软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2018,29(3):703−720 [doi: 10.13328/j.cnki.jos.005451]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

路网环境下兴趣点查询的隐私保护方法*

梁慧超, 王 斌, 崔宁宁, 杨 凯, 杨晓春

(东北大学 计算机科学与工程学院,辽宁 沈阳 110169)

通讯作者: 杨晓春, E-mail: yangxc@mail.neu.edu.cn

摘 要: 近年来,无线通信技术的迅猛发展推动了基于位置服务(location-based services,简称 LBS)的发展进程.而其中,兴趣点(point of interest,简称 POI)查询是基于位置服务最重要的应用之一.针对在路网环境下,用户查询过程中位置隐私泄露的问题,提出了位置 *k* 匿名隐私保护方法.首先,匿名服务器将兴趣点作为种子节点生成网络 Voronoi 图,将整个路网划分为相互独立且不重叠的网络 Voronoi 单元(network Voronoi cell,简称 NVC).其次,利用 Hilbert 曲线遍历路网空间,并按照 Hilbert 顺序,对路网上所有的兴趣点进行排序.当用户发起查询时,提出的匿名算法通过查找与用户所在 NVC 的查询频率相同且位置分散的 *k*−1 个 NVC,并根据用户的相对位置在 NVC 内生成匿名位置,从而保证了生成的匿名集中位置之间的相互性,克服了传统 *k*-匿名不能抵御推断攻击的缺陷.理论分析和实验结果表明,所提出的隐私保护方案能够有效地保护用户位置隐私.

关键词: 位置隐私;假位置;Hilbert 曲线;网络 Voronoi 图;兴趣点

中图法分类号: TP311

中文引用格式: 梁慧超,王斌,崔宁宁,杨凯,杨晓春.路网环境下兴趣点查询的隐私保护方法.软件学报,2018,29(3):703−720. http://www.jos.org.cn/1000-9825/5451.htm

英文引用格式: Liang HC, Wang B, Cui NN, Yang K, Yang XC. Privacy preserving method for point-of-interest query on road network. Ruan Jian Xue Bao/Journal of Software, 2018,29(3):703−720 (in Chinese). http://www.jos.org.cn/1000-9825/5451.htm

Privacy Preserving Method for Point-of-Interest Query on Road Network

LIANG Hui-Chao, WANG Bin, CUI Ning-Ning, YANG Kai, YANG Xiao-Chun

(School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China)

Abstract: In recent years, the rapid development of wireless communication technology has promoted the development of location-based services (LBS), among which the point-of-interest (POI) query is one of the most important applications. A novel privacy preserving method of *k*-anonymous model is proposed to solve the problem of leaking location privacy during the query process in road network environment. First, the anonymous server uses the set of points of interest to construct the network Voronoi diagram. Then, the whole road network is divided into independent units which are called network Voronoi cell (NVC) without overlapping. Moreover, the anonymous server uses the Hilbert curve to traverse the road network space and sort the points of interest in accordance with Hilbert order. When a user requests a query, the anonymous algorithm selects dispersed *k*−1 NVCs which have the same query frequency with the NVC that user located in, and then generates dummy locations in the relative road segments corresponding to the user's in each NVC. The reciprocity of the anonymity set can be ensured and the inference attack that traditional *k*-anonymity can't resist can be avoided through

* 基金项目: 国家自然科学基金(61572122, 61532021); 辽宁省百千万人才工程 A 类项目; 中央高校基本科研业务专项资金(N161606002)

Foundation item: National Natural Science Foundation of China (61572122, 61532021); Liaoning BaiQianWan Talents Program (level A); the Fundamental Research Funds for the Central Universities (N161606002)

本文由基于图结构的大数据分析与管理技术专刊特约编辑林学民教授、杜小勇教授、李翠平教授推荐.

收稿时间: 2017-08-01; 修改时间: 2017-09-05; 采用时间: 2017-11-07; jos 在线出版时间: 2017-12-05

CNKI 网络优先出版: 2017-12-06 16:50:36, http://kns.cnki.net/kcms/detail/11.2560.TP.20171206.1650.032.html

the proposed anonymous algorithm. Finally, the theoretical analysis and experimental results show that the proposed privacy preserving scheme can effectively protect the location privacy.

Key words: location privacy; dummy location; Hilbert curve; network Voronoi diagram; point of interest

近年来,随着 GPS 通信技术的发展和移动用户的急速增长,基于位置的服务(LBS)得到了广泛应用[1].通过具有定位功能的智能移动设备,人们可以随时随地获得方便快捷的位置服务.在众多位置服务中,与兴趣点(point of interest,简称 POI)有关的查询一直是广泛使用的一项服务.兴趣点是指电子地图上的某类特殊建筑,如医院、学校、银行等,典型的兴趣点查询包括"离我最近的 K 个医院""距我 K 公里以内的银行"等.人们在享受位置服务带来便利的同时,也面临隐私泄露的风险.如何在不影响服务质量的前提下保护用户的位置隐私,成为学术界研究的热点.

近年来的 LBS 隐私保护研究大多集中在欧式空间中[2−8],路网环境下的位置 k-匿名方案大多是基于网络拓张[9−12]的方法,其思想是:从用户所在位置按照距离从近到远向外拓张,直至匿名集中的用户数等于 k 时停止.该方法存在着匿名集不满足相互性[4]的缺陷,如图 1 所示,用二元组表示边的长度和用户数,在 k=5 且路段总长度 len=18 的条件下,即构造的匿名集中超过 k 个用户、总路段长度超过 len 的隐私保护要求下,通过网络拓张构造的用户 u_1 的匿名集 $AS(u_1)=\{(p_1,p_8),(p_1,p_4)\}$,用户 u_2 的匿名集 $AS(u_2)=\{(p_1,p_4),(p_4,p_6),(p_6,p_9)\}$.已知 u_2 在路段 (p_1,p_4) 上,u_1 在路段 (p_1,p_8) 上,显然,u_2 在 u_1 的匿名集中,u_1 却不在 u_2 的匿名集中.同时,由于网络拓张方法从用户所在位置按照距离大小向外进行拓张,当路网人数众多时,在用户所在路段或用户附近为数不多的路段内便可达到 k 匿名,产生的匿名区域极小,对用户的真实位置构成极大威胁.此外,路网上的隐私保护方法并没有考虑到路网概率信息(如不同区域的历史查询频率),因此攻击者可能根据路网的分布情况、历史查询统计情况来推测用户的实际位置.

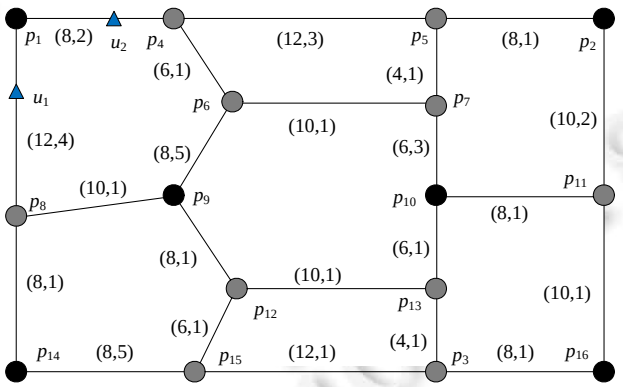


Fig.1 k-anonymous method on road network

图 1 路网上的 k-匿名方法

当前,路网环境下的位置隐私保护方法主要存在以下问题:(1) 未考虑路网中的路段历史查询概率等信息;(2) 生成较小的匿名区域;(3) 匿名集不满足相互性.为了解决上述问题,本文在路网上根据兴趣点划分网络 Voronoi 单元,利用 Hilbert 曲线填充路网空间,并根据兴趣点的 Hilbert 值对其排序,在与用户具有相同查询频率的网络 Voronoi 单元内选择与用户相对位置相同的路段生成假位置.本文生成的假位置具有与用户真实位置查询频率相同、位置分散的特点,且匿名集内各路段相互匿名,具有相互性.本文的主要贡献如下:

1) 以兴趣点为种子节点,将路网划分为独立不重叠的网络 Voronoi 单元,通过考虑每个网络 Voronoi 单元的历史查询频率,实现位置 k-匿名;
2) 利用 Hilbert 曲线对兴趣点排序,并按固定间隔选取假位置所在的 k−1 个网络 Voronoi 单元,使得生成的假位置分布广泛,避免了传统匿名方法生成较小的匿名区,能够形成较大的匿名区域,从而抵制攻击者的攻击(如单一路段攻击[13]、区域攻击[7]);

3) 通过将历史查询频率相同的网络 Voronoi 单元封装成桶,然后对每个网络 Voronoi 单元内的路段进行统一划分,从而使匿名集满足相互性;

4) 通过理论分析,证实匿名算法可以抵御重放攻击和推断攻击,通过对比实验,验证生成的假位置能够有效保护用户的真实位置.

本文第 1 节简单描述隐私保护领域的研究成果及不足之处.第 2 节介绍网络 Voronoi 图、Hilbert 空间填充曲线等本文方法需要用到的背景知识.第 3 节描述本文的攻击模型及系统框架.第 4 节详细说明隐私保护方法的算法及步骤.第 5 节对本文提出的隐私保护方法进行安全分析.第 6 节给出实验结果及分析.第 7 节总结全文,并指出未来值得关注的研究方向.

1 相关工作

在保护位置隐私方面,研究者提出了大量的研究方法.其中,Gruteser 等人[2]第 1 次提出位置 k-匿名模型,该模型指出:如果在$[t_1,t_2]$时间间隔内,出现在$[x_1,x_2]$,$[y_1,y_2]$区域中的除了移动用户的位置还包含其他 k–1 个用户的位置,那么这个区域就满足了 k 匿名,该三元组($[x_1,x_2]$,$[y_1,y_2]$,$[t_1,t_2]$)就会被发送给服务器,使得攻击者无法将用户的真实位置与其他 k–1 个位置区分开来.目前,根据隐私保护算法的应用环境,可以分为基于欧式空间的位置隐私保护方法和基于路网的位置隐私保护方法.

在欧式空间中,Kido 等人[3]提出了假位置(也称哑元)方法,即:通过匿名算法添加 k–1 个假位置,将包括用户真实位置在内的 k 个位置发送给位置服务商,以实现 k 匿名的方法.本文所提的位置隐私方案基于假位置方法.Niu 等人[3]提出了 Dummy-Location Selection(DLS)方法,该方法通过最大熵原则选择假位置,保证了 k 个位置的位置分散度;同时,该方法构造假位置时考虑到用户与假位置之间查询频率的差异性可能导致位置暴露、不能有效地实现 k-匿名的问题.图 2 中,平面区域共计有 100 个历史查询,根据每个区域的历史查询数量占所有区域历史查询数量的比例,统计出每个区域的查询频率.由于每个区域查询热度的差异,导致各个区域的查询频率的差异,用户所在区域的历史查询频率为 0.0163,添加的虚假位置的查询频率与用户所在位置相差甚远,攻击者很容易利用这种差异,将查询频率为 0 和 0.0016 的假位置排除掉.由于 DLS 方法是在 2k 个候选区域中枚举出 k 个具有最大熵的区域,因此其具有复杂度高、迭代慢、匿名集不具有相互性的缺点.

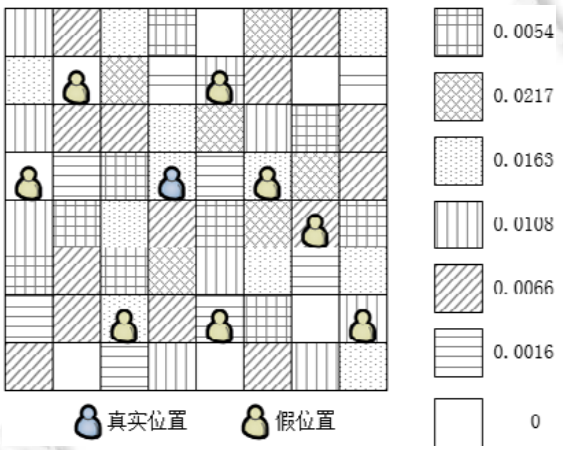


Fig.2 Selection of dummy locations in Euclidean space

图 2 欧式空间的假位置选择方法

路网环境下,基于网络拓张的思想实现 k-匿名是位置隐私保护领域的经典算法.然而路网结构复杂、数据庞大,传统的网络拓张方法需要在线对整个路网进行扫描遍历,给匿名处理和网络通信带来巨大压力.同时,该方法不能保证匿名集的相互性,攻击者很容易根据多次查询所构造的交集判断出用户的真实位置.在网络拓张的基础上,Mouratidis 等人[9]提出了基于图遍历将边上的用户进行分装的方法.其主要思想是:将路网中的点或

边进行全局编号排序,根据编号大小,把路网上的所有用户按每 k 个用户分装一个桶的方式进行分装.如图 3 所示,箭头的顺序代表边的遍历顺序,路网中有 10 个用户,假设 k=2,则会将用户分入 5 个桶中,b_1={u_1,u_2},b_2={u_3,u_4},b_3={u_5,u_6},b_4={u_7,u_8},b_5={u_9,u_{10}}.该分装方法有效地实现了相互性,即,在一个桶中的用户匿名集总是相等的,但是也存在两个问题:(1) 存在单一路径问题,即,k 个用户均在一条路段上,如在一个桶中的用户 u_1,u_2 都在边(p_{14},p_8)上,实际上就是向攻击者暴露匿名集内的用户真实所在路段就是(p_{14},p_8);(2) 在同一路段上的用户产生的匿名集不同,易遭受推断攻击[13],如在同一路段(p_5,p_4)上的用户 u_6,u_7,分属于 b_3,b_4 两个桶,其产生的匿名集分别为 $AS(u_6)$=$AS(u_5)$=$AS(b_3)$={(p_2,p_5),(p_5,p_4)},$AS(u_7)$=$AS(u_8)$=$AS(b_4)$={(p_5,p_4),(p_4,p_6)},则根据排除法可以判断出 u_5 在(p_2,p_5)上,u_8 在(p_4,p_6)上.

目前,基于假位置的方法大多只适用于欧式空间,但是路网环境更适合基于兴趣点的位置服务.因此,本文提出的隐私保护方法旨在路网环境下生成假位置,并且使添加的假位置的查询频率与用户所在区域的查询频率保持一致、假位置之间位置分散度高且匿名路段之间具有相互性.

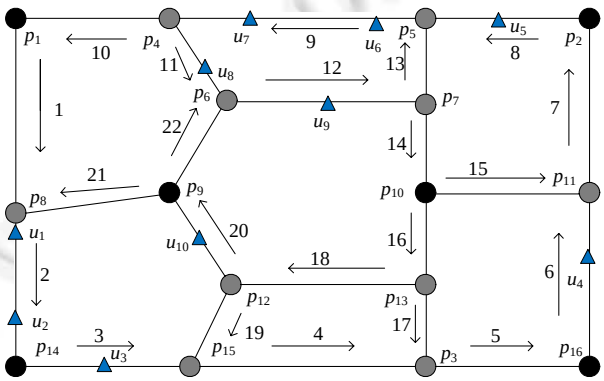


Fig.3 Location preserving method based on graph traversal

图 3 基于图遍历的位置保护方法

2 模型与框架

2.1 攻击模型

在位置隐私保护领域,攻击者可分为被动攻击者[6]和主动攻击者[14].被动攻击者主要通过窃听实体之间的通信来获取查询信息.在本文的攻击模型中,假设网络通信信道是安全的,并且实体之间的通信安全可以通过已有的加密策略实现.而主动攻击者则可以通过恶意攻击 LBS 服务器,从而获得隐私保护算法的工作机制和历史查询概率.基于 LBS 服务器拥有最多的背景知识,本文假设 LBS 服务器是主动攻击者.提出的匿名算法的安全目标是可以抵御主动攻击者的重放攻击[10]和推断攻击[7],并使得主动攻击者识别真实用户位置的概率不大于 1/k.

2.2 系统框架

在传统的基于客户端-服务器(CS)的隐私保护框架中,由客户端完成对历史查询信息的统计及对位置的匿名工作,并且需要由客户端精炼 LBS 返回的查询结果.目前,基于假位置的隐私保护方法均采用 CS 框架,大量的计算增加了客户端的计算消耗,从而降低了服务质量.本文的隐私保护框架如图 4 所示.该框架主要通过可信的第三方匿名服务器来连接用户和 LBS 服务器并实现匿名算法.与 CS 框架相比,基于第三方匿名服务器的隐私保护结构能够在很大程度上缓解客户端的计算压力.同时,由于匿名服务器执行了查询结果的精炼过程,能够快速响应查询服务请求.

用户对兴趣点的查询可以用四元组⟨uid,t,loc,$content$⟩表示,其中,每个元组分别表示用户 ID、查询时间、用户位置信息、查询内容.如图 4 所示,当用户发起查询请求时,假位置计算模块将产生 k–1 个假位置,与用户的真实位置组成位置匿名区,并将生成的位置匿名查询⟨uid,t,CR,$content$⟩发送给 LBS 服务器,其中,CR 表示位置匿名

区,由包含用户真实位置在内的 k 个位置组成.匿名服务器中的历史信息存储模块用来记录历史查询点的位置信息,该模块是为了保证匿名服务器根据匿名算法找到 k–1 个匿名路段后,不是在这些路段上随机产生假位置,而是选取该路段上的历史查询点作为匿名位置,避免了在路段中人迹稀少的区域选取假位置,从而导致具有历史位置信息的攻击者进行推断攻击.为了使生成的假位置更具真实性,规定假位置计算模块总是在 k–1 条路段上,提取最近时刻的历史查询点的位置作为假位置.LBS 服务器进行查询处理后,将结果返回给匿名服务器,匿名服务器通过查询求精模块,将用户的真实查询结果提取出来并返回给用户.

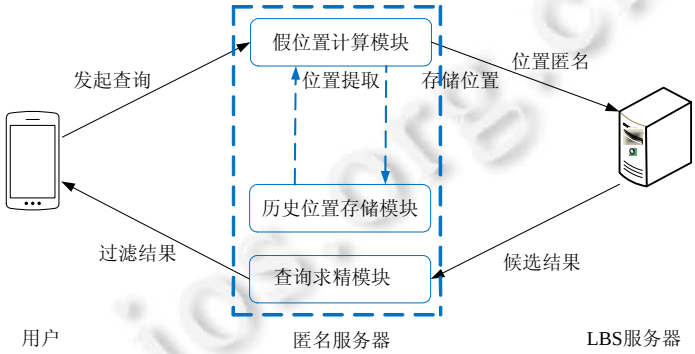


Fig.4 Overall framework of location privacy preserving

图 4 位置隐私保护框架

3 支持分散性与相互性的路网表示

通常采用无向图 $G=(V,E)$表示路网,$E=\{e_1,e_2,\ldots,e_m\}$表示路网网络拓扑中的路段,$V=\{p_1,p_2,\ldots,p_n\}$表示路段的交叉点,任意两点距离采用的是路网距离.设 V 中有 t 个顶点被用户标记为兴趣点.

本文提出的位置隐私保护方法旨在保护路网环境下兴趣点查询的用户位置.为了支持基于兴趣点查询的用户位置匿名的分散性与相互性,利用网络 Voronoi 图理论[15]对路网进行划分,以满足匿名的分散性.利用 Hilbert 空间填充曲线(Hilbert curve)[16]对兴趣点进行排序,以满足匿名的相互性.本文中使用的符号见表 1.

Table 1 Symbol definition

表 1 符号定义

名称	描述
POI	兴趣点
P	兴趣点集
$G=(V,E)$	路网图
b_i	边界点
NVC	网络 Voronoi 单元
NVD	网络 Voronoi 图

3.1 基于网络Voronoi图的路网划分

路网环境下,以兴趣点作为种子节点构造路网上的网络 Voronoi 图,用于处理相关的兴趣点查询服务.具体地,需要将路网划分为独立不重叠的子区域,并在子区域内寻找假位置.因此,首先需要将路网表示成以兴趣点为种子节点的网络 Voronoi 图,然后将路网划分为独立且互不重叠的网络 Voronoi 单元.通过该方法生成的假位置总是位于不同的子区域内且彼此距离分散,能够形成较大的匿名区域.本文的相关定义如下.

给定两个兴趣点 p_i 和 p_j,如果它们在 G 上的最短路径上不存在其他兴趣点,则在 G 中必然存在子区域 G',使得 G'中任意一点到 p_i 的路网距离都小于等于该点到 p_j 的路网距离,称 G'为 p_i 对于 p_j 的支配区域,记为 $Dom(p_i,p_j)$.

定义 1(兴趣点的网络 Voronoi 单元(NVC)). 兴趣点 p_i 的 Voronoi 单元是路网上一组路段或子路段的集合,

记做 $NVC(p_i)$.在这个集合内的路段上的任意一点到 p_i 的距离都必须小于等于该点到其他兴趣点的距离,即,$NVC(p_i)$是 p_i 相对于其他兴趣点的支配区域的并集.

定义 2(网络 Voronoi 图). 由所有兴趣点 P 的网络 Voronoi 单元共同组成的图称为该网络的网络 Voronoi 图,记为

$$NVD(P)=\{NVC(p_1),\ldots,NVC(p_t)\} \tag{1}$$

定义 3(兴趣点 p_i 与 p_j 之间的边界点). 兴趣点 p_i 与 p_j 之间的边界点是位于 p_i 与 p_j 最短路径上的某个位置点,记为 $b(p_i,p_j)$.它将该条最短路径分为两部分:一部分属于兴趣点 p_i 的 Voronoi 单元,而另一部分属于兴趣点 p_j 的 Voronoi 单元.

性质 1. 兴趣点 p_i 与 p_j 之间的边界点是这两个兴趣点之间最短路径上的中心点,即:边界点 $b(p_i,p_j)$在 p_i 与 p_j 的最短路径上,且 $b(p_i,p_j)$到 p_i 的距离与 $b(p_i,p_j)$到 p_j 的距离相等.

由于兴趣点的分布是固定的,因此,以兴趣点为种子节点的网络 Voronoi 单元就是唯一确定的,由各个独立不重叠的网络 Voronoi 单元组成的网络 Voronoi 图也是唯一确定的.

如图 5 所示,$\{p_1,...,p_{16}\}$为路网上的路段交叉点,其中,黑色节点为兴趣点 $P=\{p_1,p_2,p_9,p_{10},p_{14},p_{16}\}$,灰色节点为非兴趣点,$\{b_1,...,b_{12}\}$是兴趣点之间的边界点.根据定义 2,$NVD(P)$由 $NVC(p_1),NVC(p_2),NVC(p_9),NVC(p_{10}),NVC(p_{14}),NVC(p_{16})$组成,其中,$NVC(p_1)$是由 b_1,b_2,b_6 围成的左上角区域,$NVC(p_2)$是由 b_1,b_3,b_5 围成的右上角区域,$NVC(p_9),NVC(p_{10}),NVC(p_{14}),NVC(p_{16})$等依次排开.

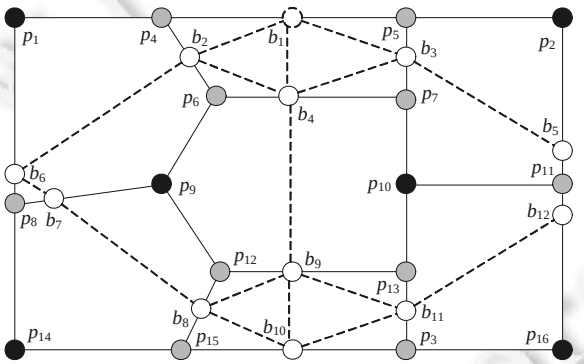


Fig.5 Network Voronoi diagram

图 5 网络 Voronoi 图

3.2 基于Hilbert空间填充曲线的兴趣点索引

定义 4(相互性[4]). 匿名集的相互性是指同一匿名集内的任意两个用户 u_i 和 u_j,其构造的匿名集分别是 AS_i 和 AS_j,如果 $AS_i=AS_j$,则匿名集满足相互性.

在假位置方法构造 k 匿名的情况下,用户 u_i 和 u_j 分别构造的匿名集分别由 k 条路段组成,匿名集具有相互性,即:当用户 u_j 所在路段出现在用户 u_i 的匿名集中时,对用户 u_j 构造的匿名集与用户 u_i 的匿名集完全相同.

大部分支持路网的 k 匿名方法不具有相互性,攻击者很容易通过多次查询构造的匿名集,推断出真正的用户位置[16].文献[17]虽然提出了支持路网并保证相互的匿名方法,但并没有实现绝对的相互性.具体地,当用户 u_i 出现在 u_j 的匿名集中时,若不能保证 u_j 也出现在 u_i 的匿名集中,则会增大 u_j,u_i 的匿名集差异性,使 u_j,u_i 的匿名集完全不同,以逃避攻击者的推断攻击.该方法需要在生成匿名集后再次检测其是否满足相互性,若不满足则需要生成完全不同的匿名集,这无疑降低了在线匿名的性能.文献[18]已经证明:利用 Hilbert 曲线遍历顺序,将匿名位置封装成桶来实现 k 匿名的方法能够满足相互性,可以抵御推断攻击.

定义 5(Hilbert 空间填充曲线(Hilbert curve)[19]). Hilbert 空间填充曲线定义为 S 维空间 R^S 与一维空间 R 之间的一一映射,记作 $H:R^S\rightarrow R$.若点 $p\in R^S$,则 $H(p)\in R$,$H(p)$称为点 p 的 Hilbert 值.对于点集$\{p_1,p_2,\ldots,p_n\}$,有:

$$H\{p_1,p_2,\ldots,p_n\}=\{H(p_1),H(p_2),\ldots,H(p_n)\}.$$

图 6 为在二维空间下进行 4×4 和 8×8 网格划分的 Hilbert 填充曲线.由 Hilbert 曲线的性质可知:在二维空间上位置相近的两个点,它们的 Hilbert 值也相近.

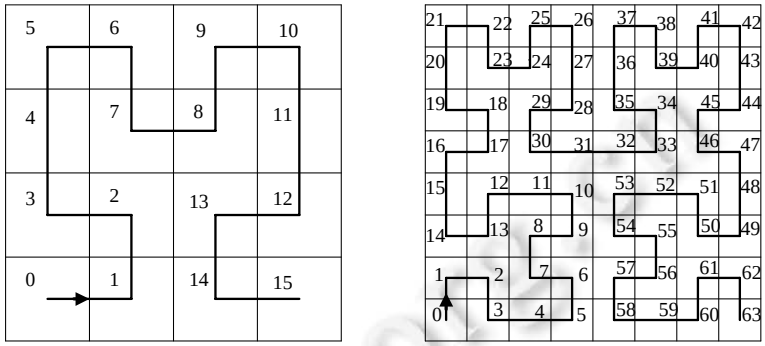


Fig.6 Hilbert curve

图 6 Hilbert 曲线

4 基于网络 Voronoi 单元查询频率的匿名方法

基于第 2.1 节中的攻击模型,本文假设攻击者拥有路网中用户在每条路段上的历史查询信息,即,攻击者可以根据不同网络 Voronoi 单元的历史查询信息推断查询用户的真实位置.本文的匿名方法使得假位置分布于不同的 NVC 中.给定两个兴趣点的 NVC:NVC(A)和 NVC(B),设查询用户的真实位置位于某个兴趣点的 NVC(A)的路段上,系统拟在某个兴趣点的 NVC(B)的路段上添加用户的虚假位置.如果系统发现在这两个 NVC 的历史用户的查询数量差距较大,系统可以根据这个差异轻易地发现添加的虚假位置,不能真正实现用户位置的 k 匿名.因此,本文的匿名算法需要统计历史用户(包括具有真实位置的用户和虚假位置的用户)在各 NVC 路段上发出查询请求的数量占历史用户在整个路网路段上发出查询请求数量的比例,即 NVC 查询频率.为了使生成的假位置更具真实性,应保证假位置所在 NVC 的查询频率与真实用户所在 NVC 的查询频率保持一致,以防攻击者根据 NVC 之间的查询频率差异推导出用户的真实位置.

4.1 网络Voronoi图的构建

为了解决基于路网的兴趣点查询的隐私保护问题,第 1 步就是要将路网构造成一个针对兴趣点集合的网络 Voronoi 图.以路网上的兴趣点作为种子节点,采用 Dijkstra 算法生成网络 Voronoi 图,算法如下.

算法 1. 网络 Voronoi 图构建算法.

输入:路网 $G(V,E)$,兴趣点集 P;

输出:网络 Voronoi 图(NVD).

1. for each $v\in V$
2. 构造标签($POI(v),d(v)$); //$P(v),d(v)$表示距离 v 最近的兴趣点及距离//
3. end for
4. foreach $(v,u)\in E$, $u\in V$ //向外拓展寻找 v 的邻接点//
5. if ($P(u)\cap P(v)=\varnothing$) //如果 u,v 属于不同的 NVC//
6. 计算边界点 b 的位置;
7. 保存边界点 $b(u,v|d(u)-d(v)-dis(u,v)/2)$; //$dis(u,v)$表示 u,v 构成的边的长度//
8. end if
9. end for
10. return NVD;

算法 1 首先对所有节点构造标签($P(v),d(v)$),其中,$P(v)$表示距离 v 最近的兴趣点,$d(v)$是 v 到 $P(v)$的最短路径

距离.对于兴趣点 $p_i \in P$,其标签为$(p_i,0)$;对于非兴趣点,通过 Dijkstra 算法在路网上拓展,拓展中遇到的第 1 个兴趣点即为最近的兴趣点,将该兴趣点与对应的最短路径距离构造标签.算法的第 3 行~第 8 行描述了寻找边界点的过程.对于每一条边$(v,u) \in E$,如果顶点 v 和顶点 u 的最近邻兴趣点不同,说明边(v,u)上必然存在边界点,记 v,u 的最近邻兴趣点为 $P(v),P(u)$,根据两个兴趣点路径之间的中点构造边界点,则边界点到 $P(v)$的距离为 $d(u)-d(v)-dis(u,v)/2$,到 $P(u)$的距离为 $d(v)-d(u)-dis(u,v)/2$,且 $d(u)-d(v)-dis(u,v)/2=d(v)-d(u)-dis(u,v)/2$.通过该方法,计算所有边界点并存储,最终实现对 NVD 的构建.根据文献[10]的分析,该算法在最坏情况下的复杂度为$O(n+(n-t)m'+(n-k)^2\log(n-t))$,其中,$n=|V|,t=|P|,m'=|V-P|$.

4.2 网络Voronoi单元的Hilbert顺序

本文提出的匿名算法需要选取距离彼此分散的 $k-1$ 个网络 Voronoi 单元并在其中添加假位置,由于兴趣点的 Hilbert 顺序即可反映其距离远近关系,进一步地,则可利用兴趣点的 Hilbert 顺序来表示网络 Voronoi 单元之间的距离远近关系,并按固定间隔依次选取 $k-1$ 个网络 Voroni 单元.

为了获取网络 Voronoi 单元的 Hilbert 顺序,本文利用 Hilbert 曲线依次遍历路网空间,并将兴趣点的 Hilbert 遍历顺序作为其所在网络 Voronoi 单元的遍历顺序.算法 2 讲述了 NVC 的 Hilbert 遍历顺序的计算过程,该算法的时间复杂度为$O(n\times\log n+t\times\log t)$.具体地,算法首先根据路网大小及兴趣点的分布情况确定划分阶数 n,将路网空间划分成 $2^n\times2^n$ 个均等的网格.随后,用 Hilbert 曲线对划分好的网格空间进行填充,按照遍历顺序,每个网格对应唯一一个 Hilbert 值.算法的第 3 行~第 6 行,对于每一个兴趣点,寻找其所在网格的 Hilbert 值;第 7 行将兴趣点的 Hilbert 值进行排序,该顺序表现了 NVC 之间的距离关系.为了避免产生较小的匿名区域而暴露用户的位置隐私,本文算法生成的假位置所在 NVC 的距离较远,从而产生较大的匿名区域.

算法 2. NVC 的 Hilbert 顺序计算算法.

输入:兴趣点集 $P=\{p_1,p_2,\ldots,p_t\}$,路网空间的划分阶数 n;

输出:NVC 的 Hilbert 遍历顺序表 HI.

1. $(2^n\times2^n)cell\leftarrow Map(G)$;
2. $AL=Hilbert((2^n\times2^n)cell)$; //将路网空间划分为 $2^n\times2^n$ 个网格,并进行 Hilbert 曲线填充//
3. foreach $(p_i \in P)$
4. $H(p_i)=AL(p_i.cell)$; //找到兴趣点所对应的 Hilbert 值//
5. $HI.add(H(p_i))$;
6. end for
7. $HI.sort()$; //对 NVC 的 Hilbert 值排序//
8. return HI;

如图 5 所示,路网上具有 6 个兴趣点 $p_1,p_2,p_9,p_{10},p_{14},p_{16}$.根据算法 2,将空间划分成 4×4 的网格,并利用 Hilbert 曲线依次遍历这些网格.如图 7 所示,每个兴趣点所在的网格对应的 Hilbert 值即为该点 NVC 的 Hilbert 值.例如 $H(NVC(p_1))=5$.

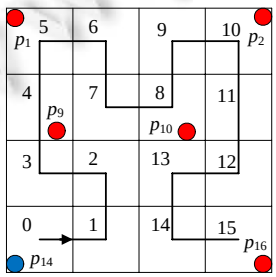


Fig.7 Hilbert value of each POI

图 7 兴趣点的 Hilbert 值

4.3 满足相互性的路段划分算法

基于概率信息的位置隐私保护方法需要统计某一区域的查询频率信息,并通过在与用户所在区域的查询频率相同的区域中添加假位置来实现位置匿名.本文提出的隐私保护方法是针对路网环境下的兴趣点查询,故需要在与用户所在 NVC 的查询频率相同的其他 NVC 中添加假位置.NVC 查询频率的定义如下.

定义 6(NVC 查询频率 P_{vi}). 设 Q_i 代表第 i 个 NVC 中的历史查询数量,网络 Voronoi 图有 n 个 NVC,则有:

$$P_{vi}=\frac{Q_i}{\sum_{i=1}^{n}Q_i} \tag{2}$$

为了保证匿名集的相互性,需要在 k–1 个与用户所在 NVC 查询频率相同的 NVC 中添加假位置,并且添加了假位置的路段在 NVC 中的排序需要与用户位置所在路段在 NVC 中的排序相同.但是,不同的 NVC 中包含的边数的差异无法满足相互匿名的要求,如用户的位置位于其所在 NVC 内的第 10 条边,为了保证匿名集的相互性,需要在其他 k–1 个 NVC 内第 10 条边上添加假位置,但当其他 NVC 内包含的边数小于 10 时,则无法找到与用户相对编号一样的边.为了解决上述问题,需要统计所有 NVC 中包含的最大边数 $L_{\max}$,并将边数小于 $L_{\max}$ 的 NVC 进行路段划分,使得每个 NVC 均拥有 $L_{\max}$ 个路段编号.

定义 7(最大边编号 $L_{\max}$). 设各个 NVC 中包含的边的数量为 L_i,i=1,2,…,t,则有:

$$L_{\max}=\mathrm{Max}(L_i),i=1,2,\ldots,t \tag{3}$$

以 $L_{\max}$ 为上界,将包含边数少于 $L_{\max}$ 的 NVC 进行边的划分,将每条边均匀划分为多个路段编号,使得划分出的路段数等于 $L_{\max}$.在考虑查询均匀分布的条件下,我们将各个边进行均匀划分,使之成为多条路段,划分后的每个路段的查询频率为

$$P_{ij}=\frac{P_{vi}}{L_{\max}},\ j=1,2,\ldots,L_{\max} \tag{4}$$

当 NVC 内的边数少于 $L_{\max}$ 时,就需要对 NVC 内的边进行再划分,算法 3 描述了路段划分算法.算法 3 为确定性算法,保证划分后的每个 NVC 包含的路段数一定为 $L_{\max}$,算法的时间复杂度为 $O(t\times L_{\max})$.

算法 3. 路段划分算法.

输入:网络 Voronoi 图(NVD);

输出:路段划分表 T.

```
1.  for each (NVC(i)∈NVD&&L_i<L_max)
2.      foreach (e_j∈NVC(i))                    //对于第 i 个 NVC 中的每一条边//
3.          if (L_max%L_i≠0 && j<(L_max%L_i))
4.              T←(⌈L_max/L_i⌉)e_j
5.          else T←(⌈L_max/L_i⌉)e_j             //对边 e_j 进行均匀划分//
6.          end if
7.      end for
8.  end for
9.  return T;
```

由于兴趣点及网络拓扑都是固定不变的,对路网进行网络 Voronoi 图的划分、对兴趣点进行 Hilbert 值排序、对各个 NVC 内路段的划分和编号都是可以离线建立的,从而缩短了匿名时间和服务响应时间.

以图 8 中的网络 Voronoi 图为例,在每个 NVC 内,根据路网中边的顺序(边的排序对算法没有影响,可以为任意顺序,本文采用路网数据集中默认的边的顺序)对 NVC 内的边进行编号.如图 8 所示,$NVC(p_1)$,$NVC(p_2)$,$NVC(p_9)$,$NVC(p_{10})$,$NVC(p_{14})$,$NVC(p_{16})$包含的边数分别为 4,5,7,9,6,4,即,$NVC(p_{10})$拥有最大边数且 $L_{\max}$=9,则需要对其他 5 个 NVC 中的边进行划分.由算法 3 可以知道,每个 $NVC(p_i)$需要划分的边数为 $L_{\max}$–L_i.在每个 $NVC(p_i)$内,从编号 1 开始划分.以 $NVC(p_1)$为例,$NVC(p_1)$需要增加 $L_{\max}$–L_1=5 条边,从第 1 条边开始进行划分,则需要将第

1 条边平均拆分为 3 条路段,将第 2 条~第 4 条边平均拆分为 2 条路段,新的路段编号紧跟边的顺序,依次递增,以此类推,将所有 NVC 都划分为 9 条路段.

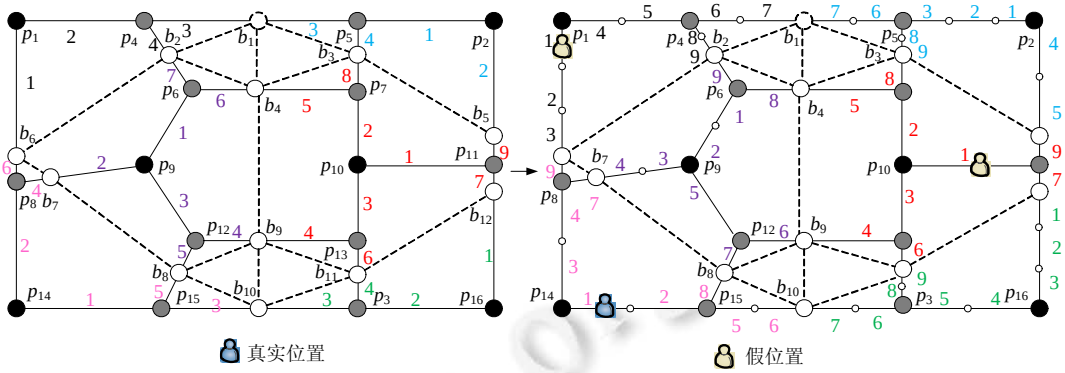


Fig.8 Road segment in each NVC

图 8 NVC 内路段划分

当路网中所有 NVC 中的路段数都与 L_{max} 相等时,用户处于任意 NVC 中的任意路段时,在其他 NVC 中总能找到与之编号相同的路段,这为匿名集满足相互性提供了前提.

4.4 生成假位置

当用户发起查询时,匿名服务器根据匿名算法生成相应的假位置,该算法根据用户的真实位置及隐私保护度 k 值的大小在线计算.算法 4 描述了匿名服务器生成假位置的过程,该算法的时间复杂度为 $O(t\times\log t)$.根据分析可知:当路网中兴趣点固定不变时,该算法的时间复杂度与隐私保护度 k 成反比关系,实验部分验证了算法 4 的平均执行时间与隐私保护度 k 的关系.

算法 4. 假位置生成算法.

输入:网络 Voronoi 图(NVD),兴趣点的希尔伯特值索引 *HI*,路段划分表 *T*,历史查询位置表 *HP*,隐私保护度 k,用户所在 NVC 的查询频率 P_{vi};

输出:假位置集合 *Pos*.

1. $count\leftarrow 0$;
2. for each ($NVC(j)\in NVD$ && $|P_{vj}-P_{vi}|<\delta$)
3. $count$++; //统计频率相同的 NVC 个数//
4. $bucket_1\leftarrow HI(NVC(j))$; //频率相同的 NVC 放入桶 $bucket_1$//
5. end for
6. $bucket_1.sort()$; //对兴趣点的 Hilbert 值排序//
7. for ($k'=0$; $k'<\lceil count/(\lfloor count/k\rfloor)\rceil$; k'++)
8. $bucket_2(k')\leftarrow(\lfloor count/k\rfloor)bucket_1$ //二级分桶,每$\lfloor count/k\rfloor$个 NVC 放入一个桶//
9. end for
10. for ($i=0$; $i<k'$; i++)
11. $Pos\leftarrow HP(bucket(i).rankR.rankL)$; //*rankR* 为用户所在 NVC 在桶内的相对位置,*rankL* 为用户所在路段在 NVC 内的编号//
12. end for
13. return *Pos*;

算法 4 描述了生成假位置的过程.首先,查找所有与用户所在 NVC 查询频率相同的 NVC 并统计其个数 *count*.在这些频率相同的 NVC 中,根据其兴趣点的 Hilbert 值大小进行排序.算法的第 7 行~第 9 行,将这些 NVC 平均划分为 k'个桶,其中 $k'=\lceil count/(\lfloor count/k\rfloor)\rceil$,根据分析,$k'$为[$k$,$k$+1]之间的整数,即 $k'=\{Z|k\leqslant k'\leqslant k+1\}$.记用户所

在NVC在桶内的相对位置为*rankR*,用户所在路段在其NVC内的编号记为*rankL*,在其他(k'−1)个桶中寻找桶内相对位置为*rankR*的NVC,并在这些NVC中寻找编号为*rankL*的路段作为假位置生成路段.需要注意的是:假位置在路段中并不是随机产生的,而是将该路段的最新时刻的历史查询点位置作为假位置,这样使得添加的假位置满足位置的真实性.

值得注意的是:在兴趣点比较多时,多个兴趣点可能位于同一个网格内.在这种情况下,有两种解决方案.

(1) 调整划分阶数 n 的大小,将整个路网空间划分为更细粒度的网格,使得一个网格内最多存在一个兴趣点;

(2) 使用文献[6]提出的修正的Hilbert曲线方法,将存在多个兴趣点的网格划分为4个子网格,如图9所示.

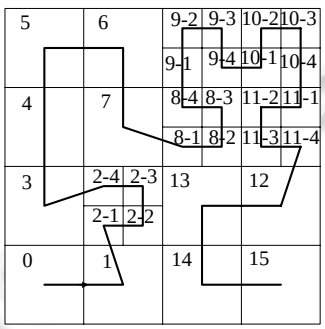


Fig.9 Modified Hilbert curve

图9 修正的Hilbert曲线

图7中假设与用户所在的$NVC(p_{14})$具有相同查询频率的NVC是$NVC(p_1)$,$NVC(p_9)$,$NVC(p_{10})$,$NVC(p_2)$,$NVC(p_{16})$,即与用户查询频率相同的NVC个数$count$=6(包含用户所在NVC),将这6个NVC根据其兴趣点的Hilbert值从小到大排序为:$NVC(p_{14})$,$NVC(p_9)$,$NVC(p_1)$,$NVC(p_2)$,$NVC(p_{10})$,$NVC(p_{16})$,以隐私保护度k=3的情况为例,将每$\lfloor count/k \rfloor$=2个NVC为一个桶,即$NVC(p_{14})$,$NVC(p_9)$为一个桶,$NVC(p_1)$,$NVC(p_2)$为一个桶,$NVC(p_{10})$,$NVC(p_{16})$为一个桶,已知$NVC(p_1)$在桶内的相对位置为1,因此选取相对位置也为1的$NVC(p_1)$,$NVC(p_{10})$作为添加假位置的NVC,之后,根据用户真实位置所在路段在其NVC中的编号,在$NVC(p_1)$,$NVC(p_{10})$的相同编号的路段添加假位置,假设用户在相对位置为1的路段,则也在$NVC(p_1)$,$NVC(p_9)$中编号为1的路段,将历史位置表HP中该路段的最近时刻的历史查询点位置作为最终的假位置点.

值得注意的是:当有用户在该匿名集内的其他任意路段内进行位置匿名时,其产生的匿名集仍然是这k条路段,即,满足匿名集的相互性.以图7为例,当k=3时,$NVC(p_1)$,$NVC(p_{10})$,$NVC(p_{14})$的编号为1的路段组成一个匿名集,则当有用户在$NVC(p_1)$的编号为1的路段发起查询请求时,匿名算法仍然会在$NVC(p_{10})$,$NVC(p_{14})$的编号为1的路段中添加假位置;当有用户在$NVC(p_{10})$的编号为1的路段发起查询请求时,算法4也会在$NVC(p_{21})$,$NVC(p_{14})$的编号为1的路段中添加假位置.

5 安全分析

5.1 抵御重放攻击

路网上的重放攻击是指攻击者具备匿名算法的原理和生成的匿名集等背景知识,攻击者利用这些背景知识对匿名集中的每个路段进行位置匿名,并与匿名集进行比较,从而判断出用户真正位置的攻击方式.

定理1. 提出的匿名算法可以抵御重放攻击.

证明:为了证明所提出的匿名算法可以抵御重放攻击,需要说明无论在哪一个路段上进行多少次查询,对其位置的匿名总是选择不变的另外k–1个路段,即,产生的匿名集满足相互性.假设本文提出的匿名算法为A(*),匿名服务器利用匿名算法A(*)为用户u的真实位置产生的路段匿名集为$S=\{l_1,l_2,\ldots,l_k\}$,则对S中的每一路段l_i(i=1,2,…,k)运行算法A(*),生成新的匿名集合S',记用户所在NVC为$NVC(v_1)$,$NVC(v_1)$的查询频率为P_{v_1},在由所

有查询频率都为 P_{v_1} 的 NVC 划分而成的 k 个桶中,$NVC(v_1)$在桶内的位置记为 $rankR$,用户所在路段在 $NVC(v_1)$内的相对位置为 $rankL$.根据匿名原则,总是在频率相同、桶内位置相同的 k–1 个 NVC 内寻找相对位置相同的路段添加假位置.因此,在本文的匿名方法中,在匿名集内任意路段上的用户,其所在 NVC 的查询频率都为 P_{v_1},所在 NVC 在桶内的位置都为 $rankR$,所在路段在 NVC 内的相对位置都为 $rankL$,因此,匿名集内的路段添加的假位置的路段也总是相互的,即 $S'=S=\{l_1,l_2,\ldots,l_k\}$.因此,无论在哪一个路段上进行多少次查询,对其位置的匿名总是选择不变的另外 k–1 个路段,即产生的匿名集满足相互性.因此,我们提出的方案能够抵御重放攻击. □

5.2 抵御推断攻击

推断攻击指攻击者通过攻陷 LBS 服务器从而获得相关背景知识(如查询频率、历史查询位置、匿名集等),通过这些知识,攻击者可以在匿名集中识别用户的真正位置.

定理 2. 提出的匿名算法可以抵御推断攻击.

证明:为了证明所提出的匿名算法可以抵御推断攻击,需要说明在攻击者拥有相关的路网背景知识,即路网中每个 NVC 的历史查询频率信息的前提下,本文提出的匿名算法仍然保证攻击者不能将用户的真实位置从 k 个位置中区分出来.假设用户所在 NVC 为 $NVC(v_1)$,用户所在 NVC 的查询频率为 P_{v_1},攻击者根据匿名算法可以获取到用户的匿名 NVC 集合为 $N=\{NVC(v_1),NVC(v_2),NVC(v_3),\ldots,NVC(v_n)\}$;同时,攻击者通过与 LBS 服务器相互勾结,可以获取路网中不同 NVC 的历史查询频率信息,则对于匿名算法寻找的 k–1 个匿名 NVC,$NVC(v_2)$,$NVC(v_3)$,…,$NVC(v_n)$,其 NVC 的查询频率分别记为 $P_{v_2},P_{v_3},...,P_{v_n}$.根据匿名原则,算法总是在与用户所在 NVC 查询频率相同的其他 k–1 个 NVC 中添加假位置,即 $P_{v_2}=P_{v_3}=...=P_{v_n}=P_{v_1}$.因此,即使攻击者具有每个 NVC 的历史查询频率信息,由于匿名集中产生的 k–1 个匿名位置与真实用户所在位置具有相同的查询频率,攻击者仍然无法通过统计信息推断和过滤掉匿名位置,即,攻击者推断出真实位置的概率不大于 1/k. □

6 实验分析

在本节中,我们根据公式(2)依次统计了每个 NVC 中的历史查询频率,并在真实路网数据集上测试本文提出的匿名方法.

6.1 度量标准

在实验中,我们利用方差、熵[20]和平均路径距离度量匿名位置的不确定性.具体地,方差值体现了 k 个位置之间的查询频率的差异,方差越小,假位置的真实性越高.熵主要用于衡量信息的不确定性,熵值越大,代表信息的不确定性越高,即,攻击者从 k 个位置中分辨出真实位置的可能性越小.平均路径距离指 k–1 个假位置到真实位置之间的平均路网距离,平均路径距离越大,代表假位置越分散,距离真实位置越远.方差 S^2、熵 H 和平均路径距离 APD 的定义如下:

$$S^2=\frac{\sum_{i=1}^{k-1}(P_{ij}-P_{uj})^2}{k-1} \tag{5}$$

$$H=-\sum_{i=1}^{k}P_{ij}\log P_{ij} \tag{6}$$

$$APD=\frac{\sum_{i=1}^{k-1}dis(loc_{dummy},loc_{user})}{k-1} \tag{7}$$

其中,P_{uj} 代表用户所在路段的查询频率,P_{ij} 代表第 i 个假位置所在路段的查询频率,$dis(loc_{dummy},loc_{user})$表示假位置与用户之间的最短路径距离.

6.2 实验环境与数据集

实验在 ubuntu 15.04 上使用 C++实现,采用 g++编译.实验运行在 Intel Core i7-6700 的 PC 上,内存和 CPU

分别为 8GB 和 3.40 GHz,硬盘大小为 1TB.基于第 3.2 节中设定网络通信是安全可信的,实验部分不考虑网络通信等带来的能耗.本文采用德国 Oldenburg 路网数据集和新加坡 Singapore 城市路网数据集来分别测试算法在稀疏路网环境下和稠密路网环境下的表现,其中,Oldenburg 数据集包含 6 105 个节点、7 035 条边,Singapore 数据集包含 20 801 个节点、55 892 条边,数据集的参数见表 2.实验从这两个数据集中随机选择 1 000 个节点作为兴趣点,默认采取 200 000 的查询频数来模拟不同用户发起的 200 000 次查询请求.为了进一步加快匿名服务的响应速度,匿名服务器离线建立兴趣点的网络 Voronoi 单元及其 Hilbert 值索引.

Table 2 Parameters of the experimental data set

表 2 实验数据集参数

Parameter	Oldenburg	Singapore
Number of nodes	6 105	20 801
Number of edges	7 035	55 892
Length of the area/m	26 915	67 544
Width of the area/m	23 572	47 774

6.3 实验结果

本节通过 11 组实验,分析和验证算法的可行性和高效性.实验中,首先测试了随着 k 值的变化,相应方差的变化情况.现有的路网环境下的隐私保护方法可分为基于网络拓张的匿名技术[9]、基于 X-star 的匿名技术[11,18]和基于 Mix Zone[21−23]的匿名技术,由于这些方法形成的匿名区域较小,因此其他 k–1 个匿名位置与用户距离也极小,很容易遭受区域攻击[7].文献[18]为 X-Star 的变种,图 10 中,DepthFirst 方法是在此基础上提出的一种针对路网环境下的位置隐私保护方法,其思想是:首先使用深度遍历方法将边进行排序,随后将路网中的边均等放入 k 个桶中,根据用户所在边在桶中的相对位置,在其他 k–1 个桶中挑选与用户在桶内相对位置相同的边,将这 k–1 条边上的最近一次历史查询位置作为添加的假位置.相较于传统的匿名技术,该方法产生的匿名位置能够拥有较大的平均路径距离.BreadthFirst[21]与 DepthFirst 相似,边的排序编号采用宽度遍历方式.Random 为路网上随机游走选择假位置的方法[5],该方法为基于假位置的位置隐私保护领域的基本算法.HilbertCurve 为本文提出的位置隐私保护方法,由于 HilbertCurve 方法考虑了真假位置间的查询频率的相似性,随着 k 值的增大,其方差总是远远小于 Random 方法.

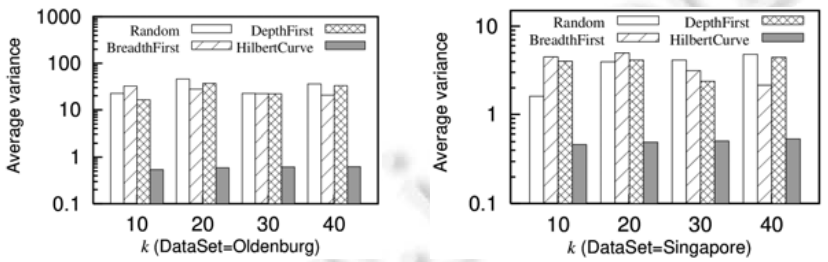


Fig.10 Effect of k on variance

图 10 k 值对方差的影响

熵值主要用于体现匿名集中各个位置间所包含信息的差异性和不确定性.由图 11 可以看出:熵值随着 k 的增大而增大,本文提出的位置保护方法的熵值略大于其他 3 种方法.即:在攻击者具有频率背景知识的前提下,本文提出的方法更能抵御攻击者的攻击.

图 12 反映了随着 k 的变化,平均路径距离的变化情况.实验结果表明:无论 k 值如何变化,本文提出的位置隐私保护方法比其他 3 种方法的平均路径距离都大,假位置分布更分散.

平均匿名时间指匿名服务器平均对一个用户查询进行位置匿名的时间,在本文中,算法 4 的平均执行时间即平均匿名时间.图 13 反映了在添加 200 000 次~600 000 次查询请求时,平均匿名时间的变化情况.随着查询数量的增大,平均匿名时间趋于下降,表现了良好的可拓展性.值得注意的是:随着 k 值的增大,平均匿名时间也趋

于下降.这是由于当与用户所在 NVC 查询频率相同的 NVC 数量 *count* 不变时,根据算法 4,匿名算法需要将这些 NVC 每隔⌊*count*/*k*⌋个装入一个桶中.随着 *k* 值增大,每个桶包含的 NVC 数量将趋于下降,分桶时间降低.

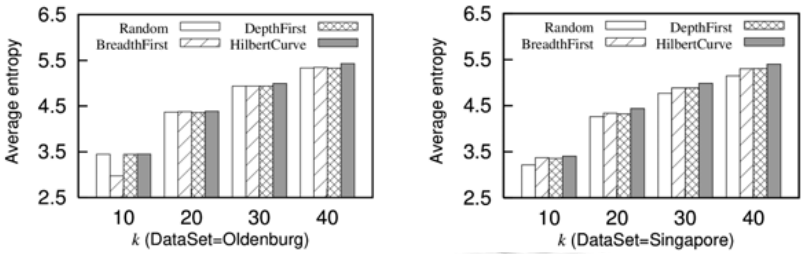


Fig.11 Effect of *k* on entropy

图 11 *k* 值对熵的影响

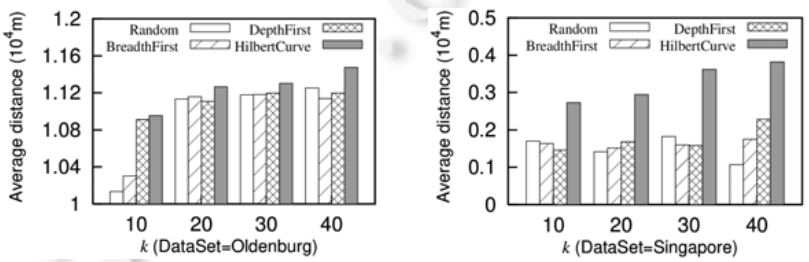


Fig.12 Effect of *k* on average path distance

图 12 *k* 值对平均路径距离的影响

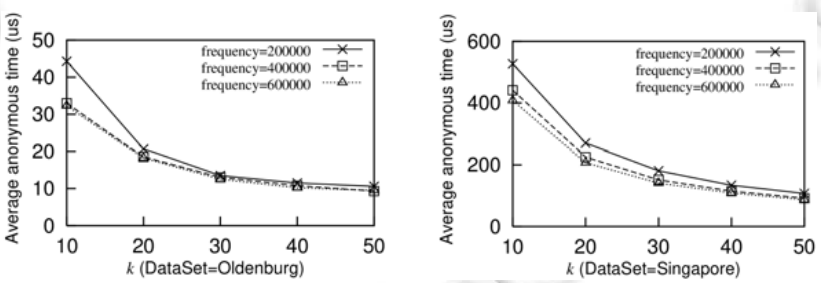


Fig.13 Effect of *k* on average anonymous time

图 13 *k* 值对平均匿名时间的影响

平均查询时间[24]指用户从发起查询到获得查询结果的平均时间.图 14 反映了在进行 200 000 次~600 000 次 NN 查询[25]时,平均查询时间随 *k* 值的变化情况.如图 14 所示:平均查询时间的变化与平均匿名时间相似,均随着 *k* 值及查询数量的增大而变小.

K-NN 查询[26]旨在查找距离用户最近的 *K* 个兴趣点,图 15 反映了在查询频数为 200 000 次~600 000 次的条件下,当用户查询距离其最近的 *K* 个兴趣点时,平均查询时间随 *K* 的变化情况.当 *K* 增大时,由于搜索路径变长,不同查询频数下的平均查询时间都会趋于增加;当 *K* 取值不变时,平均查询时间随着查询频数的增加而趋于下降,与 NN 查询的变化趋势一致.

当路网数据庞大、道路繁多时,兴趣点的数量也会大幅增加,对匿名算法的性能要求也会随之提高.图 16 反映了 POI 数目对构建 NVC(算法 1)时间的影响.随着 POI 数目的增多,构建 NVC 的时间反而大幅降低.造成这种现象的原因是随着 POI 数量的增多,顶点到最近 POI 的路径距离变小,搜索时间变短.

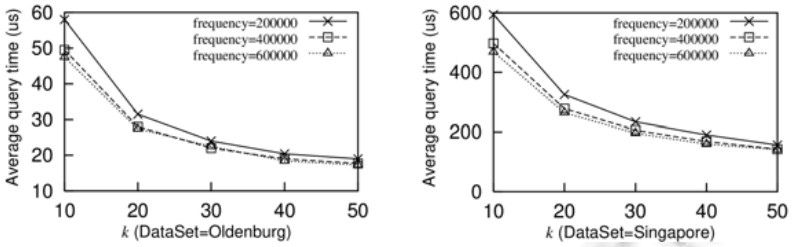


Fig.14 Effect of k on average query time

图 14 k 值对平均查询时间的影响

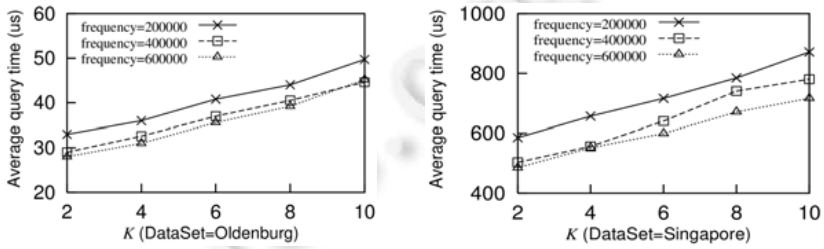


Fig.15 Effect of the number of POIs that users query K on average query time

图 15 用户查询 POI 数 K 对平均查询时间的影响

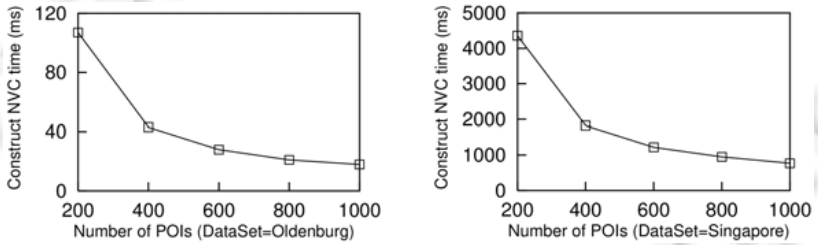


Fig.16 Effect of the number of POIs on the time of constructing each NVC

图 16 兴趣点数量对 NVC 构建时间的影响

图 17 反映了 POI 数目对计算 NVC 的 Hilbert 顺序(算法 2)时间的影响.随着 POI 数目的增加,需要计算的 NVC 相应增加,因此,NVC 的 Hilbert 值的计算时间也呈上升趋势.

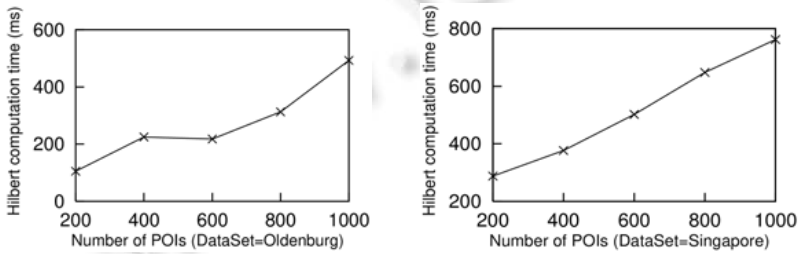


Fig.17 Effect of the number of POIs on the time of constructing the Hilbert value of calculating each NVC

图 17 兴趣点数量对 NVC 的 Hilbert 值的计算时间的影响

图 18 反映了 POI 数目对路段划分(算法 3)时间的影响.随着 POI 数目的增多,路段划分的时间趋于下降,这是因为当兴趣点增多时,每个 NVC 内包含的路段数变少,导致路段划分的基数变小,划分时间变短.

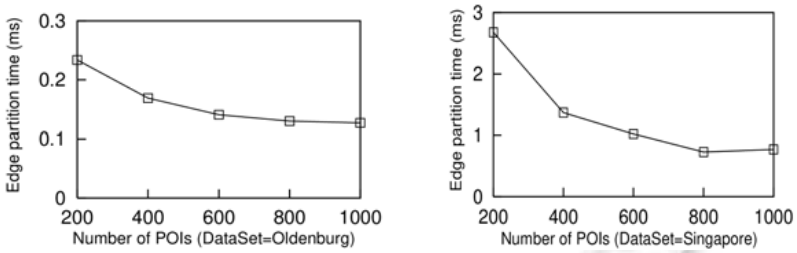


Fig.18 Effect of the number of POIs on the time of edge partition time

图 18 兴趣点数量对路段划分时间的影响

当路网中的 POI 数量增加时,由于 NVC 个数的增加,与用户查询频率相同的 NVC 个数也增加,即:匿名的候选 NVC 个数增加,则导致平均匿名时间趋于增加,如图 19 所示.但值得注意的是:与图 13、图 14 相似,图 20 中随着 POI 数量的增加,平均查询时间的上升趋势与平均匿名时间相似,说明平均查询时间的增加主要源于平均匿名时间,即,POI 数量对查询本身并不产生较大影响.

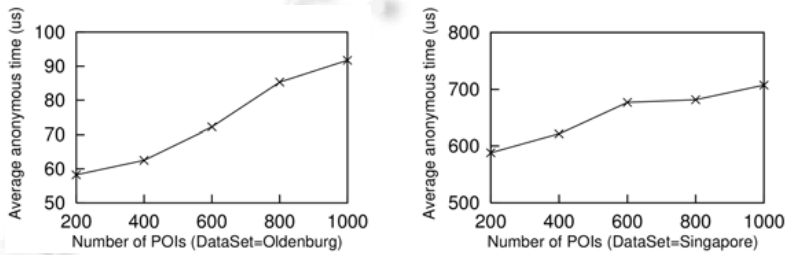


Fig.19 Effect of the number of POIs on the average anonymous time

图 19 兴趣点数量对平均匿名时间的影响

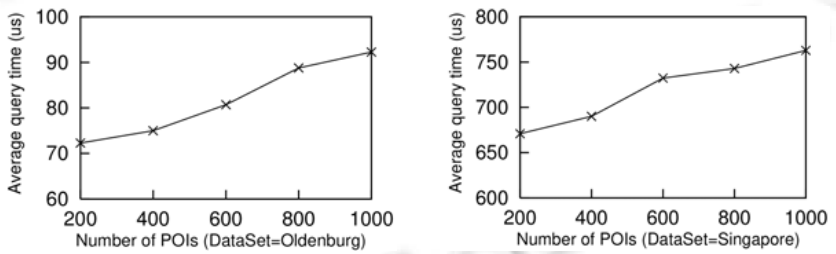


Fig.20 Effect of the number of POIs on the average query time

图 20 兴趣点数量对平均查询时间的影响

由图 16~图 20 可知:在路网复杂、POI 分布密集的情况下,本文匿名算法也能表现出良好的拓展性.根据算法在两个数据集上的表现可以得出结论:在稀疏和稠密的路网环境下,本文提出的方法都能表现出良好的性能.

7 结束语

本文提出了一种针对路网环境下兴趣点查询的隐私保护方法,该方法充分考虑到路网的历史查询频率以及匿名路段之间的相互性问题,使得生成的假位置频率相近、位置分散,能够抵御重放攻击和推断攻击.本文提出的假位置生成算法在熵值、方差、平均路径距离上都具有显著优势,因此,本文提出的位置匿名算法可以有效保护用户的位置隐私.实验结果表明:本文方法具有良好的扩展性,能够在不影响服务质量的前提下保护用户的位置隐私,对路网环境下进行兴趣点查询的位置隐私保护领域具有广泛而深远的现实意义.

当然,路网环境下的位置隐私保护领域还有许多其他问题有待深入探索和解决.在未来的工作中,我们将向两个方向进行拓展:(1) 进一步提高算法的性能,使其能够提供更高效便捷的隐私保护服务;(2) 研究更优的隐

私保护方法,将其应用到更广泛的位置服务领域,不局限于路网上的兴趣点查询范围.

References:

[1] Wang B, Yang X, Wang G, Yu G, Zang W, Yu M. Energy efficient approximate self-adaptive data collection in wireless sensor networks. Frontiers of Computer Science, 2006,10(5):936−950. [doi: 10.1007/s11704-016-4525-7]

[2] Gruteser M, Grunwal D. Anonymous usage of location-based services through spatial and temporal cloaking. In: Proc. of the Int'l Conf. on Mobile Systems, Applications, and Services (MobiSys). 2003. 163−168. [doi: 10.1145/1066116.1189037]

[3] Ido H, Yanagisawa Y, Satoh T. An anonymous communication technique using dummies for location-based services. In: Proc. of the Int'l Conf. on Pervasive Services. 2005. 88−97. [doi: 10.1109/PERSER.2005.1506394]

[4] Kalnis P, Ghinita G, Mouratidis K, Papadias D. Preventing location-based identity inference in anonymous spatial queries. IEEE Trans. on Knowledge and Data Engineering, 2007,19(12):1719−1733. [doi: 10.1109/TKDE.2007.190662]

[5] Niu B, Li Q, Zhu X, Cao G, Li H. Achieving *k*-anonymity in privacy-aware location-based services. In: Proc. of the 33rd Annual IEEE Int'l Conf. on Computer Communications. 2014. 754−762. [doi: 10.1109/INFOCOM.2014.6848002]

[6] Niu B, Li Q, Zhu X, Li H. A fine-grained spatial cloaking scheme for privacy-aware users in location-based services. In: Proc. of the Int'l Conf. on Computer Communication and Networks. 2014. 1−8. [doi: 10.1109/ICCCN.2014.6911813]

[7] Cui N, Yang X, Wang B. A novel spatial cloaking scheme using hierarchical Hilbert curve for location-based services. In: Proc. of the 17th Int'l Conf. on Web-Age Information Management. 2016. 15−27. [doi: 10.1007/978-3-319-39958-4_2]

[8] Lee HJ, Hong ST, Yoon M, Um JH, Chang JW. A new cloaking algorithm using Hilbert curves for privacy protection. In: Proc. of the ACM Sigspatial Int'l Workshop on Security and Privacy in Gis and Lbs. 2010. 42−46. [doi: 10.1145/1868470.1868480]

[9] Mouratidis K, Yiu ML. Anonymous query processing in road networks. IEEE Trans. on Knowledge and Data Engineering, 2010, 22(1):2−15. [doi: 10.1109/TKDE.2009.48]

[10] Papadias D, Zhang J, Mamoulis N, Tao Y. Query processing in spatial network databases. In: Proc. of the 29th Int'l Conf. on Very Large Data Bases. 2003. 802−813. [doi: 10.1016/B978-012722442-8/50076-8]

[11] Wang T, Liu L. Privacy-Aware mobile services over road networks. In: Proc. of the 35th Int'l Conf. on Very Large Data Bases. 2009. 1042−1053. [doi: 10.14778/1687627.1687745]

[12] Chow CY, Mokbel MF, Bao J, Xuan J. Query-Aware location anonymization for road networks. GeoInformatica, 2011,15(3): 571−607. [doi: 10.1007/s10707-010-0117-0]

[13] Li M, Qin Z. Survey of location privacy protection over road networks. Application Research of Computers, 2014,31(9):2576−2580 (in Chinese with English abstract). [doi: 10.3969/j.issn.1001-3695.2014.09.003]

[14] Zhu H, Wang J, Wang B, Yang X. Location privacy preserving obstructed nearest neighbor queries. Journal of Computer Research and Development, 2014,51(1):115−125 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2014.20130694]

[15] Erwig M, Hagen F. The graph Voronoi diagram with applications. Journal of Network, 2000,36(3):156−163. [doi: 10.1002/1097-0037(200010)36:3<156::AID-NET2>3.0.CO;2-L]

[16] Chen XH, Pang J. Measuring query privacy in location-based services. In: Proc. of the 2nd ACM Conf. on Data and Application Security and Privacy. 2012. 49−60. [doi: 10.1145/2133601.2133608]

[17] Zheng M, Wang B, Yang XC. Privacy preservation approach for location-based service on road network. Journal of East China Normal University (Natural Science), 2015,183(5):116−127 (in Chinese with English abstract). [doi: 10.3969/j.issn.1000-5641.2015.05.010]

[18] Kim YK, Hossian A, Hossian AA, Chang JW. Hilbert-Order based spatial cloaking algorithm in road network. Concurrency and Computation: Practice and Experience, 2013,25(1):143−158. [doi: 10.1002/cpe.2844]

[19] Mokbel MF, Aref WG, Kamel I. Analysis of multi-dimensional space-filling curves. GeoInformatica, 2003,7(3):179−209. [doi: 10.1023/A:1025196714293]

[20] Serjantov A, Danezis G. Towards an information theoretic metric for anonymity. In: Proc. of the 2nd Privacy Enhancing Technologies Workshop. 2003. 41−53. [doi: 10.1007/3-540-36467-6_4]

[21] Freudiger J, Shokri R, Hubaux JP. On the optimal placement of mix zones. In: Proc. of the Int'l Symp. on Privacy Enhancing Technologies. 2009. 216−234. [doi: 10.1007/978-3-642-03168-7_13]

[22] Palanisamy B, Liu L. MobiMix: Protecting location privacy with mix-zones over road networks. In: Proc. of the Int'l Conf. on Data Engineering. 2011. 494−505. [doi: 10.1109/ICDE.2011.5767898]

[23] Freudiger J, Raya M, Félegyházi M, Papadimitratos P, Hubauxet JP. Mix-Zones for location privacy in vehicular networks. In: Proc. of the 1st Int'l Workshop on Wireless Networking for Intelligent Transportation Systems. 2007.

[24] Yang X, Wang B, Yang K, Liu C, Zheng B. A novel representation and compression for queries on trajectories in road networks. IEEE Trans. of Data Engineering (TKDE), 2018. [doi: 10.1109/TKDE.2017.2776927]

[25] Zhu H, Yang X, Wang B, Lee WC. Range-Based obstructed nearest neighbor queries. In: Proc. of the ACM Int'l Conf. on Management of Data (SIGMOD). 2016. 2053−2068. [doi: 10.1145/2882903.2915234]

[26] Wang B, Zhu R, Yang X, Wang G. Top-*K* representative documents query over geo-textual data stream. World Wide Web-Internet & Web Information Systems, 2017,20(8):1–19. [doi: 10.1007/s11280-017-0470-0]

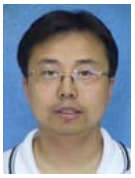
附中文参考文献:

[13] 李敏,秦志光.路网环境下位置隐私保护技术研究进展.计算机应用研究,2014,31(9):2576–2580. [doi: 10.3969/j.issn.1001-3695.2014.09.003]

[14] 朱怀杰,王佳英,王斌,杨晓春.障碍空间中保持位置隐私的最近邻查询方法.计算机研究与发展,2014,51(1):115–125. [doi: 10.7544/issn1000-1239.2014.20130694]

[17] 郑淼,王斌,杨晓春.路网环境下基于位置服务的隐私保护方法.华东师范大学学报(自然科学版),2015,183(5):116–127. [doi: 10.3969/j.issn.1000-5641.2015.05.010]

梁慧超(1993－),女,河南郑州人,硕士生,主要研究领域为位置隐私保护.



王斌(1972－),男,博士,副教授,主要研究领域为数据科学,数据流管理,数据质量管理.

崔宁宁(1988－),男,博士生,主要研究领域为空间数据库,隐私保护.

杨凯(1992－),男,硕士生,主要研究领域为时空数据管理.

杨晓春(1973－),女,博士,教授,博士生导师,CCF 高级会员,主要研究领域为数据库理论与技术,大数据管理与分析,数据隐私保护.