同的 λ 值时，MHC 和 SHC 映射技术采用的 σ、曲线阶数 D 及生成兴趣点索引的时间见表 1。

表 1 参数选择与索引生成时间

算法	索引生成时间/ms			
	$\lambda=1$	$\lambda=1.5$	$\lambda=2.7$	$\lambda=4.9$
	$\sigma=1,D=13$	$\sigma=2,D=11$	$\sigma=5,D=10$	$\sigma=10,D=9$
SHC	1 237	1 067	988	903
MHC	892	521	300	214

由表 1 中可以看出，在相同 λ 下，MHC 映射技术生成索引的效率大大高于 SHC，而且随着 λ 的增大，MHC 的效果更明显。因为 MHC 考虑了兴趣点的密度分布，对不同密度的区域采用不同的曲线阶数，使得对密度较低区域的划分不使用过高的曲线阶数，从而提高了索引的计算效率。下面的实验取参数 $\lambda=4.9$、$\sigma=10$、$D=9$ 进行。

4.3 k 匿名集生成效率评价

图 4 给出了采用 MHC 和 SHC 映射技术的 k 匿名集生成时间 t 的比较。

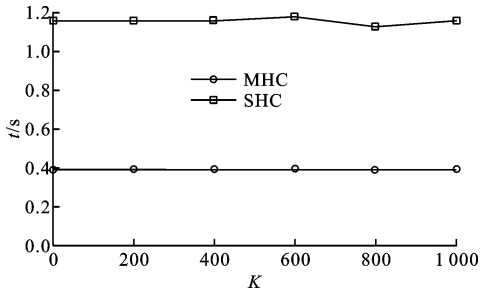


图 4 k 匿名集的生成时间的比较

从图 4 中可以看出，随着 k 值的增加，使用 MHC 和 SHC 映射技术生成 k 匿名集的时间没有显著地变化。这是因为在扰动算法中，选择满足互惠性的 k 个兴趣点仅需遍历四分树中很小的局部子树，而且由于四分树中间节点的数据结构中包含其子树中所有兴趣点的数目，所以不需要再遍历其子树来得到这些信息。在相同 k 值的情况下，SHC 生成匿名集的时间远多于 MHC，平均多出了约 66%，这是因为 SHC 映射对所有区域都采用统一的粒度进行划分，没有考虑用户的上下文信息。在生成匿名集时，相同情况下 MHC 要比 SHC 遍历更少的子树，从而匿名集的生成时间有了较大幅度的降低。

4.4 位置扰动性能评价

阈值 σ 决定了四分树的大小，也决定了序列化地图存储文件的大小。随着 σ 的增大，生成的序列化地图文件仅需很小的存储空间。$\sigma=10$ 时，MHC 生成扰动的时间小于 1 ms，检索时间不超过 2 s。

隐形区可以确保查询结果中包含对应用户真实位置的检索结果，而由扰动位置查询的结果则不一定包含对应用户真实位置的检索结果，这种检索结果的差异是否存在取决于扰动位置和真实位置间的距离。本文采用以下 3 个指标评价算法的有效性。

(1)接近度。接近度指标表示扰动接近于真实位置的比率。

(2)相似度。假设 $o=\{o_1,o_2,\cdots,o_K\}$ 是相对于用户真实兴趣点位置 l_r 的 K 最近邻检索对象结果集，$o'=\{o'_1,o'_2,\cdots,o'_K\}$ 是相对于扰动兴趣点位置 l_p 的 K 最近邻检索对象结果集，则查询结果的相似度 ρ 可表示为 o 和 o' 间共同对象的比率，定义为

$$\rho=\frac{|o\cap o'|}{|o|} \tag{13}$$

式中：$|o|$ 是真实结果集 o 中的查询对象的数量，

$|o \cap o'|$ 是 o 和 o' 间共同对象的数量。

(3)偏移度。偏移度 μ 度量 K 最近邻检索的实际检索结果和真实检索结果在距离方面的差异,定义如下式

$$\mu = \begin{cases} \frac{1}{K}\left(\sum_{i=1}^{K} \mathrm{dist}(o'_i, q) - \sum_{i=1}^{K} \mathrm{dist}(o, q)\right), & o \neq o' \\ 0, & o = o' \end{cases} \tag{14}$$

式中:q 为查询用户的真实查询兴趣点,dist(·)为欧式距离函数。

对于接近度,记扰动位置邻近真实位置的距离为 d,计算 ε 为不同值时 d 小于等于 1 000 m、500 m、100 m 的平均扰动百分比,结果见表 2。$\varepsilon=0.01$ 说明 2 个用户产生扰动的概率相同($e^{0.01}=1.01$),但这对大多数 k 值来说很难实现。当 ε 为 0.5($e^{0.5}=1.65$)时,超过 90%的扰动点在真实位置 1 000 m 以内,超过 60%的扰动点在真实位置 500 m 以内。扰动点的数量随着 ε 的增加而增加,但是较高的 ε 值会降低方法的实际意义。例如,当 $\varepsilon=2.0$,意味着必须接受 7 倍概率($e^{2.0}=7.39$)差异的结果。然而,具有较小 ε 值的高接近度也是可能的。

表 2　扰动位置邻近用户真实位置的接近度

ε	接近度/%		
	$d\leqslant$1 000 m	$d\leqslant$500 m	$d\leqslant$100 m
0.01	2.39	0.95	0.01
0.1	28.20	16.27	8.13
0.3	89.30	61.24	7.18
0.5	94.26	62.20	11.48
1.0	99.04	73.68	27.27
2.0	99.99	91.87	33.97

对 K 最近邻检索,取 $\varepsilon=0.5$ 产生扰动,图 5、图 6 分别给出了对应不同 K 值的平均相似度和偏移度的实验结果。

由图 5 可以看出,增加 K,便提高了检索结果对象的相似度,随着 K 的增加,K 最近邻检索的相似度由约 60%增加到接近 90%。因为更多数量的检索对象可以看作扩大了搜索半径,在这种情况下,一个对象就变得更有可能在更多数量查询的 K 最近邻对象集中。例如,从一个城市的两个不同地方查找最近的电影院,我们期望这两个位置在它们最近的 10 个电影院查询结果列表中有更高的重叠。重叠的程度依赖这两个位置彼此邻近的程度。在这方面,添加到位置上的噪声至关重要。由 MHC 产生

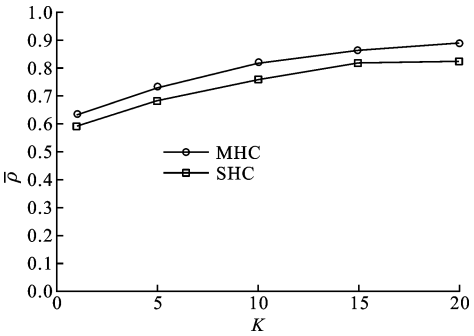


图 5　近似 KNN 检索的平均相似度 $\bar{P}$

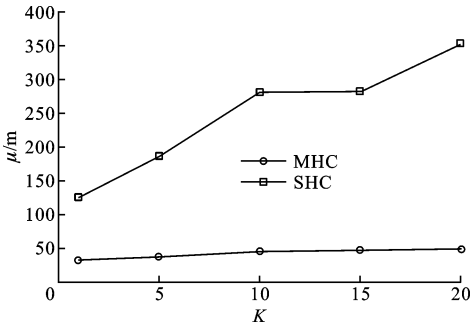


图 6　近似 KNN 检索的偏移度

扰动的 K 最近邻检索结果的相似度比由 SHC 产生扰动的 K 最近邻检索结果的相似度平均提高了 15.4%,这是因为 MHC 考虑了兴趣点的密度分布,在稠密区域仅需添加较小的扰动就可以获得较高的隐私保护度。

偏移度指标表示相对于真实位置的 K 最近邻查询的检索对象与相对于扰动位置的 K 最近邻查询的检索对象的距离差异,其可以更有效地度量检索结果的质量。由图 6 可以看出,由 MHC 产生扰动的 K 最近邻查询的偏移度要比由 SHC 产生扰动的 K 最近邻查询的偏移度小,这是因为 MHC 考虑了上下文信息,产生的扰动要比 SHC 小。

5　结　论

本文提出了一种以用户为中心的差分扰动位置隐私保护方法(Ulp2mDP)。Ulp2mDP 方法不依赖任何 TTP,在产生扰动时考虑了用户所在位置的上下文,能够有效阻止背景信息攻击。通过安全性分析和实验评估,可以发现 Ulp2mDP 方法能够抵御近似位置知识的攻击,使用扰动位置不会显著提高攻击者关于用户位置的先验知识,具有很强的隐私保护强度。但是,一些不合理扰动的识别仍然是一个问题,后续工作中,将考虑放弃 k 匿名集来解决这一问题。

参考文献：

[1] 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 Xuejun，GUI Xiaolin，WU Zhongdong. Privacy preservation for location-based service：a survey [J]. Journal of Software，2015，26(9)：2373-2395.

[2] 张学军，桂小林，冯志超，等. 位置服务中的查询隐私度量框架研究 [J]. 西安交通大学学报，2014，48(2)：8-13.
ZHANG Xuejun，GUI Xiaolin，FENG Zhichao，et al. A quantifying framework of query privacy in location-based service [J]. Journal of Xi'an Jiaotong University，2014，48(2)：8-13.

[3] GRUTESER M，GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking [C]// Proceedings of the First International Conference on Mobile Systems，Applications，and Services. New York，USA：ACM，2003：31-42.

[4] SCHLEGEL R，CHOW C Y，HUANG Q，et al. User-defined privacy grid system for continuous location-based services [J]. IEEE Transactions on Mobile Computing，2015，14(10)：2158-2172.

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

[6] GHINITA G，KALNIS P，SKIADOPPULOS S. MOBIHIDE：a mobile peer-to-peer system for anonymous location-based queries [C]// Proceedings of the 10th International Symposium on Advances in Spatial and Temporal Databases. Berlin，Germany：Springer-Verlag，2007：221-238.

[7] KIDO H，YANAGISAWA Y，SATOH T. An anonymous communication technique using dummies for location-based services [C]// Proceedings of the Second International Conference on Pervasive Services. Piscataway，NJ，USA：IEEE，2005：88-97.

[8] NIU Ben，LI Qinhua，ZHU Xiaoyan，et al. Achieving *k*-anonymity in privacy-aware location-based services [C]// Proceedings of the 33th IEEE Conference on Computer Communications. Piscataway，NJ，USA：IEEE，2014：754-762.

[9] NIU Ben，LI Qinhua，ZHU Xiaoyan，et al. Enhancing privacy through caching in location-based services [C]// Proceedings of the 34th IEEE Conference on Computer Communications. Piscataway，NJ，USA：IEEE，2015：754-762.

[10] PINGLEY A，YU W，ZHANG N，et al. A context-aware scheme for privacy-preserving location-based services [J]. Computer Networks，2012，56(11)：2551-2568.

[11] DEWRI R. Local differential perturbations：location privacy under approximate knowledge attackers [J]. IEEE Transactions on Mobile Computing，2013，12(12)：2360-2372.

[12] XIAO Yonghui，LI Xiong. Protecting location with dynamic differential privacy under temporal correlations [C]// Proceedings of the 22nd ACM Conference on Computer and Communications Security. New York，USA：ACM，2015：1298-1309.

[13] GHINITA G，ZHAO Keliang，PAPADIAS D，et al. A reciprocal framework for spatial *k*-anonymity [J]. Information System，2010，35(3)：299-314.

[14] 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.

[15] GAMBS S，KILLIIJIA MO，CORTEZ M. Show me how you move and I will tell you who you are [J]. Transactions on Data Privacy，2011，2(4)：103-126.

[16] JIANG K，SHAO Dongxu，BRESSAN S，et al. Publishing trajectories with differential privacy guarantees [C]// Proceedings of the 25th International Conference on Scientific and Statistical Database Management. New York，USA：ACM，2013：1-12.

（编辑　武红江）