面向基于位置服务的一种改进型隐私感知虚假位置选择机制

吴 荻, 张 玉, 刘银龙, 马 伟*, 朱大立, 孙 鑫

中国科学院信息工程研究所, 北京 中国 100093

摘要 基于位置的服务已经逐渐成为人们生活中的重要部分, 然而在无线信道中传输位置信息容易受到各种攻击, 导致严重的隐私泄露问题。为此, 本文考虑隐私保护等级以及实际虚假位置区域, 将虚假位置选择问题建模为多目标优化问题, 进而提出一种低复杂度的隐私程度可控的虚假位置选择机制。本文首先从候选虚假位置中选择出请求率差异在指定范围内的虚假位置, 保护一定的隐私等级, 然后从中找出 *K-1* 个虚假位置, 最大化总泛化面积。为了更准确确定总泛化面积, 本文推导出两位置区域的相交面积。安全分析验证了本文提出的算法可以对抗主动攻击以及被动攻击。与其他算法相比, 仿真结果也证明了本文提出的算法可以在保护用户隐私等级的情况下增大总泛化面积。

关键词 基于位置的服务; 隐私保护; 请求率; K 匿名
中图法分类号 TN92, TN918.91 DOI 号 10.19363/j.cnki.cn10-1380/tn.2018.03.07

An Improved Privacy-Aware Dummy Location Selection Scheme in LBSs

WU Di, ZHANG Yu, LIU Yinlong, MA Wei*, ZHU Dali, SUN Xin

Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

Abstract Location-Based Service (LBS) becomes increasingly important for our daily life. However, the localization information in the air is vulnerable to various attacks, which results in serious privacy concerns. To overcome this problem, we formulate a multi-objective optimization problem with considering both the query probability and the practical dummy location region. A low complexity dummy location selection scheme with the controllable privacy level is proposed. For preserving the privacy level, we first find several candidate dummy locations with various query probabilities, whose differences from the query probability of the real user is within a specified threshold. Among these selected candidates, a cloaking area based algorithm is then offered to find the remaining *K-1* dummy locations to maximize the cloaking area. The intersected area between two dummy locations is also derived to assist to determine the total cloaking area. Security analysis verifies the effectiveness of our scheme against the passive and active adversaries. Compared with other methods, simulation results show that the proposed dummy location scheme can improve the privacy level and enlarge the cloaking area simultaneously.

Key words location based services, privacy preservation, query probability, K anonymity.

1 引言

无线定位是一种新型技术, 已广泛应用在工业界各个领域, 比如导航、机器人和监控[1-3]等。基于位置服务(LBS)成为人们生活中的重要部分。在 LBS 的帮助下, 用户可以随时随地得到该位置区域的相关服务, 因此受到学术界与工业界的关注[4]。

随着定位精度变得越来越高, 位置信息的使用变得越来越频繁。因为 LBS 服务器知道用户所有位置信息, 包括现在以及历史位置, 用户位置相关的隐私有可能遭到严重泄露[5]。同时, 由于相关位置信息还包含着时间, 识别号等信息, 因此攻击者可能推理出用户更多的敏感信息[6]。位置隐私的重要性往往被人们低估, 对大多数人来说, 位置隐私看似远不及其他的个人隐私重要, 但实际上并非如此。随着普适计算和移动计算技术的蓬勃发展, 位置隐私的泄露有可能直接导致用户个人隐私的泄露, 如利用用户位置信息分析获取用户的行为习惯, 可能危及

通讯作者: 马伟, 博士, 副研究员, Email: mawei@iie.ac.cn。
收稿日期: 2017-04-27; 修改日期: 2017-06-26; 定稿日期: 2017-08-08

用户的生命财产安全。因此, 保护用户的个人隐私数据越来越受到重视, 已成为近年来的研究热点。

通过将用户的时间、地点以及身份信息相关联, 不法分子可以发掘出更多敏感信息, 对此最直接的解决方法是将用户身份隐藏, 即采用匿名技术将用户的定位信息与身份信息分离[7], 其中最为经典的方法就是 K 匿名技术[8]。该技术使用 *K-1* 个不同的元素与真实用户位置融合一起发给 LBS 服务器, 实现匿名效果。LBS 服务器很难从这 K 个虚假位置集合中找出真实用户位置。尽管使用虚假位置的方法可以实现 K 匿名, 但是如何选择虚假位置具有挑战性。现有大部分研究[9-11]虽然取得了较好的匿名效果，但均假设攻击者不具有边信息(如历史请求率等)[12-13]。而当攻击者攻击拥有全局边信息或者用户全局边信息的LBS服务器时, 现有虚假位置选择算法不能得到很好的 K 匿名效果。这是因为有些虚假位置位于可能性较小的位置, 比如湖边、建筑旁等, 从这些区域存在用户请求的概率很低, 所以很容易被攻击者排除。文献[14, 15]考虑了边信息的影响, 提出基于信息熵的虚假位置选择算法保护用户隐私。但一方面, 该方法是在不同候选虚假位置集合中选择具有最大信息熵的集合, 并不针对给定候选虚假位置集合内的选择。另一方面, 用户上报的虚假位置通常是一定范围的区域, 而不是数学上抽象的点。但现有研究工作很少考虑实际虚假位置区域对隐私保护水平的影响。

针对上述问题, 本文考虑虚假位置请求率以及泛化面积, 将虚假位置选择问题建模为多目标优化问题, 进而提出一种低复杂度的隐私程度可控的虚假位置选择机制, 首先从候选虚假位置中至少选择出 $M(M \geqslant K)$个请求率差异在指定范围内的虚假位置, 保证隐私等级①, 然后从中找出 *K-1* 个虚假位置最大化总泛化面积, 保护真实用户隐私。

本文的主要结构如下: 第 2 章综述并对比相关研究工作; 第 3 章描述相关背景知识, 系统模型以及攻击模型; 第 4 章, 将虚假位置选择问题建模为多目标优化问题, 并设计两步骤虚假位置选择算法求解。为了显示提出算法的优越性; 在第 5 章分析了该算法针对不同类型攻击时的安全性能; 并在第 6 章利用仿真验证了算法的性能。最后一章总结了全文, 并确定了未来研究方向。

2 相关研究工作

为了减少 LBS 中位置隐私的泄露程度, 近年来研究者们致力于研究如何保护用户的位置隐私, 并针对经典的 K 匿名方法提出了多种解决方案, 主要包括时间泛化和空间泛化两种。时间泛化是通过牺牲时间精度来换取空间精度[16], 其基本思想是当中介接收到用户的位置报告后, 并不立刻转发给服务提供商, 而是先缓存起来, 直到缓存了 K 个以上用户的位置报告, 再一起发给服务提供商以换取服务。文献[17]设计了一种可定制的 K 匿名模型保护位置信息, 可以提供可变的 K 匿名, 允许更多用户受益于位置隐私保护。研究者针对该模型提出一种时空泛化算法为移动用户提供 K 匿名隐私保护, 但该算法需要在可信第三方上运行, 完成后上报给 LBS 服务器。时间泛化方式虽然可以获得 K 匿名效果, 但实时性比较差, 且必须要有可信第三方的存在。

已有很多研究工作者着手于如何选择空间泛化区域内的虚假位置区域。其基本思想是选取一块足够大的区域, 使其覆盖至少 K 个用户区域。当该区域内的用户需要 LBS 时, 向提供服务的服务器报告用户真实位置以及虚假位置区域, 而不是用户真实位置区域, 从而实现 K 匿名效果。文献[9]提出一种在线个性化的机制产生匿名区域以保护移动用户的移动设备隐私, 其设计思想是合并几何转换和动态假名更改机制, 用户可控地产生虚假位置, 保护用户隐私。该方法不需要引入第三方可信机构, 可直接融入现有设备, 并可以有效对抗强攻击模型的推理攻击。文献[10]为了减小用户请求开销, 提出一种轻量级方案根据虚拟网格或者是圆形区域, 产生虚假位置保证用户隐私。在文献[11]中, 学者使用虚假位置在一定范围内隐藏用户位置。该方法可以弥补用户由于意外泄露的位置信息。文献[18]基于短期泄露、长期泄露以及距离偏移等因素提出两种虚假位置方法, 有效保护用户位置隐私。为了减小由于用户自私性引起的隐私保护副作用, 文献[19]从博弈论的角度研究虚假位置选择问题, 将其建模为静态和时间感知的贝叶斯博弈模型, 并分析了该问题解的存在性。基于此, 学者提出一种虚假位置选择策略帮助用户得到成本与隐私之间的平衡。文献[20]则提出一种新颖的合作系统, 合并 K 匿名技术与缓存技术以保护用户位置隐私。通过合理选择虚假位置, 保护用户隐私的同时改善缓存命中率。文献[21]使用频繁改变的假名和虚假位置保护位置隐私。每当用户改变假名时, 虚假位置选择机制将随即触发, 使得攻击者很难追踪真实用户。研究者联合最近邻居搜索法与空

① 隐私等级用于衡量隐私保护的强度。隐私等级越高，越难以随机方式猜测出真实用户位置；隐私等级越低，越易以随机方式猜测出真实用户位置。

间泛化区域来满足用户匿名请求[22]。文献[23]为了解决较大泛化空间的低效率以及高成本问题，设计了一种方法联合 K 匿名与虚假位置选择机制，可获得效率与隐私保护之间的折中。

上述研究从各角度考虑虚假位置选择问题，也获得了较好的匿名效果，但均假设攻击者不具有边信息[12-13]。当攻击者攻击拥有全局边信息，或者用户全局边信息的 LBS 服务器的时候，现有虚假位置选择算法不能得到很好的 K 匿名效果。文献[18]通过利用边信息对位置数据进行长期的收集和分析，位置服务商能以很大概率恢复出用户的真实轨迹。针对此问题，文献[14-15]考虑边信息的影响，提出基于信息熵的虚假位置选择算法，在保护用户隐私的同时最大化虚假位置之间的距离。该算法可以有效防止攻击者利用边信息进行攻击，保护用户隐私等级并最大化泛化范围。但该方法是在不同候选虚假位置集合中选择具有最大信息熵的集合，并不针对给定候选虚假位置集合内的选择。文献[24]提出使用基于滑动窗口的 K 匿名算法，选择 K-1 个虚假位置。同时为了保护用户隐私，引入最大熵和选择机制以防止攻击。

另一方面 用户上报的虚假位置通常是一定范围的区域，而上述研究均认为虚假位置为数学上抽象的点。当攻击者得知泛化区域内所有虚假位置的具体区域时，将影响 K 匿名效果。实际的虚假位置区域可能存在各种影响因素，比如部分被物体遮挡，或者是虚假位置区域相互重叠的情况，均将导致有效隐私保护区域减小。

3 背景介绍

3.1 研究动机

现有 LBS 应用中，当用户有需求时向 LBS 服务器提出请求(包括用户信息以及所在区域)，然后基于位置服务的服务器发送相应服务给该用户。为了保护用户隐私，用户通常会随机选择 K-1 个虚假位置，并将这些虚假位置与自己真实的位置区域一起打包发给 LBS 服务器。因此理论上，用户真实位置被暴露的概率应为 $1/K$，即所谓的 K 匿名技术。然而，因为攻击者也许有所有位置的历史请求信息，就可能根据请求信息的概率推测并排除比如 K_d 个虚假位置，这样用户真实位置被暴露的概率将变成 $1/(K\text{-}K_d)$，不如预期匿名效果。同时，现有研究考虑的虚假位置是数学中模型的抽象点，如图 1 所示，而实际中考虑可能是任意形状的虚假位置区域，如图 2 虚线所示圆形区域。这样就可能存在所选虚假位置区域互相相交，或者所选虚假位置区域被物体遮挡的情况。一旦发生上述现象，使用 K 匿名技术实现的隐私保护面积将小于预期面积，相应的隐私保护程度也将降低。

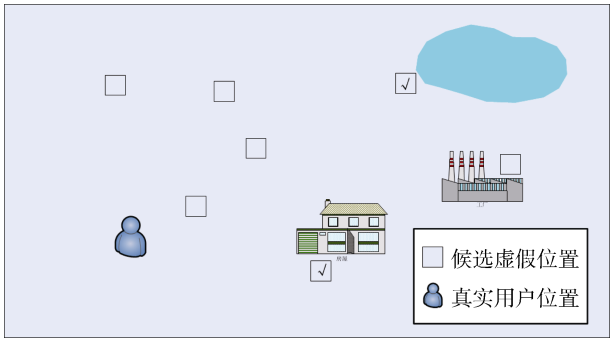


图 1 虚假位置选择：理论位置点模型

Figure 1 Dummy location selection: ideal location point model

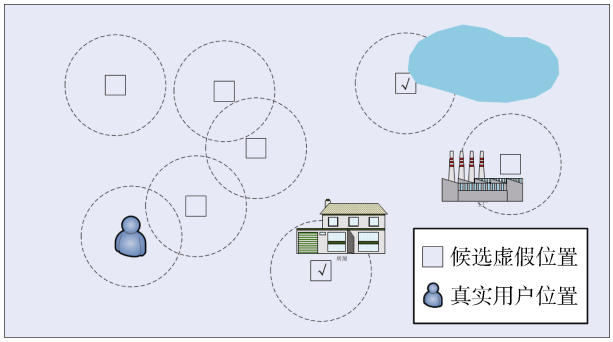


图 2 虚假位置选择：实际位置区域模型

Figure 2 Dummy location selection: practical location region model

3.2 系统模型

本文考虑无线网络，包含有 N 个候选虚假位置区域随机分布在无线网络中，本文研究圆形区域，而不是数学上抽象的点模型。每个圆形区域半径为 R。每个虚假位置可能被物体遮挡，令第 n 个虚假位置 δ_n 的区域被遮挡，$\delta_n \in [0,1]$，$n \in \mathcal{N}$，$\mathcal{N}$ 为无线网络中 N 个候选虚假位置区域的集合。在实际中，每个虚假位置区域被遮挡的区域可以预先测量得到。每个虚假位置的历史请求率可以通过服务器记录历史请求次数确定，即

$$q_n = \frac{\#_n}{\sum_{n\in\mathcal{N}} \#_n} \tag{1}$$

其中 $\#_n$ 为虚假位置 n 的历史请求次数。

为了保护真实用户的隐私安全，使用 K 匿名技术。当用户向 LBS 服务器提出请求，用户选择 K-1 个虚假位置，并将其与自己真实位置区域一起发送

给 LBS 服务器并等待回复位置服务。为了研究方便, 本文仅考虑最多两个虚假位置区域相交的情况。更多区域相交的情况可以用类似的方法求解。令虚假位置 i 和 j 的相交面积为 A_{ij}^I, $i,j\in\mathcal{N}$, 虚假位置区域中被物体遮挡的面积为 $A_n^O=\pi R^2\delta_n$, $n\in\mathcal{N}$, $0\leqslant\delta_n\leqslant 1$。当 δ_n=1表示全部被物体遮挡, 当 δ_n=0表示未被遮挡。

当虚假位置区域远离其他虚假位置, 或者未被任何物体遮挡时, 所选虚假位置区域总泛化面积应为 $K\pi R^2$, 则实现了 K 匿名应有的总泛化面积。然而, 当虚假位置区域与其他虚假位置区域相交, 或者被物体部分遮挡, 总泛化面积将会受到影响。有以下几种情况将导致总泛化面积的减少。

i. 当两个候选虚假位置相距较远时, 它们不相交, 如图 3(a)所示。在这种情况下, 两个虚假位置区域组成的总泛化面积为 $2\pi R^2-A_i^O-A_j^O$;

ii. 当两个候选虚假位置相互靠近的时候, 它们将相交, 如图 3(b)所示。在这种情况下, 两个虚假位置区域组成的总泛化面积为 $2\pi R^2-A_{ij}^I-A_i^O-A_j^O$;

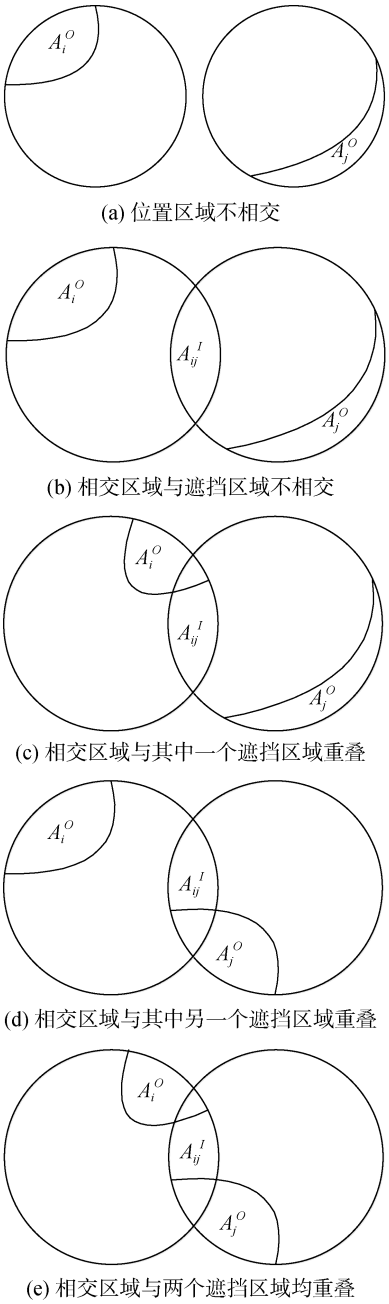
iii. 当两个候选虚假位置相交且与虚假位置区域 i 的遮挡面积重叠, 如图 3(c)所示。由于遮挡面积有可能是不规则的, 因此为了数学上可计算, 假设相交和遮挡均等地对总泛化面积产生影响。在这种情况下, 两个虚假位置区域组成的总泛化面积为 $2\pi R^2-A_j^O-(A_i^O+A_{ij}^I)/2$。同理, 当两个候选虚假位置相交且与虚假位置区域 j 的遮挡面积重叠, 如图 3(d)所示, 两个虚假位置区域组成的总泛化面积为 $2\pi R^2-A_i^O-(A_j^O+A_{ij}^I)/2$;

iv. 两个候选虚假位置相交且与虚假位置区域 i 和虚假位置区域 j 的遮挡面积同时重叠, 如图 3(e)所示。假设相交和遮挡均等地对总泛化面积产生影响。在这种情况下, 两个虚假位置区域组成的总泛化面积为 $2\pi R^2-(A_j^O+A_i^O+A_{ij}^I)/3$。

此外, 对于更多虚假位置区域存在于指定范围内时, 将会出现多于两个虚假位置区域相交的情况出现。虽然这种概率相对较小, 但一旦出现也将对性能有所影响。由于当相交面积较多时, 总泛化面积将进一步缩小。具体情况可以通过类似上述方法: 首先求解出多个虚假位置区域相交的面积, 然后与多个虚假位置的遮挡区域面积共同求解总泛化面积。

3.3　攻击模型

攻击者的目的是得到真实用户的敏感信息比如位置。因此他需要知道所有候选虚假位置区域的信息。用户通常可以得到两种类型的边信息: 局部边信息和全局边信息。局部边信息指的是部分用户的历史请求信息; 全局边信息指的是所有用户的历史请求信息。攻击方式可分为被动式和主动式两种。被动式攻击者可以监测监听无线信道, 或攻破用户得到相关的历史请求信息。主动式攻击者可以攻破 LBS 服务器, 从而得到所有候选位置历史请求信息。被动式攻击意味着攻击者能够得到局部边信息; 主动式攻击意味着攻击者能够得到全局边信息, 并时刻监测所有位置的请求情况。同时, 攻击者还知道系统使用 K 匿名技术



(a) 位置区域不相交
(b) 相交区域与遮挡区域不相交
(c) 相交区域与其中一个遮挡区域重叠
(d) 相交区域与其中另一个遮挡区域重叠
(e) 相交区域与两个遮挡区域均重叠

图 3　导致总泛化面积减少的 5 种情况示意图

Figure 3　Sketches of five cases leading to area reduction

保护真实用户隐私, 以及使用的虚假位置选择算法。

4 虚假位置选择机制

本文的主要目标是在考虑虚假位置历史请求以及圆形位置区域和被物体部分遮挡状态下最大化隐私等级和总泛化面积。实际虚假位置区域有可能与其他虚假位置区域相交, 也有可能被物体部分遮挡。因此该虚假位置选择问题可以建模为多目标优化问题。一方面, 我们期望最大化隐私等级, 这样可以得到 K 匿名效果, 否则攻击者将很容易排除不可能的候选虚假位置。实现 K 匿名效果等效于选出的虚假位置请求率与真实用户位置的请求率差异尽可能小, 即最小化被选中位置集合中请求率之间的差异值。这样即使攻击者得知所有被选中虚假位置的请求率, 仍然无法从中找到真实用户的位置, 从而得到 K 匿名效果。另一方面, 我们期望尽可能增大实际虚假位置区域的总泛化面积, 这样可以增大匿名区域。

对于第一个目标, 提高隐私等级的优化问题可以建模为:

$$\text{P1:}\ \min_{n\in\mathcal{N}}\left(q^{\max}-q^{\min}\right) \tag{2}$$

$$\text{s.t.}\ \ q^{\max}=\max_{n\in\mathcal{K}} q_n \tag{3}$$

$$q^{\min}=\min_{n\in\mathcal{K}} q_n \tag{4}$$

$$\mathcal{K}\in N$$

其中 $\mathcal{K}$ 为 *K-1* 个所选虚假位置的集合。目标函数是从 N 个候选虚假位置之中选出 *K-1* 个虚假位置, 其虚假位置与用户真实位置组成的集合 $\mathcal{K}$ 具有最小的历史请求率差异。限制条件(3)和(4)分别表示集合 $\mathcal{K}$ 中最大历史请求率以及最小历史请求率。

对于第二个目标, 增大总泛化面积的优化问题可以建模为:

$$\text{P2:}\ \max_{i\in\mathcal{N}}\left(K\pi R^2-\sum_{i\in\mathcal{K}}\sum_{j=i+1}^{K}\phi_{ij}\right) \tag{5}$$

$$\text{s.t.}\ \ \mathcal{K}\in\mathcal{N} \tag{6}$$

其中 ϕ_{ij} 表示虚假位置 i 和 j 需要去除的面积。根据上述导致总泛化面积减少的不同情况, ϕ_{ij} 的值可以表示为:

$$\phi_{ij}=\begin{cases}A_i^O+A_j^O, & D_{ij}^I=\varnothing,\\ \dfrac{A_i^O+A_{ij}^I}{2}+A_j^O, & D_{ij}^I\cap D_i^O\neq\varnothing, D_{ij}^I\cap D_j^O=\varnothing,\\ \dfrac{A_j^O+A_{ij}^I}{2}+A_i^O, & D_{ij}^I\cap D_j^O\neq\varnothing, D_{ij}^I\cap D_i^O=\varnothing,\\ \dfrac{A_j^O+A_i^O+A_{ij}^I}{3}, & D_{ij}^I\cap D_j^O\cap D_i^O\neq\varnothing,\end{cases} \tag{7}$$

其中 D_{ij}^I 定义为候选位置区域 i 和 j 之间的相交区域, D_i^O 为虚假位置 i 被遮挡的区域。由此可见, ϕ_{ij} 的大小取决于 A_i^O、A_j^O、A_{ij}^I。A_i^O 和 A_j^O 可以直接测量得到, 而 A_{ij}^I 可以通过计算得到:

$$A_{ij}^I=\frac{\arcsin\left(\dfrac{\sqrt{R^2-d_{ij}^2/4}}{R}\right)}{90}\pi R^2-d_{ij}\sqrt{R^2-d_{ij}^2/4} \tag{8}$$

其中为 d_{ij} 为虚假位置 i 和 j 之间的距离。该公式的详细推导请参见附录。

通过以上描述, 可以得到多目标优化问题为 $\max\{P1',P2\}$, 其中

$$P1':\ \max_{n\in\mathcal{N}}\left(\frac{1}{q^{\max}-q^{\min}}\right) \tag{9}$$

$$\text{s.t.}\ \ q^{\max}=\max_{n\in\mathcal{K}} q_n$$

$$q^{\min}=\min_{n\in\mathcal{K}} q_n$$

$$\mathcal{K}\in\mathcal{N}$$

由于目标函数以及限制条件中的变化相互影响, 同时优化问题的非凸性, 通常很难直接求解[25]。当使用穷举法时, 相当于从 N 个候选节点中选择 *K-1* 个合适的虚假位置区域, 最大化隐私等级和总泛化面积, 其搜索复杂度为 $\mathcal{O}(C_N^{K-1})$, 其中 $C_N^K=\dfrac{N!}{K!(N-K)!}$。

为了快速求解上述优化问题, 本文提出一种两步骤虚假位置选择机制。首先从 N 个候选虚假位置中选择出至少 M 个请求率差异小于一定阈值的候选虚假位置, 称为“基于请求率的虚假位置选择”算法。然后从选定的候选虚假位置集合中, 根据虚假位置实际情况选择 *K-1* 个虚假位置, 使得其总泛化面积最大, 称为“基于泛化面积的虚假位置选择”算法。

算法 1. 基于请求率的虚假位置选择

1: 输入 $\mathcal{N}$、M、Q、q_0、ε 和 ε_Δ
2: 初始化 $\mathcal{M}=\varnothing$
3: WHILE $|\mathcal{M}|<M$ DO
4: $\varepsilon=\varepsilon+\varepsilon_\Delta$
5: IF $\varepsilon\geqslant 1$
6: $\varepsilon=1$
7: ENDIF
8: 初始化 $q'=q_0$
9: WHILE $(|q'-q_0|<\varepsilon)\cap(\varepsilon\leqslant\varepsilon_{up})$ DO
10: IF $\mathcal{N}=\varnothing$
11: Break
12: ENDIF

13:　$m' = \arg\min_{m\in\mathcal{N}} |q_0 - q_m|$

14:　$\mathcal{M} = \mathcal{M}+\{m'\}$

15:　$\mathcal{N} = \mathcal{N} - \{m'\}$

16:　$q' = q_{m'}$

17:　END WHILE

18: IF　ε=1

19:　Break

20: ENDIF

21: END WHILE

22: 输出 $\mathcal{M}$

算法 1 描述了“基于请求率的虚假位置选择”算法, 其中 q_0 是真实用户请求率, Q 是 N 个候选虚假位置的请求率集合, M 是算法 1 需要的候选虚假位置的最小个数, $M \leqslant N$, ε 为虚假位置与真实用户位置之间请求率差异容忍值, ε_Δ 是 ε 的步长, $\varepsilon_\Delta > 0$。由于隐私等级受请求率直接影响, 因此当请求率差异较小时, 更能有效实现 K 匿名效果, 也就能更好保护隐私等级; 相反, 当请求率差异较大时, 实现 K 匿名效果的可能性也较小, 也就不能更好保护隐私等级。M 的具体值可根据实际需要保证的隐私等级确定, 与虚假位置与真实用户位置之间请求率差异容忍值直接相关, 这是系统可控的。当虚假位置与真实用户位置之间请求率差异容忍值越大, M 值越大, 反之亦然。

该算法的主要思想是从 N 个候选虚假位置中选择出至少 M 个虚假位置集合用于后续“基于泛化面积的虚假位置选择”中, 这样可以保证一定的隐私等级, 同时可以选出足够的候选虚假位置。在第 9 行与第 17 行之间的 WHILE 循环是为了从所有 N 个候选虚假位置找到与真实位置请求率差异小于请求率差异容忍值 ε 的所有虚假位置。在每次循环中, 只要此次虚假位置请求率与真实位置请求率之间的差异满足限制, 即将该虚假位置纳入 $\mathcal{M}$ 集合中, 并将该虚假位置从 $\mathcal{N}$ 集合中去除。在第 3 行与第 21 行之间的 WHILE 循环是为了保证至少有 M 个候选虚假位置供后续使用。如果在内循环完成后得到的虚假位置个数不足 M 个, 则将请求率差异容忍值增加 ε_Δ, 然后再次执行第 9 行与第 17 行之间的 WHILE 循环, 直至找到至少 M 个候选虚假位置供后续选择或者所有候选位置被选中。

基于请求率的虚假位置选择的具体流程下所示。首先输入主要参数 $\mathcal{N}$、M、Q、q_0、ε、ε_Δ 和 ε_{up}。然后依次从所有候选虚假位置中选出请求率小于请求率差异容忍值 ε 的所有虚假位置。当个数大于预置阈值 M 时, 则停止“基于请求率的虚假位置选择”算法, 否则更新请求率差异容忍值, 直至选出的虚假位置个数大于 M 或请求率差异容忍值 ε 达到给定上限 ε_{up}。当选择出请求率接近的候选虚假位置集合, 就保证了一定的隐私等级。之后则将重点致力于最大化泛化面积。

算法 2. 基于泛化面积的虚假位置选择

1: 输入 $\mathcal{M}$

2: 初始化 $\mathcal{L} = \{L_0\}$

3: FOR *k=2 to K* DO

4:　$j^* = \arg\max_{j\in\mathcal{M}} \left(\theta(\mathcal{L}+\{j\})\right)$

5:　$\mathcal{L} = \mathcal{L}+\{j^*\}$

6:　$\mathcal{M} = \mathcal{M} - \{j^*\}$

7: END FOR

8: 输出 $\mathcal{L}$

算法 2 为执行“基于泛化面积的虚假位置选择”算法的主要思路, 其中 $\mathcal{L}$ 表示已选出的虚假位置集合, $\theta(\mathcal{L})$ 代表集合 $\mathcal{L}$ 中虚假位置组成的总泛化面积, L_0 表示真实位置。在第 3 行与第 7 行之间的 FOR 循环是为了从集合 $\mathcal{M}$ 中找到带有最大总泛化面积的 *K-1* 个虚假位置。首先将真实位置区域选入集合 $\mathcal{L}$ 中, 然后将具有最大泛化总面积的虚假位置依次添加入集合 $\mathcal{L}$ 中, 并依次将选出的虚假位置从集合 $\mathcal{M}$ 中去除。重复上述过程直到找到 *K-1* 个虚假位置。

基于泛化面积的虚假位置选择的具体流程如下所示。首先将真实用户位置纳入已选集合中。然后从候选集合 $\mathcal{M}$ 中遍历所有的虚假位置, 并于已选集合 $\mathcal{L}$ 中的已选的集合组成临时集合 $\mathcal{L}+\{j\}$, $(j \in \mathcal{M})$, 计算集合 $\mathcal{L}+\{j\}$ 的总泛化面积。从中选择出具有最大总泛化面积的虚假位置 j*, 将该位置选入已选集合 $\mathcal{L}$ 中, 并从候选集合 $\mathcal{M}$ 中移除。直至已选集合中包含 K 个虚假位置。

算法 2 的关键是计算 $\theta(\mathcal{L})$, 这依赖于本文第 3.2 章归纳的几种情况, 算法 3 详细描述了 $\theta(\mathcal{L})$ 的计算过程。首先从真实位置的泛化面积算起, 记初始面积为 $S = \pi R^2 - A_1^O$。然后依次将集合 $\mathcal{L}$ 中的虚假位置参与计算, 如第 3 行与第 23 行之间的循环。由于候选虚假位置与已选虚假位置之间可能存在相交或遮挡的情况, 因此根据上述不同情况需要分别计算。当上述循环完成后, 即可得到集合 $\mathcal{L}$ 中总泛化面积。

算法 3. 计算 $\theta(\mathcal{L})$

1:　输入 $\mathcal{L}$

2:　初始化 $S = \pi R^2 - A_1^O$

3:　for i=2 to $|\mathcal{L}|$ do

4:　　for j=1 to i-1 do

5: if $D_{ij}^I = \varnothing$ then

6: $S = S + \pi R^2 - A_i^O$

7: else

8: if $D_{ij}^I \cap D_i^O = \varnothing, D_{ij}^I \cap D_j^O = \varnothing$ then

9: $S = S + \pi R^2 - A_i^O - A_{ij}^I$

10: else

11: if $D_{ij}^I \cap D_i^O \neq \varnothing, D_{ij}^I \cap D_j^O = \varnothing$ then

12: $S = S + \pi R^2 - \frac{A_i^O + A_{ij}^I}{2}$

13: end if

14: if $D_{ij}^I \cap D_j^O \neq \varnothing, D_{ij}^I \cap D_i^O = \varnothing$ then

15: $S = S + \pi R^2 - A_i^O + A_j^O - \frac{A_j^O + A_{ij}^I}{2}$

16: end if

17: if $D_{ij}^I \cap D_i^O \cap D_j^O \neq \varnothing$ then

18: $S = S + \pi R^2 + A_i^O + A_j^O - \frac{A_j^O + A_i^O + A_{ij}^I}{3}$

19: end if

20: end if

21: end if

22: end for

23: end for

24: $\theta(\mathcal{L}) = S$

25: 输出 $\theta(\mathcal{L})$

计算 $\theta(\mathcal{L})$ 的具体流程如下所示。首先计算真实用户位置的泛化面积。然后从集合 $\mathcal{L}$ 中任选一个虚假位置开始计算泛化面积。判断与现有计算位置区域是否相交, 如果不相交, 即可根据情况(a)计算总泛化面积; 如果相交, 需要判断该虚假位置区域与相交位置区域是否存在遮挡, 如果不存在遮挡, 则即可根据情况(b)计算总泛化面积; 如果存在遮挡, 则存在 3 种可能: 相交区域与已计算区域的遮挡区域重叠, 可根据情况(c)计算总泛化面积; 相交区域与待计算区域的遮挡区域重叠, 可根据情况(d)计算总泛化面积; 相交区域与已计算区域的遮挡区域、待计算区域的遮挡区域均重叠, 可根据情况(e)计算总泛化面积。当根据具体情况计算出总泛化面积后, 将该位置从集合 $\mathcal{L}$ 中移除。重复上述过程直至 $\mathcal{L}$ 集合中不包含任何虚假位置。

5 安全分析

因为密码学的方法(比如公钥基础设施)可以较容易应用到本文提出的系统中, 因此本文忽略通过无线信道窃听的攻击方式, 主要关注被动攻击和主动攻击引起的合谋攻击和推理攻击。

首先分析合谋攻击, 即攻击者具有一个或多个虚假位置的相关位置请求信息, 并从中猜测真实位置的攻击方式。被动攻击很可能与其他用户或者 LBS 服务器合谋套取其他用户隐私信息。如果被攻击的用户不位于被选中的虚假位置, 那么猜测出真实用户位置区域的概率为*(K-1)/N*。当所有候选虚假位置个数 *N* 较大的时候, 则被猜中的概率很低。如果被攻击的用户正好位于所选被选中的虚假位置之一, 那么攻击者猜对真实用户位置的概率不会大于 *1/K*, 因为真实位置位于 *K* 个虚假位置之间。接下来研究主动攻击方式, 即如果攻击者获得了这 *K* 个位置的历史请求率信息。但即使在该情况下, 攻击者仍然无法有效猜出真实用户位置, 因为被选中的每个虚假位置具有的请求率与真实用户位置的请求率差异很小, 攻击者唯一能做的是在所有覆盖的区域搜寻。然而一方面, 攻击者需要攻破所有 *K* 个虚伪位置并获得相关信息, 但这难度比较大; 另一方面, 算法 2 保证了本文提出的虚假位置选择算法能得到最大的总泛化面积, 即使偶然被猜中, 攻击者也不能具体定位真实位置。因此, 即使攻击者知晓本文提出的虚假位置选择算法及内容, 仍然只能从 *K* 个虚假位置中随机猜测真实用户的位置区域, 且概率不会高于 *1/K*。这一结论也会在下文的仿真结果中得到验证。综上所述, 本文提出的算法可以对抗合谋攻击。

其次分析推理攻击, 即攻击者利用已有的虚假位置信息, 通过推理的方式猜测真实位置的一种攻击方式。被动攻击者仅具有局部信息且是被动的, 不具有强威胁, 因此本文直接分析主动攻击方式。主动攻击者可以通过监测区域内所有的用户并可得到所有相关信息, 包括所有位置的历史请求率信息、用户身份、混合真实用户与虚假位置区域的请求信息等。通过这些信息, 攻击者可以使用推理攻击得到有关真实用户的敏感信息。对于主动攻击者, 假设他知道系统采用了本文提出的虚假位置选择算法, 并知晓该算法的具体流程以及主要参数, 因此攻击者将会尝试反向破解该算法。但本文提出的虚假位置选择算法也能够很好地对抗此类攻击, 原因如下。首先, 本文提出的“基于请求率的虚假位置选择”算法选择出至少 *M* 个具有与真实用户位置请求率差异在一定范围内的候选虚假位置集合。按照算法 1 的原则, 将会出现有些选出的虚假位置请求率略高于真实用户的请求率, 其余的则略低于真实用户的请求率。其

中高于或低于真实用户位置请求率的虚假位置的个数和具体位置均是随机的。因此，由于上述随机性，即使攻击者能够获得所有候选虚假位置集合以及请求率，也无法以高于 *1/M* 的概率从中猜测出真实用户的位置，从而保证了用户的隐私等级。其次，本文提出的“基于泛化面积的虚假位置选择”算法在上述算法选出的至少 *M* 个候选虚假位置中，继续选择出 *K-1* 具有最大总泛化面积的虚假位置，同时保证了用户的隐私面积。即使攻击者采用主动攻击，也不能有效猜出实际用户的真实位置区域。因此，本文提出的算法还可以对抗推导攻击。

6　数值仿真

本章使用数值仿真来证明所提出的虚假位置选择算法的性能。一个真实用户以及 *N* 个候选虚假位置区域随机均匀地分布在 *l***w* 的矩形区域。由于候选虚假位置可能处于特殊位置，将存在零请求的情况。令每个候选虚假位置出现零请求的概率为 ρ_0。为了验证本文提出的两步骤虚假位置选择算法，本章节使用蒙特卡洛方法分别对“基于请求率的虚假位置选择”算法和“基于泛化面积的虚假位置选择”算法进行验证。

对于“基于请求率的虚假位置选择”算法，本文用检测率来衡量。检测率定义为攻击者猜对真实用户位置的概率，检测率越高说明匿名效果越差，反之亦然。在仿真中，检测率可以由攻击者猜对真实位置的总次数与总仿真次数的比值确定。由于通过“基于请求率的虚假位置选择”算法得到的虚假位置集合能有效保证隐私等级，之后执行“基于泛化面积的虚假位置选择”算法从上述候选集合中选择出的最终虚假位置区域亦能保证预期的隐私等级，因此对于“基于泛化面积的虚假位置选择”算法的性能衡量，本文使用归一化总泛化面积，即将总泛化面积除以 *K* 个没有被遮挡也没与其他位置相交的虚假位置区域的总泛化面积。归一化总泛化面积的取值范围为 0~1。数值越大代表总泛化面积越大，隐私保护效果越好。

本章节主要考虑两种仿真场景。第一种是用户静止不动的场景，假设用户静止在区域内某个位置，该位置在研究区域内服从随机分布。当有服务请求时，向基于位置服务器发送自身相关位置信息以及虚假位置信息。图 4~7、图 9~12 给出了用户静止情况下的仿真结果。第二种场景为用户移动的场景，本文使用现实中广泛应用的航路点(waypoint)模型[26]来描述该场景。假设用户初始位置在研究区域内服从随机分布，然后用户可以朝 0~2π 中某个方向以一定速度移动，该移动速度在 0 和 $v_{\max}$ 之间随机分布，其中 $v_{\max}$ 为用户的最大移动速度。如果用户移动到研究区域边缘则可以选择朝另一个方向以另一个移动速度移动。该场景主要是为了研究当用户移动时对虚假位置选择算法的影响。图 8 和图 13 给了用户移动情况下的仿真结果。

6.1　平均检测率

首先验证“基于请求率的虚假位置选择”算法的性能。默认的仿真参数如表 1。

表 1　主要仿真参数

Table 1　Main simulation parameters

仿真参数	值
ε	0.02
ε_{up}	0.03
ε_{Δ}	0.01
ρ_0	0.3
M	20
N	40

该算法的目标是保证 *K* 匿名的隐私等级。按照算法 1 的描述，使用“基于请求率的虚假位置选择”算法选择出至少 *M* 个候选虚假位置集合以供后续进一步选择。为了显示提出虚假位置选择算法的优势，本文将对比几种相关的虚假位置选择算法对保护隐私等级的效果。本文提出的虚假位置选择算法记作“proposed”算法。第二种方法是从所有候选虚假位置集合中随机选择出 *M-1* 个虚假位置，记作“random”算法。最后将 *1/M* 作为衡量性能的基准，考察虚假位置选择算法是否能保证一定的匿名效果，记作“baseline”。

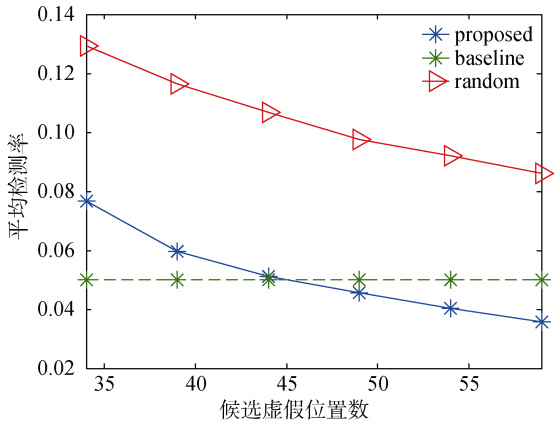


图 4　不同候选虚假位置数初始值时平均检测率

Figure 4　Average detection rate versus the number of candidate dummy locations

图 4 为不同候选虚假位置个数时平均检测率结果。随着候选虚假位置个数的逐渐增大，“proposed”算法和“random”算法的检测率逐渐减小，对真实用户的隐私保护效果也越来越强。这是由于更多的选择也带来了更多的性能增益，可能性小的虚假位置将不会被选出。“baseline”曲线仅与 M 值有关，因此在该仿真中保持不变。当候选虚假位置个数超过 45 的时候，提出的虚假位置选择算法的检测率低于“baseline”基准，说明超过 M 个虚假位置满足给定差异阈值，因此能得到更好的性能。同时发现“proposed”算法优于“random”算法，也证明了提出的虚假位置选择算法在保护隐私等级方面的优势。

接下来在图 5 中评估不同 M 个数对平均检测率的影响。当 M 值增加时，所有曲线得到的检测率都存在不同程度的下降。“proposed”算法整体优于“random”算法，但当 M 值为候选虚假位置总个数的时候，“proposed”算法得到的检测率与“random”算法重合，这是因为当所有候选虚假位置都被选中时，没有选择增益的存在。同样，当候选虚假位置个数超过一定数目的时候，提出的虚假位置选择算法的检测率低于“baseline”基准，说明“proposed”算法能得到更好的性能。由于“proposed”算法还受到阈值影响，因此曲线整体降低比较平缓。

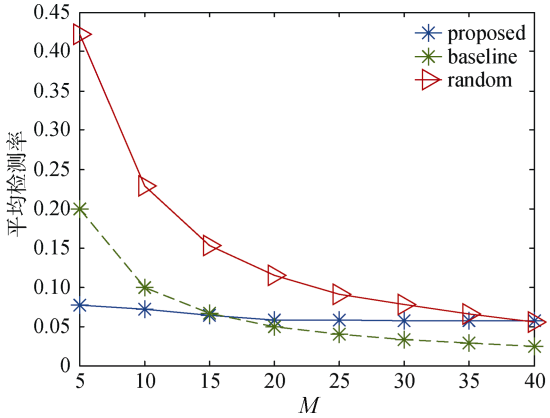


图 5　不同 M 值时平均检测率

Figure 5　Average detection rate versus M

图 6 的横坐标为初始阈值，纵坐标为平均检测率。由于“proposed”算法考虑选择相似的请求率并排除了可能性小的虚假位置，因此保护隐私的效果比“random”算法好。当给定阈值比较小的时候，“proposed”算法不如“baseline”基准，说明满足阈值的虚假位置不足 M 个，而当放大阈值到 0.04 以上，“proposed”算法能选择出不止 M 个供后续算法使用，因此得到了更低的检测率性能。

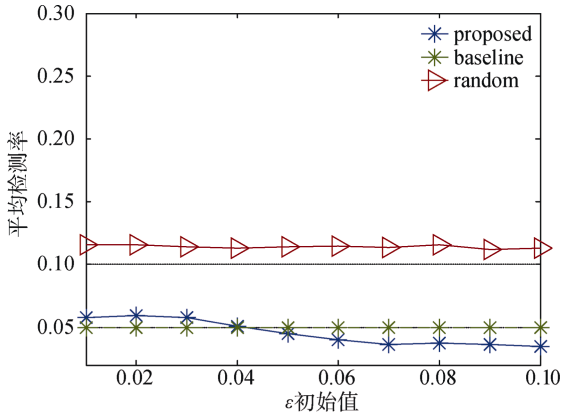


图 6　不同 ε 初始值时平均检测率

Figure 6　Average detection rate versus the initial value of ε

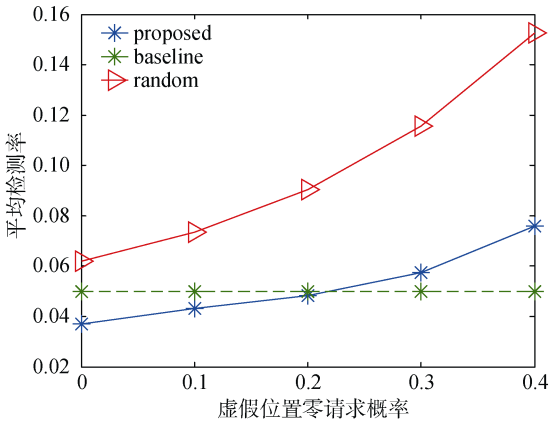


图 7　不同虚假位置零请求概率初始值时平均检测率

Figure 7　Average detection rate versus the initial value of zero probability query

最后评估虚假位置无请求概率对平均检测率的影响。如图 7 所示，当有更多的虚假位置存在零请求率的时候，“proposed”算法和“random”算法的检测率逐渐增大，这是因为可能性小的虚假位置不断增加，更难选出合适的虚假位置。同时由于提出的虚假位置选择算法能更好地考虑每个候选位置的请求率，观察到它们之间的间隔也变大。当无请求概率大于一定值以后，提出的虚假位置选择算法的检测率高于“baseline”基准，无法找出 M 个合适的虚假位置。

图 8 给出了用户移动情况下的仿真结果，横坐标为用户最大移动速度 $v_{\max}$，纵坐标为平均检测率。主要仿真参数如表 1 所示。该仿真主要是为了研究用户移动对虚假位置选择算法的影响，因此仅考虑本文提出的算法以及其在移动场景中的应用，即“proposed: moving scenario”。从图中可以看出，随着用户最大移动速度的变化，平均检测率的性能几乎不变。由于用户移动后所处位置的不同，可能该位

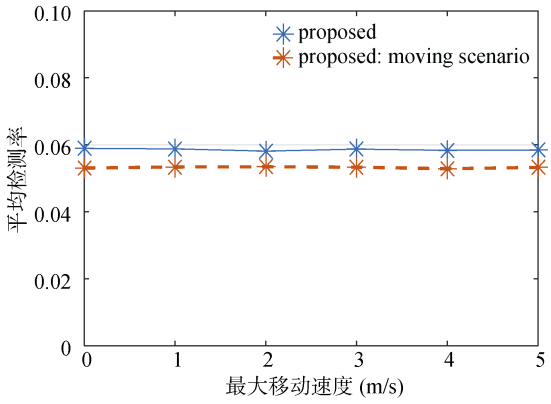


图 8　不同 v_{max} 时平均检测率

Figure 8　Average detection rate versus v_{max}

置的历史请求率有所变化。因此在用户移动场景中，本文提出的虚假位置选择算法在初始位置处选出的候选虚假位置集合并不一定能很好适用于用户移动后所处位置，导致用户移动场景的检测率性能低于用户静止的场景。但因为在不同位置处请求率相差有限，因此本文提出的算法在移动场景也具有一定的鲁棒性。

6.2　归一化总泛化面积

验证“基于泛化面积的虚假位置选择”算法的性能。该算法的最终目标是获得最大的总泛化面积，使得隐私保护区域最大化，隐私保护的效果最好。为此考察该算法对总泛化面积的影响。默认的仿真参数如表 2。

表 2　主要仿真参数

Table 2　Main simulation parameters

仿真参数	值
$\delta_n, n \in \mathcal{N}$	0.5
M	10
R	10
K	5
l	400 m
w	200 m

为了显示提出虚假位置选择算法的优势，本文将对比几种相关的虚假位置选择算法对总泛化面积的效果。本文提出的虚假位置选择算法记作“proposed”算法。第二种方法是从所有候选虚假位置集合中选择具有最大距离乘积的虚假位置[14]，记作“dist-prod”算法。第三种方法是从所有候选虚假位置集合中选择具有最大距离和的虚假位置，记作“dist-sum”算法。最后一种方法是从所有候选虚假位置集合中随机选择出 *K-1* 个虚假位置，记作“random”算法。

图 9 我们评估不同候选虚假位置个数对归一化总泛化面积的影响。从图中可以看出，随着候选虚假位置区域数目的增加，提出的虚假位置选择算法可以得到更大的归一化总泛化面积。这是因为在 K 值一定的情况下，候选虚假位置集合中的虚假位置区域可能会相交或被物体遮挡，达不到预期的总泛化面积，提出的算法中有考虑该因素影响，因此可以获得更多的选择增益。由于其余算法没有考虑实际虚假位置区域的情况，因此提出算法的性能优于其他算法。此外，我们观察到“dist-prod”算法比“dist-sum”算法的性能好，也验证了当不考虑实际虚假位置区域时，“dist-prod”算法可以获得最大的归一化总泛化面积。

图 10 的横坐标为虚假位置区域的半径，纵坐标为归一化总泛化面积。归一化总泛化面积随着虚假位置区域半径的增加而增加，这是由于半径增加会导致每个虚假位置区域的泛化面积增大。我们观察

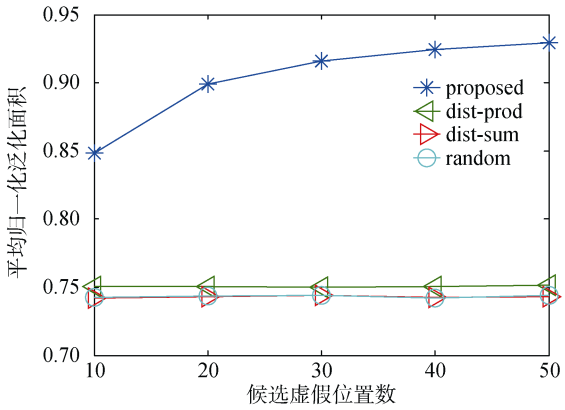


图 9　不同虚假位置数时平均归一化泛化面积

Figure 9　Average normalized cloaking area versus the number of candidate dummy locations

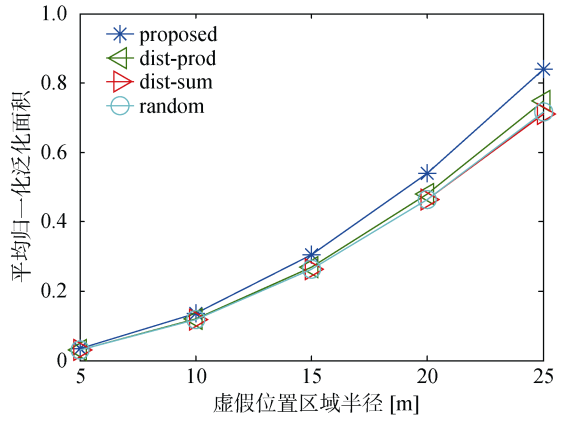


图 10　不同虚假位置区域半径时平均归一化泛化面积

Figure 10　Average normalized cloaking area versus the radius of location region

到本文提出的虚假位置选择算法得到了最好的性能。尽管“dist-prod”算法性能优于“dist-sum”算法和“random”算法, 但该算法的性能仍然受到虚假位置区域实际的情况限制, 比如当虚假位置区域被物体遮挡或相交的情况。

我们也评估 K 值对归一化总泛化面积的影响。如图 11 所示, 当 K 值增加时, 所有算法的归一化总泛化面积也逐渐增加。这是由于 K 值的增加意味着更多的匿名区域参与, 也带来了更大的泛化面积。同时也能发现提出的算法由于考虑了虚假位置区域的实际情况, 得到了最好的性能。注意到当 $K=1$ 和 $K=N$ 的情况下, 所有算法得到的结果相同。当 $K=1$ 时, 意味着只能上报一个位置区域, 那就是真实用户位置, 因此所有算法得到的性能一致。而当 $K=N$ 时, 相当于所有的虚假位置区域都被选为参与匿名隐私保护中, 所以所有算法获得的性能一致。

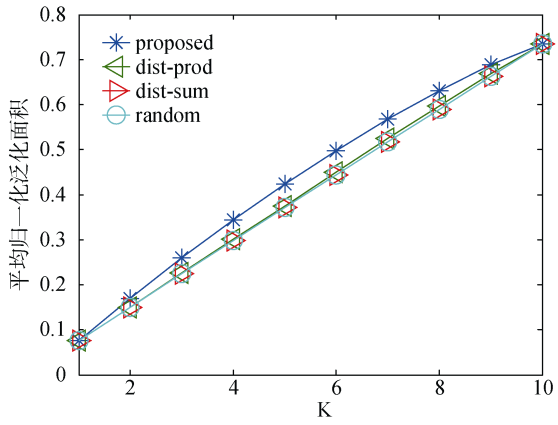


图 11 不同 K 时平均归一化泛化面积

Figure 11 Average normalized cloaking area versus K

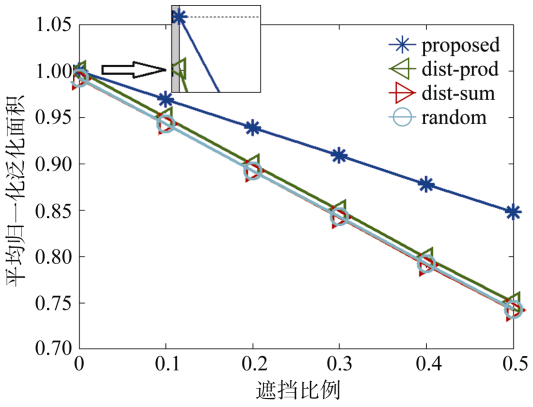


图 12 不同遮挡比例时平均归一化泛化面积

Figure 12 Average normalized cloaking area versus occupied percentage

最后, 图 12 中比较了不同遮挡比例对归一化总泛化面积的影响。由于越大的遮挡比例, 每个虚假位置实际的泛化面积变小, 因此归一化总泛化面积也相应减小。本文提出的算法能在不同的遮挡比例情况下获得最优的性能, 证明了该算法能够有效考虑遮挡比例的影响, 选出最合适的虚假位置集合。同时从图中也观察到当遮挡比例为零的时候, 即不存在遮挡的情况下, 所有算法性能之间的差异几乎消失, 说明了遮挡对算法的影响。在局部放大的图形中可以看出提出的算法仍然优于其他算法, 因为其考虑了相互重叠的情况, 但相交的不多。当虚假位置区域之间重叠比例越大时, 该算法的优势会体现更明显。

图 13 给出了用户移动情况下的仿真结果, 横坐标为用户最大移动速度 $v_{\max}$, 纵坐标为平均归一化泛化面积。主要仿真参数如表 2 所示。该仿真主要

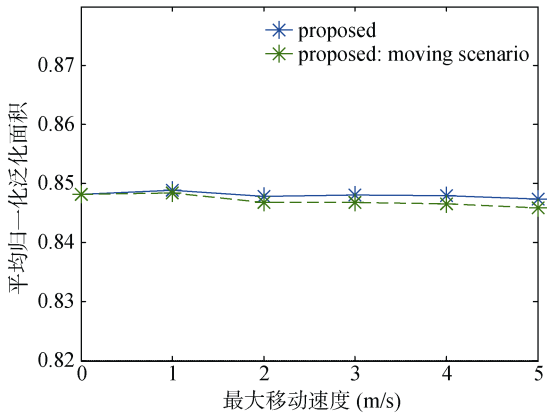


图 13 不同 $v_{\max}$ 时平均归一化泛化面积

Figure 13 Average normalized cloaking area versus $v_{\max}$

是为了研究用户移动对虚假位置选择算法的影响, 因此仅考虑本文提出的算法以及其在移动场景中的应用, 即“proposed: moving scenario”。从图中可以看出, 随着用户最大移动速度的变化, 在用户静止场景下的平均归一化泛化面积几乎不变。当用户最大移动速度为 0 的时候, 静止场景和移动场景的算法性能一致。而随着移动速度的增加, 在用户移动场景下的平均归一化泛化面积有所降低, 这说明在初始位置处选出的虚假位置并不能很好适用于用户移动后所处位置。这是因为用户移动后, 可能会改变原有的相交或遮挡的情况。另一方面, 从图中还可以看出, 在最大移动速度有限的情况下, 本文提出的算法在移动场景也具有一定的鲁棒性。即当最大移动速度在一定范围内的情况下, 对本文提出的虚假位置选择算法性能影响不大。

7　结论与展望

本文研究了*K*匿名技术中的虚假位置选择问题。为降低求解复杂度, 本文提出了两步骤虚假位置选择算法, 首先通过选择出与真实用户请求率差异在给定阈值内的一组候选虚假位置集合以保证隐私等级, 然后考虑虚假位置区域的实际情况在候选集合中选择出 *K-1* 个虚假位置区域, 以最大化总泛化面积。为了更加准确衡量总泛化面积, 本文针对圆形虚假位置区域相交的情况进行推导。最后通过安全分析以及数值仿真, 验证了提出算法在隐私等级保护以及总泛化面积最大化方面的性能。

由于本文内容有限, 提出的虚假位置选择算法对于多目标优化问题仅是次优解。为了更深入了解该优化问题能达到的理论性能, 则需要经过适当的变换设计更准确的最优解。同时, 对于虚假位置区域实际相交或被遮挡情况研究应根据实际情况细化, 这也是未来研究的方向之一。另一方面, 本文的研究假设用户是处于静止状态, 或者以较小的速度移动。但当用户移动一定距离后所处新位置的历史请求率、可选的虚假位置集合以及与可选虚假位置集合的重叠情况均有所不同, 这都将对虚假位置选择算法产生影响。因此如何在移动场景下研究有效的虚假位置选择算法也很有意义。

附录: 圆形位置区域相交面积计算

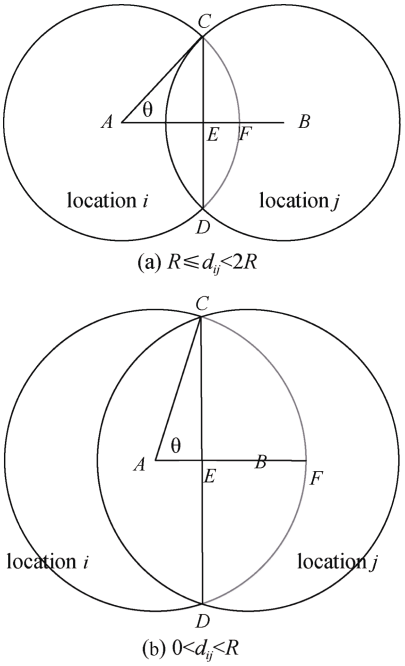


图 14　两位置区域相交示意图

Figure 14　Sketch of intersected area between two location regions

当两个选定的虚假位置区域 *i* 和 *j* 相互靠近的时候, 它们将彼此相交。令两个虚假位置区域圆心之间的距离为 d_{ij} 。为便于描述, 本文使用图 14(a)和图 14(b)分别描绘了两虚假位置区域相交面积的两种情况。令 $\widetilde{xyz}$ 表示图中 *x*、*y*、*z* 围成的区域。从图中可以得知, 虚假位置区域 *i* 和 *j* 相交面积等于 $\widetilde{CEF}$ 区域面积的 4 倍。因此, 计算相交面积的关键即求解 $\widetilde{CEF}$ 区域面积。为了使计算方法适用于 $R \leqslant d_{ij} < 2R$ (图 14(a))和 $0 < d_{ij} < R$ (图 14(b))两种情况, $\widetilde{CEF}$ 区域面积等于 $\widetilde{ACF}$ 的扇形面积减去三角形 ΔACE 的面积, 即:

$$A_{ij}^{I} = 4 \cdot (S_{\widetilde{ACF}} - S_{\Delta ACE}) \tag{10}$$

$\widetilde{ACF}$ 的面积可以表示为:

$$S_{\widetilde{ACF}} = \frac{\arcsin\left(\dfrac{\sqrt{R^2 - d_{ij}^2/4}}{R}\right)}{360}\pi R^2 \tag{11}$$

三角形区域 ΔACE 的面积为:

$$\begin{aligned} S_{\Delta ACE} &= \frac{1}{2}\overline{AE} \cdot \overline{CE} \\ &= \frac{d_{ij}}{4}\sqrt{R^2 - d_{ij}^2/4} \end{aligned} \tag{12}$$

将公式(11)和(12)代入公式(10)中, 即可得到相交面积为:

$$A_{ij}^{I} = \frac{\arcsin\left(\dfrac{\sqrt{R^2 - d_{ij}^2/4}}{R}\right)}{90}\pi R^2 - d_{ij}\sqrt{R^2 - d_{ij}^2/4} \tag{13}$$

参考文献

[1] J. Prinsloo and R. Malekian, “Accurate vehicle location system using rfid, an internet of things approach,” *Sensors*, vol. 16, no. 6, pp. 1-24, 2016.

[2] M. Lukic and I. Mezei, “Localised querying and location update service in wireless sensor and robot networks with arbitrary topology,” *International Journal of Ad Hoc & Ubiquitous Computing*, vol. 22, no. 1, pp. 48–61, 2016.

[3] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, “Tagoram: realtime tracking of mobile rfid tags to high precision using cots devices,” *in Proceedings of the 20th annual international conference on Mobile computing and networking. ACM*, pp. 237–248, 2014.

[4] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, “Landmarc: indoor location sensing using active rfid,” *Wireless networks*, vol. 10, no. 6, pp. 701–710, 2004.

[5] M. Conti, J. Willemsen, and B. Crispo, “Providing source location privacy in wireless sensor networks: A survey,” *Communications Surveys & Tutorials IEEE*, vol. 15, no. 3, pp. 1238–1280, 2013.

[6] L. Zhang, C. Fang, Y. Li, H. Zhu, and M. Dong, “Optimal strate-

gies for defending location inference attack in database-driven crns," 2015 IEEE International Conference on Communications (ICC), pp. 7640–7645, 2015.

[7] C. Bettini, X. S. Wang, and S. Jajodia, "Protecting privacy against location-based personal identification," *Secure data management. Springer*, pp. 185–199, 2005.

[8] L. SWEENEY, "k-anonymity: A model for protecting privacy," International Journal of Uncertainty, *Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 1–14, 2012.

[9] M. Guo, N. Pissinou, and S. S. Iyengar, "Pseudonym-based anonymity zone generation for mobile service with strong adversary model," *in Consumer Communications and NETWORKING Conference*, pp. 335–340, 2015.

[10] H. Lu, C. S. Jensen, and M. L. Yiu, "Pad: privacy-area aware, dummy based location privacy in mobile services," *in Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, pp. 16–23, 2008.

[11] T. Hara, A. Suzuki, M. Iwata, and Y. Arase, "Dummy-based user location anonymization under real-world constraints," *IEEE Access*, vol. 4, pp. 673 - 687, 2016.

[12] C. Y. T. Ma, D. K. Y. Yau, N. K. Yip, and N. S. V. Rao, "Privacy vulnerability of published anonymous mobility traces," *IEEE/ACM Transactions on Networking*, vol. 21, no. 3, pp. 720–733, 2013.

[13] X. Liu, K. Liu, L. Guo, and X. Li, "A game-theoretic approach for achieving k-anonymity in location based services," *Proceedings – IEEE INFOCOM*, vol. 12, no. 11, pp. 2985–2993, 2013.

[14] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving k-anonymity in privacy-aware location-based services," *INFOCOM*, 2014 Proceedings IEEE, pp. 754–762, 2014.

[15] B. Niu, Z. Zhang, X. Li, and H. Li, "Privacy-area aware dummy generation algorithms for location-based services," *2014 IEEE International Conference on Communications (ICC)*, pp. 957–962, 2014.

[16] M. Gruteser and D. Grunwald. "Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking." *International Conference on Mobile Systems, Applications, and Services*, pp. 31-42, 2003.

[17] G Bu, L. Liu. "A Customizable k-Anonymity Model for Protecting Location Privacy". *Icdcs*, pp. 620--629, 2004.

[18] T. H. You, W C Peng, W C Lee. "Protecting Moving Trajectories with Dummies". *International Conference on Mobile Data Management*, pp. 278-282, 2007.

[19] X Liu, K Liu, L Guo, et al. "A game-theoretic approach for achieving k-anonymity in Location Based Services", *INFOCOM*, 2013 Proceedings IEEE, pp. 2985-2993, 2013.

[20] X Zhu, H Chi, B Niu, et al. "MobiCache: When k-anonymity meets cache". *GLOBECOM 2013 - 2013 IEEE Global Communications Conference*, pp. 820-825, 2013.

[21] P K Sahu, S K Chandra. "Location Privacy Using User Anonymity and Dummy Locations". *International Journal of Innovative Technology & Creative Engineering*, 2012, 2(5).

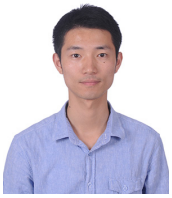
[22] K Liu. "Dummies and Nearest Neighbor Based Location Privacy Protection". *Journal of Information & Computational Science*, vol. 10, no. 12, pp. 3831-3839, 2013.

[23] N Xu, D Zhu, H Liu, et al. "Combining Spatial Cloaking and Dummy Generation for Location Privacy Preserving". *International Conference on Advanced Data Mining and Applications*, pp. 701-712, 2012.

[24] D Liao, H Li, G Sun, et al. "Protecting User Trajectory in Location-Based Services". *GLOBECOM 2015 - 2015 IEEE Global Communications Conference*, pp. 1-6, 2015.

[25] B Korte, J Vygen. "Combinatorial Optimization: *Theory and Algorithms*". Springer, 2000.

[26] W Navidi, T Camp. "Stationary Distributions for the Random Waypoint Mobility Model". *IEEE Transactions on Mobile Computing*, vol. 3, no.1, pp. 99-108, 2004.



吴荻 于 2014 年在北京交通大学通信与信息系统专业获得博士学位。现任中科院信息工程研究所助理研究员。研究领域为移动网络、传感网等。研究兴趣包括：物联网、隐私保护、无线资源分配等。Email: wudi@iie.ac.cn

张玉 于 2013 年在北京交通大学通信与信息系统专业获得工学博士学位。现在在中国科学院信息工程研究所攻读博士后。研究领域为移动互联网络。研究兴趣包括移动自组织网络、内容中心网络、移动互联网安全。Email: yuzhang1984@iie.ac.cn

刘银龙 于 2011 年在北京邮电大学通信与信息系统专业获得博士学位。现任中国科学院信息工程研究所副研究员。研究领域为：通信与信息系统。研究兴趣包括：无线通信系统与安全、大数据与网络安全。Email: liuyinlong@iie.ac.cn

马伟 于 2010 年在北京邮电大学通信与信息系统专业获得博士学位。现任中国科学院信息工程研究所副研究员。研究领域为：通信与信息系统。研究兴趣包括：无线通信系统与安全、大数据与网络安全。Email: mawei@iie.ac.cn

朱大立　于 2007 年在华中科技大学获得计算机应用技术专业博士学位。现任中国科学院信息工程研究所正研级高级工程师，博士生导师。研究领域移动互联网安全和无线网络攻防技术，研究兴趣包括：智能终端安全、应用安全、无线管控等技术。Email: zhudali@iie.ac.cn

孙鑫　现任中科院信息工程研究所高级工程师。研究领域为移动互联网安全、数据分析、工程管理等。研究兴趣包括：终端安全、智能分析、车联网等。Email: sunxin@iie.ac.cn