基于位置语义和查询概率的假位置选择算法

王洁[1]，王春茹[1]，马建峰[2]，李洪涛[1]
（1. 山西师范大学数学与计算机科学学院，山西 临汾 041099；
2. 西安电子科技大学网络与信息安全学院，陕西 西安 710071）

摘 要：针对传统位置隐私保护方案中未充分考虑攻击者拥有背景知识而导致的隐私泄露问题，基于位置语义和查询概率提出一种假位置选择算法。在假位置集中的位置之间满足语义差异性、查询概率相近且地理位置尽量分散的条件下，避免了攻击者结合背景知识过滤假位置，同时保证了查询结果的精确性。仿真实验验证了所提算法能有效保护用户的位置隐私。

关键词：基于位置的服务；假位置隐私；位置语义；查询概率
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2020061

Dummy location selection algorithm based on location semantics and query probability

WANG Jie[1], WANG Chunru[1], MA Jianfeng[2], LI Hongtao[1]
1. College of Mathematics & Computer Science, Shanxi Normal University, Linfen 041099, China
2.School of Cyber Engineering, Xidian University, Xi'an 710071, China

Abstract: Aiming at the problem of privacy leakage caused by attackers possessing background knowledge in traditional location privacy protection schemes, a dummy location selection algorithm based on location semantics and query probability was proposed. Under the conditions that the locations in the dummy location set satisfied semantic difference, similar query probability, and geographically dispersed, it avoided attackers who filter dummy locations by combining background knowledge, and the accuracy of query results was guaranteed. Simulation experiments verify that the proposed algorithm can effectively protect the user's location privacy.

Key words: location based service, dummy location privacy, location semantics, query probability

1 引言

近年来，随着手机等移动智能终端的普及和移动互联网技术的发展，基于位置的服务（LBS, location based service）在日常生活中得到广泛的应用。用户下载 LBS 应用程序后，使用这些应用程序可以很方便地向 LBS 服务器发送请求，从而获取与某个兴趣点相关的信息，如交通导航信息、附近的餐厅等。然而，人们在享受 LBS 提供便利的同时，也面临着严重的隐私泄露风险。不受信任的 LBS 服务器会获取用户的查询数据，如用户的位置、提交的查询类型等。根据这些数据，LBS 服务提供者可以进一步推断出用户的一些个人敏感信息，如目标用户的社会关系、家庭住址、日常行为等；也可以直接追踪用户或泄露用户的个人信息，这将导致用户的隐私泄露，因此，位置隐私的保护变得越来越重要。

针对位置隐私保护问题，国内外研究学者提出

收稿日期：2019−08−27；**修回日期**：2020−02−25
基金项目：国家自然科学基金资助项目（No.61702316）；山西省自然科学基金资助项目（No.201901D111280, No.201801D221177）；山西省软科学基金资助项目（No.2017041016-4）
Foundation Items: The National Natural Science Foundation of China (No.61702316), The Natural Science Foundation of Shanxi Province (No.201901D111280, No.201801D221177)；Soft Science Project of Shanxi Province (No.2017041016-4)

了多种用户位置隐私保护方法，其中 *k*-匿名是最常用的方法之一。现有的位置 *k*-匿名技术通常依靠可信第三方服务器将用户的准确位置泛化为一个包含 *k* 个用户（包含目标用户）的区域[1-3]。这样不受信任的 LBS 服务器很难从此区域中区分出目标用户的真实位置。但是上述隐私保护模型存在一定的局限性。首先，存在单点故障。因为它依赖于一个受信任的匿名服务器，一旦匿名器服务器被对手控制，那么所有用户的隐私都将受到损害；其次，很难平衡服务可用性和位置隐私，如果泛化区域过大，会影响服务质量，泛化区域过小，又容易造成位置泄露。为了在不受第三方影响的情况下保护用户的位置隐私，有研究者提出了同样能实现 *k*-匿名的假位置技术，但是如何选择合适的假位置是一个难题。现有的假位置选择方案考虑到对手已掌握了位置查询概率，提出基于熵度量选择虚拟位置的方法，保护了用户的位置隐私[4-6]。然而当真实位置和 *k*−1 个假位置都是同一语义类型时，对手还是可以很容易地推断出用户的个人信息，例如所有的地点都位于学校时，对手就能推断出用户的身份可能是教师或者学生，因此，在选择假位置时要充分考虑位置的语义信息，生成的假位置集需满足语义差异性。现有的方案[7-8]大多采用欧氏距离或余弦距离度量语义差异度，计算量庞大，影响算法的执行效率，降低用户的服务质量。文献[9]采用树形结构组织所有位置，并根据位置节点间的跳数来计算语义距离，将语义距离作为语义差异度的度量标准，使生成的假位置集满足语义差异性，但是当假位置集中包含湖泊、森林等访问概率较低的位置时，对手很容易将它们过滤；此外，如果生成的假位置都在真实位置附近，对手就会将真实位置锁定在一个小区域，进而推测出真实位置。

本文充分考虑了用户边信息和背景知识可能被攻击者手利用的情况，提出了一种基于位置语义和查询概率的最大最小假位置选择（MMDS, maximum and minimum dummy selection）算法。在选择假位置时，首先保证最后生成的假位置集中 *k* 个位置的语义满足差异性；然后，使假位置集中 *k* 个位置间的查询概率尽量相近；最后，使假位置间的地理位置尽量分散，并针对考虑位置之间在地图上分散且查询概率尽可能相近这 2 个因素进行了多目标优化。本文算法解决了攻击者具有边信息可以排除部分假位置的问题，能有效保护用户的位置隐私。仿真实验分别从假位置集生成时间、物理分散度、语义差异性以及位置熵这 4 个方面将本文算法与已有算法进行对比，验证了本文算法的有效性。

2 预备知识

2.1 系统架构

本文采用的系统架构如图 1 所示，主要由 Wi-Fi 接入点（AP, access point）、LBS 服务器及智能终端三部分组成。

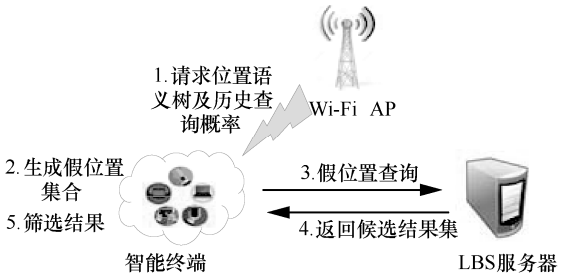


图 1 系统结构

目前，Wi-Fi 覆盖面较广，计算能力和存储能力也较强。本文系统中 Wi-Fi AP 可提供网络支持，计算并存储当前覆盖范围内所有位置的历史查询概率，在其无线电范围内收集位置语义信息，并生成和保存位置语义树。对任意 Wi-Fi AP，它所覆盖范围内的位置是相对稳定的，所以位置语义树和历史查询概率不需要频繁变化。

各种智能终端设备用于执行假位置选择算法，生成假位置集，并筛选由 LBS 服务器返回的候选结果集。

LBS 服务器提供基于位置的服务。本文假设攻击者能够成功攻破 LBS 服务器，并且攻击者所掌握的边信息包括用户的匿名机制、位置的历史查询概率以及位置的语义信息和地图信息。

查询前，智能终端用户首先请求 Wi-Fi AP，获取当前 Wi-Fi AP 覆盖范围内的地图信息、位置语义树及历史查询概率；其次，根据假位置选择机制生成满足自身隐私保护需求的 *k*−1 个假位置；然后，用户将真实位置和 *k*−1 个假位置发送至 LBS 服务器请求查询，LBS 服务器根据用户发送的位置返回候选结果集；最后，用户筛选结果，得到当前查询的目标结果。

2.2 假位置集构造原则

定义 1 用户的真实位置 l_{real}。l_{real} 包含用户位置

的地理坐标及语义信息。

定义 2　用户隐私需求 S。以二元组(k,u)表示某用户的隐私需求 S，其包含以下 2 个方面含义。

1) 匿名度 k，表示每次查询时都发送一个真实位置和至少 k–1 个假位置给服务器，且判断出真实位置的概率为 $\frac{1}{k}$。

2) 语义差异度 u，表示假位置集中任意 2 个位置间的语义距离的最小可接受值。本文设置 $u=4$，即对假位置集中任意 2 个位置 l_i 和 l_j，满足 $[d_{\text{sem}}(l_i,l_j)]\min \geqslant u$ (1)，则生成的假位置集中位置之间满足语义差异性。

2.3　位置地图距离计算方法

令 Map_{cur} 表示当前 Wi-Fi AP 覆盖范围内的地图信息，对于任意 2 个位置 $l_i,l_j(i\neq j)$，位置地图距离 $d_{\text{phy}}(l_i,l_j)$ 为 2 个位置在地图信息 Map_{cur} 上的距离，其取值范围为几米到几千米。

2.4　位置查询概率距离计算方法

定义 3　位置查询概率（LQP, location query probability）。将每个 Wi-Fi AP 将其覆盖范围内的地图作为样本空间按照网格进行划分，如图 2 所示，每个网格中包含一个位置单元，确定每个位置单元的坐标，通过数据训练集来训练每个位置的查询概率。使用 $q_i=\frac{n_i}{\sum_{i=1}^{m^2}n_i}$ 计算每个位置单元的历史查询概率并保存，其中，$i=1,2,\cdots,m^2$，n_i 表示某个位置单元的查询次数，$\sum_{i=1}^{m^2}n_i$ 表示所有位置单元的总查询次数，$\sum_{i=1}^{m^2}q_i=1$。

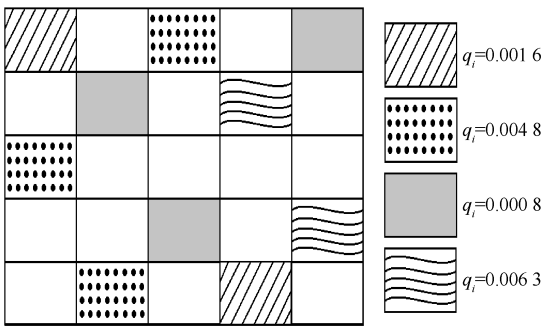


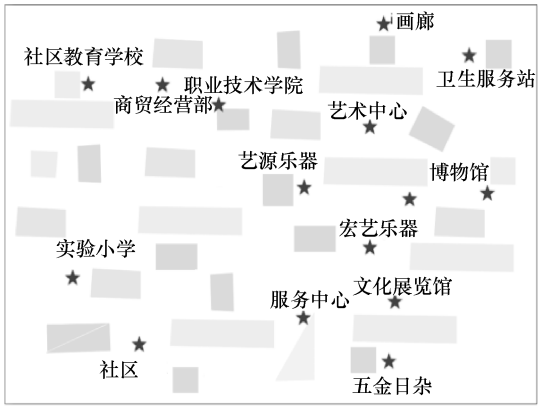
图 2　样本空间划分

对于任意 2 个位置 $l_i,l_j(i\neq j)$，位置查询概率距离即 2 个位置查询概率之间的差值，用 $d_{\text{que}}(l_i,l_j)$ 表示。

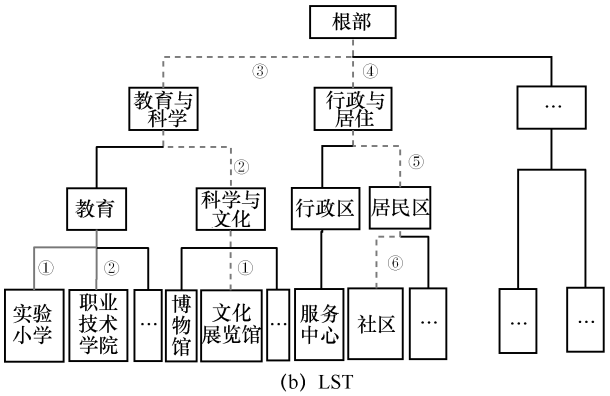
2.5　位置语义距离计算方法

定义 4　位置语义树（LST, location semantic tree）。将当前 Wi-Fi AP 覆盖的所有位置根据其语义组织为一种树形结构，每个叶节点表示地图上的真实位置 l_{real}，每个非叶节点表示其子节点的类别，LST 的深度为 h，h 的值等于 LST 中分类的层数加 1。

对于任意 2 个位置 $l_i,l_j(i\neq j)$，语义距离 $d_{\text{sem}}(l_i,l_j)$ 即位置语义树中 2 个位置叶节点之间的跳数。下面举例说明语义距离计算方法。以某一 Wi-Fi AP 覆盖的一块真实的地图信息为例，图 3(a)为该区域的示意图，★表示真实位置。将图 3(a)中位置分为教育与科学、行政与居住等类型，在图 3(b)中显示所形成的部分 LST，其叶节点表示真实位置，非叶节点表示子节点的类别。根据此 LST 可以得出实验小学与职业技术学院（图中灰色实线路径）的语义距离为 2，文化展览馆与社区（图中灰色虚线路径）的语义距离为 6。



(a) 真实地图



(b) LST

图 3　真实地图信息及其 LST

根据上述分析，最后的假位置集需要满足位置在地图上尽量分散且其历史查询概率尽可能相近，即最后的假位置集 DLS 需同时满足式(2)

和式(3)。

$$\mathrm{DLS}=\arg\max\{\min[d_{\mathrm{phy}}(l_i,l_j)]\},l_i,l_j\in\mathrm{DLS},i\neq j \quad (2)$$

$$\mathrm{DLS}=\arg\min\{\max[d_{\mathrm{que}}(l_i,l_j)]\},l_i,l_j\in\mathrm{DLS},i\neq j \quad (3)$$

由于需要同时考虑 2 个因素，可将其转化为多目标优化问题，由此本文提出一种最大最小假位置选择（MMDS）算法，该算法确保 DLS 满足式(4)。

$$\mathrm{DLS}=\arg\max\left\{\frac{\min[d_{\mathrm{phy}}(l_i,l_j)]}{\max\{r[d_{\mathrm{que}}(l_i,l_j)+1]\}}\right\} \quad (4)$$

其中，$l_i,l_j\in\mathrm{DLS}$，$i\neq j$。$d_{\mathrm{que}}(l_i,l_j)+1$ 避免了两位置概率相同，即差值为 0 的情况。为了平衡地理距离和概率差值这 2 个因素，设置参数 r=100。通过式(4)使假位置在地图上尽量分散且其历史查询概率尽可能相近。

3 最大最小假位置选择算法

根据假位置集构造原则，选择合适的假位置构成假位置集 DLS。本文采用边选择边判断的策略，其主要思想是，选择与当前假位置集中已有位置满足语义差异的所有位置作为候选假位置集；然后，在候选假位置集中选择一个最优位置，最优位置是指由该位置构成的 DLS 满足式(3)；最后，生成一个由用户所在真实位置和 k–1 个假位置构成的集合 DLS。算法 1 给出了最大最小假位置选择算法的伪代码。

算法 1 最大最小假位置选择（MMDS）算法

输入 真实位置 l_{real}，用户隐私需求 S，当前 Wi-Fi AP 覆盖范围内的地图信息 $\mathrm{Map}_{\mathrm{cur}}$，位置语义树 LST，位置查询概率 LQP

输出 假位置集 DLS

1) 将 $\mathrm{Map}_{\mathrm{cur}}$ 中所有位置生成候选假位置集 CLS；

2) 由 LST 得到语义距离矩阵 **SDM**；由 $\mathrm{Map}_{\mathrm{cur}}$ 得到地理距离矩阵 **GDM**；由 LQP 得到概率距离矩阵 **PDM**；

3) 将真实位置 l_{real} 加入 DLS，并将其从 CLS 中移除；

4) while |DLS|< k do //当假位置集中的位置个数小于 k 时

5) if(CLS=Φ) then //如果 CLS 为空

6) 匿名失败；

7) else

8) max=0；BestLoc=Φ；

9) for each Loc in CLS

10) if(d_{sem}(Loc,DLS.last) ⩽ u) then

11) 将 Loc 从 CLS 中移除；//将不满足语义差异性的位置去掉

12) continue；

13) else

14) $m=\dfrac{d_{\mathrm{phy}}(\mathrm{Loc},\mathrm{DLS.last})}{r[d_{\mathrm{que}}(\mathrm{Loc},\mathrm{DLS.last})+1]}$；

15) 用 max 记录 m 的当前最大值，并将对应的 Loc 赋值给 BestLoc；//从当前 CLS 中找出最佳位置

16) end if

17) end for

18) 将 BestLoc 放入 DLS 中；

19) 将 BestLoc 从 CLS 中移除；

20) end if

21) end while

22) return DLS

算法 1 的输入参数为用户的真实位置 l_{real}、用户隐私需求 S（包含隐私需求参数 k 和语义差异性需求参数 u）、当前 Wi-Fi AP 覆盖范围内的地图信息 $\mathrm{Map}_{\mathrm{cur}}$、位置语义树 LST、位置查询概率 LQP。首先，将地图信息 $\mathrm{Map}_{\mathrm{cur}}$ 中的所有位置生成候选假位置集 CLS（第 1）行），然后由位置语义树 LST 得到语义差异度矩阵 **SDM**，由 $\mathrm{Map}_{\mathrm{cur}}$ 得到位置地理距离矩阵 **GDM**，由位置查询概率 LQP 生成概率距离矩阵 **PDM**（第 2）行），将真实位置加入假位置集 DLS 中，并将其从 CLS 中移除，根据 SDM 将 CLS 中与 DLS 新加入的假位置（DLS.last）语义距离小于 u 的位置过滤掉（第 3)～11)行），再根据 LDM 及 QDM 从 CLS 剩余位置中选择与 DLS.last 满足 $\max\left\{\dfrac{d_{\mathrm{phy}}(\mathrm{Loc},\mathrm{DLS.last})}{r[d_{\mathrm{que}}(\mathrm{Loc},\mathrm{DLS.last})+1]}\right\}$ 的位置 Loc 作为新的假位置放入 DLS（第 14)～18)行），并将此位置从 CLS 中移除（第 19)行），循环执行以上语句（第 5) ～20)行），直到 DLS 的个数等于 k，最后返回最优假位置集 DLS。

4 算法理论分析

4.1 安全性分析

由系统架构可知，本文的假位置选择算法在智

能终端设备上进行时，智能终端设备不会将其准确位置坐标发送给任何实体。所以攻击者无法通过链路攻击来获取用户的准确位置。

攻击能力较强的强攻击者企图根据已获得的地图和语义信息以及历史查询概率数据来推测用户的真实位置，本文算法可以有效抵抗此类攻击者。首先，在选择假位置时保证了其与当前假位置集 DLS 中已有位置的语义均不同，即对于任意 2 个位置 l_i、l_j，满足 $d_{\text{sem}}(l_i,l_j)\geqslant u$，使攻击者无法推断用户的位置语义信息；其次，假位置间的查询概率接近，即满足 $p_i\approx p_j$（p_i,p_j 表示位置 l_i、l_j 的历史查询概率），根据位置熵的计算式 $H=-\sum_{i=1}^{k}q_i\text{lb}q_i$（$q_i=\frac{p_i}{\sum_{i=1}^{k}p_i}$）可知，查询概率越接近时信息熵越大，且对于含 k 个位置的假位置集 DLS，攻击者推出真实位置的概率约为 $\frac{1}{k}$，因此本文算法可以有效抵抗强攻击者，保护用户的位置隐私。

4.2　算法复杂度分析

MMDS 算法开始时，假位置集 DLS 中只有用户的真实位置 l_{real}，假设候选位置有 n 个。MMDS 算法分为 2 个阶段，第一阶段是通过比较将候选位置中与 DLS 中新加入的假位置语义相同的位置过滤掉，第二阶段是通过式(3)从剩余候选位置中选择出最优假位置加入 DLS。每个阶段的时间复杂度均为 $O(n)$，因此总的时间复杂度为 $O(n)$。

5　实验结果与分析

5.1　实验设置

实验选用某城市真实地图数据，其中心城区已经被 Wi-Fi 全面覆盖且拥有大量的 LBS 用户。每个 Wi-Fi AP 覆盖范围约为 700~800 m，样本空间被均匀划分为 28×28 的矩形网格，共 13 579 个样本轨迹点作为历史数据，计算每个网格内地理位置的历史查询概率。实验将位置语义主要分为六大类，分别为教育与科学、行政与居住、医疗救护、商城、公共场所和餐饮娱乐。

实验采用 MyEclipse 开发平台，以 Java 编程语言实现。硬件环境为 Windows 7 操作系统，3.40 GHz Intel Core i7 处理器，4 GB 内存。实验参数如表 1 所示。

表 1　实验参数

参数	默认值	取值范围
k	5	[2,10]
语义差异度 u	4	[3,8]
网格数	28×28	
语义位置分类	教育与科学、行政与居住、医疗救护、商城、公共场所、餐饮娱乐	
LST 的深度 h	5	[3,7]
Wi-Fi AP 覆盖范围/m	700	
地图位置的数目 Map_i.LN /个	50	[20,60]

5.2　实验结果

首先，通过分析匿名成功率对 MMDS 评价；然后，将 MMDS 样考虑位置语义的 MaxMinDistDS 算法[9]和 SimpMaxMinDistDS 算法[9]从假位置集生成时间、物理分散度、语义差异性比较，以及位置熵四方面进行比较，验证 MMDS 算法的有效性。

5.2.1　匿名成功率

图 4 给出了匿名成功率相对于地图中位置的数目 Map_i.LN、匿名度 k 以及语义差异度 u 的变化情况。图 4(a)结果显示，地图中位置数越多，越有利于匿名执行，匿名成功率越高。这是因为位置数越多，

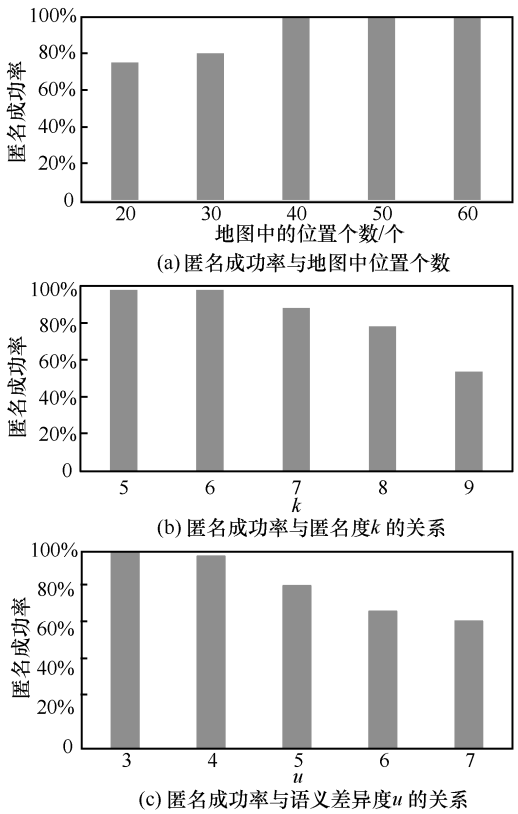


图 4　匿名成功率的变化情况

位置语义的种类数就越多，更有利于构造满足隐私需求的假位置集，匿名成功率得到提高。反之，当地图中位置的数目较少时，很难使匿名集位置间的语义互不相同，导致匿名失败。在图 4(b)中，随着隐私需求的匿名度 k 的增加，匿名成功率有所下降，因为满足语义差异度的位置数不能满足匿名度 k 的要求。图 4(c)中，随着语义差异度 u 的增加，匿名成功率同样呈下降趋势，因为语义差异度要求越高，假位置集的语义差异性越难满足。综上所述，实际操作中匿名度 k 与语义差异度 u 不能设置得过大。

5.2.2　假位置集生成效率

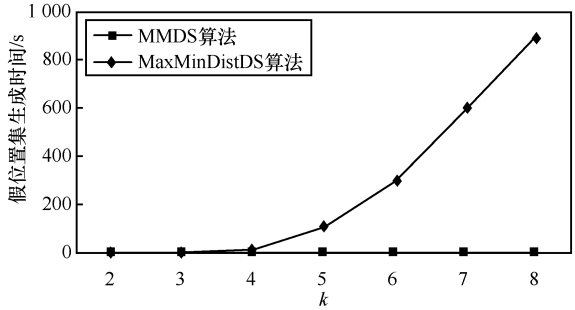
在考虑位置语义的假位置选择算法中，MaxMinDistDS 算法、SimpMaxMinDistDS 算法和 MMDS 算法生成假位置集的平均生成时间如表 2 所示。图 5 为这 3 种算法假位置集生成时间的对比结果，其中 k 的取值范围为 2~8。

表 2　　假位置集平均生成时间

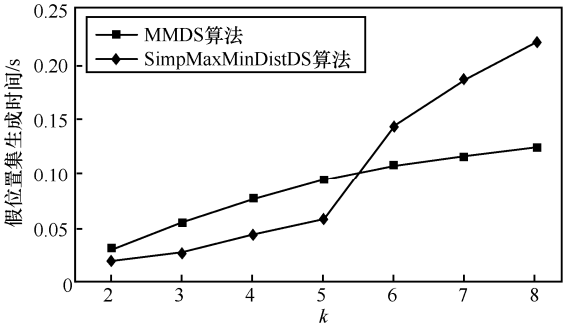
k	平均生成时间/s		
	MaxMinDistDS 算法	SimpMaxMinDistDS 算法	MMDS 算法
2	0.07	0.020	0.030
3	1.69	0.028	0.055
4	13.00	0.044	0.076
5	106.27	0.058	0.094
6	295.41	0.144	0.107
7	592.91	0.186	0.116
8	899.45	0.220	0.124

从图 5(a)可以看出，随着 k 值的增加，MMDS 算法的假位置集平均生成时间远小于 MaxMinDistDS 算法，MMDS 算法生成假位置集的效率更高。因为 MMDS 算法在选择候选位置时，过滤了与当前假位置集已有位置语义距离相近的位置，避免了不必要的时间开销。从图 5(b)可以看出，当 $k \leqslant 5$ 时，MMDS 算法的假位置集生成时间要高于 SimpMaxMinDistDS 算法，当 $k \geqslant 6$ 时，MMDS 算法的假位置集生成时间低于 SimpMaxMinDistDS 算法。

通过对比还可以看出，随着 k 值的增加，3 种算法生成假位置集所花费的时间都增加，但 MMDS 算法明显比 SimpMaxMinDistDS 算法和 MaxMinDistDS 算法的增加幅度小。也就是说，随着 k 值的增大，MMDS 算法所花费的时间更小，优势更加明显，更具有实用性。



(a) MMDS算法与MaxMinDistDS算法假位置集生成时间比较



(b) MMDS算法与SimpMaxMinDistDS算法假位置集生成时间比较

图 5　假位置集生成时间随 k 的变化

5.2.3　物理分散性比较

假位置间的距离越大则位置越分散，本文通过比较假位置集中任意两位置间的最小距离来衡量假位置集的物理分散性，最小距离越大说明越分散。图 6 表示本文 MMDS 算法、MaxMinDistDS 算法、SimpMaxMinDistDS 算法这 3 种算法在不同 k 值下假位置间的最小距离。

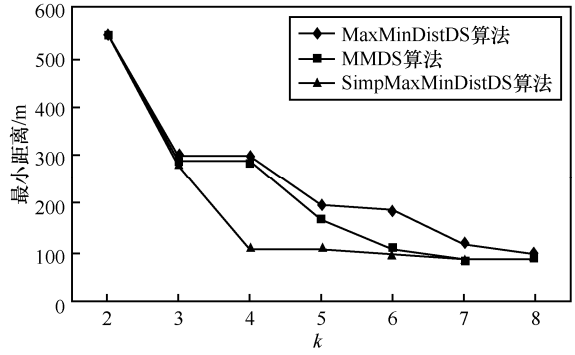


图 6　假位置间的最小距离

通过实验对比可以看出，当 $k \leqslant 4$ 时，MMDS 算法与 MaxMinDistDS 算法的最小距离接近；$k \geqslant 5$ 时，MaxMinDistDS 算法的最小距离要大于 MMDS 算法。在相同 k 值下，MMDS 算法的最小距离略大于 SimpMaxMinDistDS 算法。这 3 种算法假位置间的最小距离都随 k 值的增加而呈下降趋势，且随着 k 值的增大，最小距离趋近相同，均

能保持良好的物理分散性。

5.2.4 语义差异性比较

本实验采用 θ-安全值来度量假位置集的语义差异性。θ-安全值的计算如下

$$\theta\text{-安全值}=1-\frac{|\mathrm{SEM}|}{\mathrm{C}_k^2}$$

其中，$\mathrm{SEM}=\{d_{\mathrm{sem}}\mid d_{\mathrm{sem}}(l_i,l_j)<u\}$，$k=|\mathrm{DLS}|$，DLS 是包含真实位置在内的假位置集，$\mathrm{C}_k^2$ 表示组运算式。当 θ 无限接近 1 时，则说明假位置集满足语义差异性。

MMDS 算法、MaxMinDistDS 算法、Simp-MaxMinDistDS 算法的 θ-安全值如图 7 所示。可以看出，3 种算法的 θ-安全值始终接近 1，这是因为它们在选择假位置时均考虑到了位置的语义信息，从而保证了语义差异性。

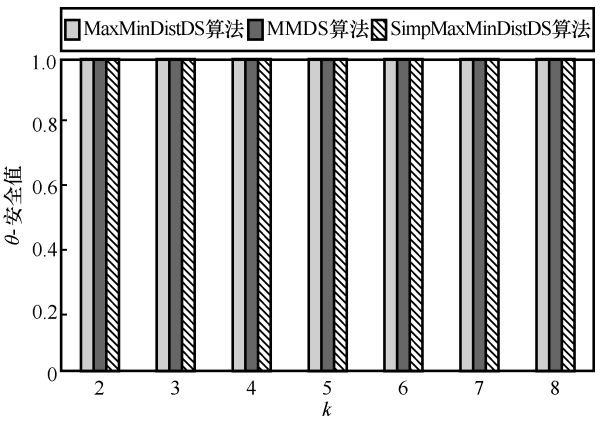


图 7　假位置集的 θ-安全值

5.2.5 位置熵比较

在位置隐私保护中位置熵主要用来衡量真实位置的不确定性，熵值越大表明匿名化程度越高，反之则匿名化程度越低。

3 种算法在不同 k 值下的位置熵如图 8 所示，可以看出，随着 k 值的增大，3 种算法的位置熵都呈整体增大趋势，但是 MMDS 算法的位置熵要明显大于另外 2 种算法，这是因为 MMDS 算法不仅考虑到位置语义差异性，还考虑到位置的访问概率，用户真实位置的不确定性更大，可以更有效地保护用户的位置隐私。

通过以上实验对比发现，MMDS 算法在尽可能满足地理位置之间分散和语义多样化的同时，还具有较高的假位置生成效率和位置熵值，能有效提高位置服务质量。

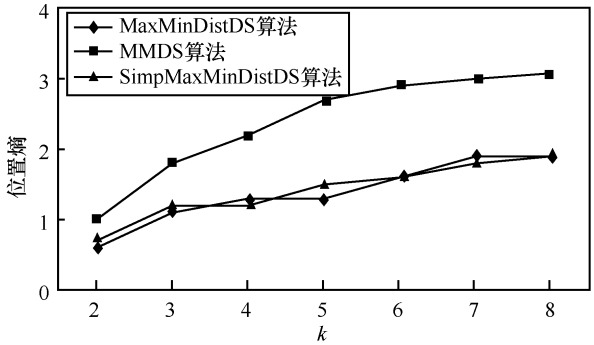


图 8　3 种算法在不同 k 值下的位置熵

6 相关工作

现有的基于位置服务的隐私保护技术主要有空间转换、匿名区域和伪造数据 3 类。文献[10]提出一种基于空间转换的位置隐私保护方法。该方法采用 Hilbert 曲线来转换用户坐标，使匿名服务器无法获得用户的位置信息，但匿名服务器可以从中得到用户的运动方向、运动速度等信息，造成用户位置信息泄露。文献[11]提出基于差分隐私的位置隐私方案，考虑由于轨迹中位置之间的相关性，独立应用噪声会泄露隐私，利用一个简单的预测函数和 2 个预算支出策略，改善独立应用噪声易造成隐私泄露的问题。文献[12]中提出基于查询范围的匿名区构造方案，考虑位置服务提供商（LSP, location service provider）的查询区域面积，将用户的查询范围引入匿名区的构造中，在保护用户隐私的同时有效降低 LSP 的查询区域面积。文献[13]提出基于用户偏好选择的假位置生成方案，根据历史查询概率选取用户发出请求较多的位置生成假位置，并考虑其速度、行驶方向，选取与用户位置相似度较高的假位置构造匿名区域。

文献[10-13]在一定程度上保护了位置隐私，但未考虑位置的语义信息，无法应对具有一定语义背景知识的攻击。文献[14]为了防止敏感语义信息泄露，充分考虑了每种语义位置类型的普及度与敏感度，提出一种满足个性化需求的位置语义保护方法。文献[15]通过 Voronoi 分割区域“有选择”的扩展来构建 θ-语义安全的隐匿区域，θ 越小，其保护程度越高。文献[16]提出一种针对路网环境下的语义位置隐私保护方法，使匿名集中用户所处语义位置类型所占比例尽可能小，从而增加用户所处语义位置的不确定性。文献[17]提出一种面向连续查询的敏感语义位置隐私保护方案，为同时抵抗连续

查询追踪攻击和语义推断攻击，构建满足-隐私模型的匿名区域。文献[14-17]均有效保护了用户的位置语义信息，但均存在匿名区域过大的问题，影响用户的服务质量，且大部分都需要第三方服务器的参与，容易引发单点故障。文献[18]提出一种基于假位置的 k-匿名位置隐私保护方法。该方法采用独立式架构，并充分考虑了位置的语义信息等特征，选取语义相似度最小的 $k-1$ 个位置点作为假位置集，保证了位置的语义安全。但因未考虑位置的查询概率信息，所以无法应对具有查询概率背景知识的攻击。

通过以上分析可知，本文所提的 MMDS 算法充分考虑了位置语义和概率信息，能够抵抗攻击者具有位置语义和查询概率信息的背景知识攻击，可以提供较强的位置隐私保护。

7 结束语

针对当前大多数基于假位置的 k-匿名位置隐私保护方案没有充分考虑攻击者拥有边信息或者背景知识等问题，本文的假位置集构造方法可以从 3 个方面达到位置隐私保护的效果。首先，保证假位置集 k 个位置之间满足语义差异性，提高用户位置语义的不可区分性，防止因语义推断造成的位置语义泄露；然后，使位置之间查询概率尽量相近，防止因过滤查询概率较低的位置，影响用户隐私需求的实现；最后，使假位置之间的地理位置尽量分散，防止因匿名区域过小造成的位置隐私泄露。实验分别从假位置集生成时间、物理分散度、语义差异性比较以及位置熵四方面将本文所提 MMDS 算法与 MaxMinDistDS 算法和 SimpMaxMinDistDS 算法进行对比，结果表明，MMDS 算法能有效提高位置熵值，且具有更好的生成效率，可以有效保护用户的位置隐私。本文方案主要考虑了快照查询的位置隐私保护，而快照查询可以看作连续查询的特殊情况，因此下一步将研究连续查询的位置隐私保护。

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, 2007, 7(1): 1-18.

[2] 倪巍伟, 马中希, 陈萧. 面向路网隐私保护连续近邻查询的安全区域构建[J].计算机学报, 2016, 39(3): 628-642.
NI W W, MA Z X, CHEN X. Safe region for privacy-preserving continuous nearest neighbor query on road networks [J]. Journal of Computer Science, 2016, 39(3): 628-642.

[3] 叶阿勇, 李亚成, 马建峰, 等. 基于服务相似性的 k-匿名位置隐私保护方法[J].通信学报, 2014, 35(11): 162-169.
YE A Y, LI Y C, MA J F, et al. Location privacy-preserving method of k-anonymous based on service similarity[J]. Journal on Communications, 2014, 35(11): 162-169.

[4] NIU B, LI Q, ZHU X, et al. Achieving k-anonymity in privacy-aware location-based services[C]//IEEE INFOCOM 2014-IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2014: 754-762.

[5] WU D, ZHANG Y, LIU Y. Dummy location selection scheme for k-anonymity in location based services[C]// Dummy Location Selection Scheme for K-Anonymity in Location Based Services. Piscataway: IEEE Press, 2017, 441-448.

[6] 李璐璐, 华佳烽, 万盛, 等. 基于高效信息缓存的位置隐私保护方案[J]. 通信学报, 2017, 38(6): 148-157.
LI L L, HUA J F, WAN S, et al. Achieving efficient location privacy protection based on cache[J]. Journal on Communications, 2017, 38(6): 148-157.

[7] 周长利, 马春光, 杨松涛. 基于敏感位置多样性的 LBS 位置隐私保护方法研究[J].通信学报, 2015, 36(4): 129-140.
ZHOU C L, MA C G, YANG S T. Research of LBS privacy preserving based on sensitive location diversity [J]. Journal on Communications, 2015, 36(4): 129-140.

[8] ZHAO W H, JING Y. Trajectory privacy protection based on location semantic perception[J]. International Journal of Cooperative Information Systems, 2019, 28(3): 1950006.

[9] CHEN S, SHEN H. Semantic-aware dummy selection for location privacy preservation[C]//2016 IEEE Trustcom/BigDataSE/ISPA. New York: IEEE Press, 2016, 752-759.

[10] PENG T, LIU Q, WANG G J. Privacy preserving for location based services using location transformation[M]. Berlin: Springer International Publishing, 2013.

[11] CHATZIKOKOLAKIS K, PALAMIDESSI C, STRONATI M. A predictive differentially-private mechanism for mobility traces[M]. Berlin: Springer International Publishing, 2014.

[12] 裴卓雄, 李兴华, 刘海, 等. LBS 隐私保护中基于查询范围的匿名区构造方案[J].通信学报, 2017, 38(9): 125-132.
PEI Z X, LI X H, LIU H, et al. Anonymizing region construction scheme based on query range in location-based service privacy protection[J]. Journal on Communications, 2017, 38(9): 125-132.

[13] 李畅, 张兴, 颜飞, 等. 基于用户偏好选择的假位置生成方案[J].计算机工程与设计, 2019, 40(4): 914-919.
LI C, ZHANG X, YAN F, et al. False position generation scheme based on user preference [J]. Computer Engineering and Design, 2019, 40(4): 914-919.

[14] 陈慧, 秦小麟. 基于位置语义的路网位置隐私保护[J].通信学报, 2016, 37(8): 67-76.
CHEN H, QIN X L. Location-semantic-based location privacy protection for road network [J]. Journal on Communications, 2016, 37(8): 67-76.

[15] LI M, QIN Z, WANG C. Sensitive semantics-aware personality cloaking on road network environment[J]. International Journal of Security and Its Applications, 2014, 8(1): 133-146.

[16] 曾海燕, 左开中, 王永录, 等. 路网环境下的语义多样性位置隐私保护方法[J].计算机工程与应用, (2019-07-25)[2020-02-25].
ZENG H Y, ZUO K Z, WANG Y L, et al. Semantic diversity location-privacy protection in road network environment[J]. Computer Engineering and Applications, (2019-07-25)[2020-02-25].

[17] 王永录, 左开中, 曾海燕, 等. 面向连续查询的敏感语义位置隐私保护方案[J].计算机工程与应用, (2019-07-25)[2020-02-25].
WANG Y L, ZUO K Z, ZENG H Y, et al. Sensitive-semantic location privacy protection for continuous query[J]. Computer Engineering and Applications, (2019-07-25)[2020-02-25].

[18] 张永兵, 张秋余, 李宗义, 等. 基于近似匹配的假位置 k-匿名位置隐私保护方法[J]. 控制与决策, 2020, 35(1): 65-73.
ZHANG Y B, ZHANG Q Y, LI Z Y, et al. Privacy protection method of pseudo-location k-anonymous location based on approximate matching[J]. Control and Decision, 2020, 35(1): 65-73.

[作者简介]

王洁（1977– ），女，山西霍州人，博士，山西师范大学副教授、硕士生导师，主要研究方向为网络信息安全、数据隐私保护。

王春茹（1990– ），女，山西洪洞人，山西师范大学硕士生，主要研究方向为数据隐私保护。

马建峰（1963– ），男，陕西西安人，博士，西安电子科技大学教授、博士生导师，主要研究方向为网络信息安全、模式识别。

李洪涛（1984– ），男，山东临沂人，博士，山西师范大学副教授、硕士生导师，主要研究方向为网络信息安全、大数据安全和隐私保护。