软件学报 ISSN 1000-9825, CODEN RUXUEW
*Journal of Software*,2016,27(7):1805−1821 [doi: 10.13328/j.cnki.jos.005053]

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

# 保护位置隐私近邻查询中隐私偏好问题研究*

倪巍伟[1,2], 陈 萧[1,2]

[1](东南大学 计算机科学与工程学院,江苏 南京 211189)

[2](计算机网络和信息集成教育部重点实验室(东南大学),江苏 南京 211189)

通讯作者: 倪巍伟, E-mail: wni@seu.edu.cn

**摘 要**: 近年来,位置服务中的隐私保护问题得到了研究者的持续关注,特别是近邻查询中位置隐私保护问题更是得到了广泛的研究.已有工作缺少对查询者个性化隐私偏好约束的系统研究,位置隐私与查询服务质量的兼顾,在隐私偏好约束下尤为困难:(1) 偏好强调个性与隐私模型侧重共性存在矛盾;(2) 偏好对查询中间结果动态可控依赖与查询简化中间结果的思想相抵触;(3) 连续查询中,支持隐私偏好存在基于候选解集攻击的风险.结合上述问题,提出保护位置隐私近邻查询中的隐私偏好问题,从位置隐藏原理及近邻查询性能与保护位置隐私内在制约机理的角度,对已有的位置隐藏与查询处理方法的性能及其对隐私偏好支持能力进行论述分析.进一步地,对支持隐私偏好与保护位置隐私查询内在制约机理进行了剖析,分析保护位置隐私近邻查询中支持隐私偏好需解决的主要问题,并对所归纳问题的可能解决方法进行了展望.

**关键词**: 位置服务;近邻查询;位置隐私保护;隐私偏好

**中图法分类号**: TP309

中文引用格式: 倪巍伟,陈萧.保护位置隐私近邻查询中隐私偏好问题研究.软件学报,2016,27(7):1805−1821. http://www.jos.org.cn/1000-9825/5053.htm

英文引用格式: Ni WW, Chen X. User privacy preference support in location privacy-preserving nearest neighbor query. Ruan Jian Xue Bao/Journal of Software, 2016,27(7):1805−1821 (in Chinese). http://www.jos.org.cn/1000-9825/5053.htm

## User Privacy Preference Support in Location Privacy-Preserving Nearest Neighbor Query

NI Wei-Wei[1,2], CHEN Xiao[1,2]

[1](College of Computer Science and Engineering, Southeast University, Nanjing 211189, China)

[2](Key Laboratory of Computer Network and Information Integration of Ministry of Education (Southeast University), Nanjing 211189, China)

**Abstract**: Privacy protection problem in location based service receives continuous attentions in recent years, especially for location privacy protection in location based nearest neighbors query. Existing work however often neglects or fails to cultivate in accommodating query users' privacy preference requirements. The constraint of privacy preference burdens the trade-off between location privacy protection and quality of service in privacy-aware location based query service. Several issues need to be addressed: (1) Privacy preferences contradict with privacy models diametrically in terms of personality and commonality they focus on; (2) There is a dilemma between privacy preferences and query performance that preferences require intermediate query results be dynamic and adjustable while simplified intermediate query results commonly promise good performances; and (3) Privacy preferences incur the attack originated from intersection inferring to candidate answer sets in continuous location based queries. In this survey, the privacy preference problem in location based nearest neighbor query is identified and presented. The performance of existing location obfuscation and query techniques,

* 基金项目: 国家自然科学基金(61370077, 61003057)
Foundation item: National Natural Science Foundation of China (61370077, 61003057)
收稿时间: 2014-11-26; 修改时间: 2015-08-17, 2016-02-02; 采用时间: 2016-02-23; jos 在线出版时间: 2016-03-16
CNKI 网络优先出版: 2016-03-17 09:57:19, http://www.cnki.net/kcms/detail/11.2560.TP.20160317.0957.004.html

as well as their ability in accommodating users' privacy preferences, are discussed in terms of location obfuscation principle and inherent restricting mechanism between nearest neighbor query performance and location protection. Subsequently, inherent restricting mechanism between location privacy preserving nearest neighbors query and privacy preference support is detailed, and some major problems originated from location privacy preference are demonstrated. Finally, some possible solutions to these problems are elaborated and the future research work is suggested.

**Key words**: location based service; nearest neighbor query; location privacy protection; privacy preference

无线通信和智能移动终端的广泛应用,推进了基于位置服务(location based service,简称 LBS)的快速发展.目前,位置服务已广泛应用于军事、交通、物流、医疗等领域[1,2].基于位置的近邻查询作为位置服务的支撑应用,有广泛的应用前景.LBS 服务提供方根据移动对象当前位置,为其提供距其最近的若干近邻目标对象信息.尽管 LBS 为各类用户提供了诸多便捷服务,但其对位置的共享和传播也引发了隐私安全方面的争论.2010 年 7 月,互联网安全服务企业 Webroot 的一项调查显示:超过半数的 LBS 用户担心位置服务可能导致自身隐私的泄露.如何在不侵犯移动对象位置隐私的情况下,为移动对象与数据使用方提供基于位置的近邻查询服务,已成为数据库与信息安全领域的研究热点.

保护位置隐私查询(如图 1 所示)是数据库领域隐私保护位置服务研究的主要内容,其应用场景为查询者不信任 LBS 服务提供方(查询响应方),但希望获得 LBS 提供的关于自身位置的服务,保护位置隐私查询侧重于在客户/服务器模式下,位置信息安全交互机制及查询处理策略的研究.例如,旅行者希望无需告知服务提供方其准确位置,即可获取距其最近的旅馆的信息.其处理特征为联机实时处理,不同于已有的隐私保护数据发布以数据为中心,保护位置隐私查询以提供服务为中心.

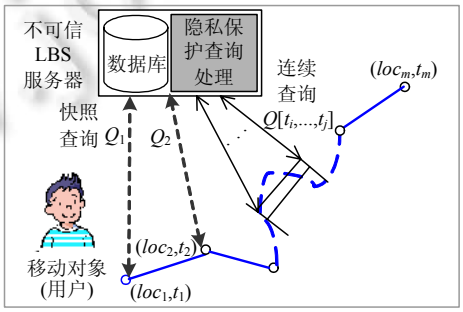



Fig.1 Frameworks of privacy-preserving location based query

图 1 保护位置隐私查询框架

近年来,国内外研究者在保护位置隐私近邻查询方面取得了显著成果[3−35],提出了一系列保护位置隐私快照查询(snap-shot query)和连续查询(continuous query)解决方法.所采用的主要技术方法归纳如下[4]:

(1) 位置干扰(location obstruction):查询者持续地向 LBS 服务提供方提交关于所选假位置的查询请求,直到返回满足其查询精度与隐私安全要求的结果[12];

(2) 空间变换(space transformation):设计数据变换方法将原始二维平面坐标转换为新空间坐标系下的编码,以保护位置隐私安全,要求转变后,空间尽可能地保持原二维坐标点间的几何位置关系,以保证查询的精度;

(3) 空间混淆(spatial cloaking):将查询者的位置扩展为包含该位置的泛化区域,并提交 LBS 服务器,由查询发起者或可信第三方从返回的候选解中甄别出查询结果(时间混淆技术相对简单,采用延时等待满足隐私保护条件移动对象出现的思路,故而只列出空间混淆).

保护位置隐私近邻查询需兼顾查询者位置的安全与查询性能.目前,隐私保护中的偏好问题开始引起研究者的关注.2010 年,文献[22,23]提出了社会网络隐私保护中的个性化保护和用户偏好问题,强调保护强度和模式

的可调控性.在保护位置隐私查询领域同样存在偏好问题,仅追求位置隐私保护的强度并不能满足应用需要.

(1) 攻击者所掌握的背景知识不可预知,使得位置隐私保护机制所提供的保护强度具有相对性.考虑到存在不同能力的攻击者,用户希望对自身位置保护强度具有灵活的调控能力;

(2) 不同用户对自身的位置安全有不同的要求,即便同一用户,不同场景下,其位置保护的强度要求也不尽相同.例如:商业用户对其位置保护的强度要求通常高于普通用户;同一用户在娱乐场所附近时的位置保护强度要求往往高于在公共区域时,等等;

(3) 较高的隐私保护强度往往以服务质量(查询性能)的折损为代价,移动对象对位置保护强度与服务质量的偏好具有动态性,偏好的动态性势必要求移动对象对隐私保护强度具有灵活的调控能力.

我们用隐私偏好(privacy preferences)表示保护位置隐私近邻查询中,不同场景下,查询者对其位置保护强度、查询效率与查询准确性的个性化调控要求.隐私偏好的具体内容包括:某轮查询中,位置隐私保护强度、查询效率与查询准确性这 3 要素中,查询者优先关注的要素;是否接受牺牲准确性换取查询效率的提升;是否接受牺牲效率换取查询准确性或位置保护强度的提升;希望获得的位置隐私保护强度的等级等.

例如,旅行者旅途劳顿,途经城市准备投宿,发起距其最近的 3 个旅馆的查询请求.由于城市道路较拥挤,车辆及行人速度都相对较慢,而城市内环境复杂,对个人行踪的安全性要求相对提高,这时,旅行者往往能够接受适当降低查询效率以获取查询准确性和位置隐私保护强度的提升,其偏好可对应优先关注位置隐私保护强度,可以适当牺牲查询效率,不接受牺牲查询准确性;第二天重新启程离开城市,在向下一个城镇进发的路途中,同伴突患阑尾炎,这时,旅行者需要发起距其最近的 3 家医院的查询,由于病情紧急,旅行者对查询效率和结果准确性的要求自然会提高,而愿意牺牲位置隐私保护强度,其偏好表现为优先关注效率,接受降低隐私保护强度获取效率或准确性的提升,不接受牺牲查询准确性.由于两次查询的环境和查询者的心态不同,查询者选择了截然不同的偏好约束,这就要求相应查询机制具有动态支持隐私偏好的能力.

虽然已有的研究在隐私模型构建、高强度隐私保护机制等方面已取得显著成果,但从已有的模型和位置隐私保护机制对隐私偏好的支持方面考虑,主要存在隐私模型和位置隐私保护机制对隐私偏好的支持能力不足的问题.

数据隐私安全通常用隐私模型(privacy model)描述,隐私模型对隐私形式和需预防攻击的类型进行定义.目前,*k*-匿名模型在隐私保护位置服务的研究中得到了广泛的应用.保护位置隐私查询过程中,隐私偏好完全决定于查询发起者个体,具有个性化和动态性的特征;而基于 *k*-匿名的位置隐私模型严重依赖于查询者所在区域的不可控、不可预知的移动对象分布.存在隐私偏好的灵活可控需求与 *k*-匿名类位置隐私模型的不可控现实间的矛盾(例如,文献[11]提出了支持用户对查询位置时空约束的个性化匿名模型,查询者位置时空约束的动态性与匿名服务器信息采集固有的时效约束间的冲突,加剧了响应失效风险),难以兼顾隐私偏好与查询服务质量.目前,保护位置隐私查询研究对查询者位置的隐藏可以归纳为借助空间变换及加密、隐私信息检索(private information retrieval,简称 PIR)[36]和匿名思想,匿名思想具体又可分为基于空间混淆匿名与假位置匿名.空间变换及加密、隐私信息检索能够提供较强的位置隐私保护,但通常以较大的时间开销为代价.当查询者在隐私安全强度与查询性能间更倾向于后者时,由于该类隐藏机制的位置隐私安全往往完全依赖于固定密钥的安全性,查询者与 LBS 服务提供方均缺少灵活的调控能力,难以提供对查询者隐私偏好的有效支持.基于假位置的匿名(对应位置干扰技术)中,客户端和 LBS 服务器间往往需要多轮迭代,以获取满足精度要求的查询结果,不可控的多轮迭代加剧了对查询者提供隐私偏好支持的难度.

位置服务正在潜移默化地改变着人们的工作和生活模式,其核心在于“服务”.而人们对服务需求由“无差异”到“个性化”的演化是事物发展的必然,在这一背景下,对保护位置隐私近邻查询中支持隐私偏好与保护位置隐私安全、兼顾服务质量间的内在制约机理与实现技术进行研究,有助于推进位置服务的继续深入和服务的安全化、个性化.

本文总体结构如图 2 所示.首先引入保护位置隐私近邻查询中存在的隐私偏好问题;第 1 节从位置隐私模型和查询机制的角度概述保护位置隐私近邻查询研究的现状.在此基础上,第 2 节分析和论述主要位置隐藏与

查询技术对隐私偏好的支持能力.第 3 节解析隐私偏好施加于保护位置隐私近邻查询的约束.第 4 节分析支持保护位置隐私近邻查询中隐私偏好约束面临的主要问题.第 5 节对隐私偏好研究及可能的解决方案进行展望.最后总结全文.

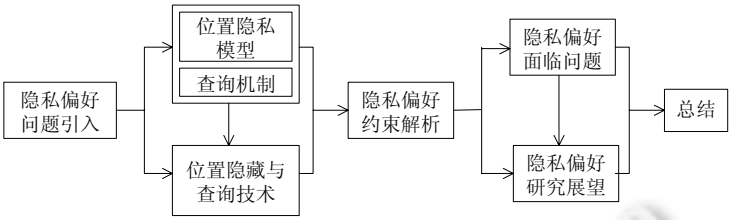



Fig.2 Architectural structure of the paper

图 2 本文总体结构

## 1 保护位置隐私近邻查询研究现状

本节对保护位置隐私近邻查询领域国内外的研究现状进行分析,主要从已提出的隐私模型、保护位置隐私查询技术方面加以展开.

### 1.1 位置隐私模型

目前,保护位置隐私查询研究领域的位置隐私模型主要基于 $k$-匿名思想,采用空间混淆技术实现.模型要求混淆区域至少包含与查询者类似的其余 $k$–1 个对象,以保证查询者被逆推的概率小于 1/$k$;文献[10]进一步提出了($k$,$A_{\min}$)模型,增加混淆区域面积至少为 $A_{\min}$ 约束;文献[11]提出了支持查询者对查询位置添加时空约束的个性化匿名模型,这种基于 $k$ 匿名的个性化模型难以同时兼顾时间、空间与匿名度 $k$ 的三元约束,不可避免地导致响应失效风险.该类位置隐私模型需要区域内移动对象位置信息的协调,区域内大量移动对象的实时位置信息往往需要由可信第三方采集(尽管局部区域内移动对象的分组自治协调可以避免对可信第三方的依赖,但增加了隐私泄露的风险和通信代价),对可信第三方及移动对象实时位置分布的依赖,使得该类模型难以有效地提供隐私偏好支持.

近年来,差分隐私模型以其强健的数学理论基础和对攻击者背景知识的不敏感得到了研究者的持续关注,差分隐私由隐私参数$\varepsilon$控制隐私保护强度与数据可用性,可以实现某个数据集中单条记录的添加或删除不影响隐藏后数据集的任何计算结果.差分隐私要求待隐藏数据相互独立,主要基于静态数据集进行隐藏.而在保护位置隐私查询中,查询者的位置是动态的,具有时效性,查询者不同时刻的位置与其他移动对象的位置具有关联性,限制了差分隐私技术的应用,目前还未见基于差分位置隐私模型的保护位置隐私近邻查询工作.目前,差分隐私模型的研究主要集中在数据隐藏发布方面,在查询应用中,主要集中于范围计数查询以及部分批量查询.

### 1.2 保护位置隐私近邻查询研究

保护位置隐私近邻查询包括快照查询和连续查询两类.快照查询对应查询者在运动过程中,向 LBS 服务提供方发起关于自身实时位置的近邻查询(例如,查找距离查询者当前位置最近的若干旅馆等);连续查询对应查询者在移动过程中,连续向 LBS 服务提供方发起关于其实时位置的相同查询.保护位置隐私快照查询侧重于保护位置隐私前提下,查询效率与查询结果反馈的实时性;保护位置隐私连续查询研究侧重于查询者运动模式(速度、方向等)与历史查询信息(如历史查询中位置隐匿结构等)对位置隐匿机制安全性的影响.代表性方法分别见表 1 和表 2.

表 1 中,文献[11,12]采用空间混淆技术,通过可信第三方将查询者位置隐藏为包含该位置的混淆区域,要求混淆区域满足 $k$-匿名隐私模型,并将隐藏后混淆区域提交服务器进行查询处理.服务器端要进行复杂的区域近邻查询,导致服务器计算量激增,其支持隐私偏好能力只能通过对匿名度 $k$ 的设置来体现.$k$ 值越大,通常位置隐私保护强度也相对较高,但也将导致服务器处理代价的增加,难以兼顾查询效率.因此,这两种方法支持隐私偏

好的能力较弱.文献[13]采用空间变换技术,利用 Hilbert 填充曲线的连续贯穿性对平面区域及 POI 坐标进行 Hilbert 编码,实现兼顾位置隐私的近邻查询.然而,DCQR 方法所提供的位置隐私保护强度完全依赖于 Hilbert 填充曲线的五元参数的安全性,不支持位置隐私保护强度的可调控.SpaceTwist 方法[14]采用位置干扰技术实现保护位置隐私近邻查询,查询者不断地向服务器发送关于假位置的查询请求,因为服务器与查询者之间交互的信息无关查询者的真实位置,因此无需可信第三方介入.SpaceTwist 方法可以通过提前终止假位置迭代查询以提高查询效率,但难以兼顾查询结果的正确性.此外,所提供的位置隐私保护安全性难以度量和调节.基于 PIR 技术的文献[15, 17]同样由于所提供的位置隐私保护强度依赖于固化的密钥安全性,不能提供隐私保护强度的动态调控.在保护位置隐私连续查询方面,已有的研究多数采用空间混淆技术,扩展传统的 $k$ 匿名位置隐私模型来避免各类基于查询者历史运动信息的攻击.例如,文献[30]提出了历史 $k$ 匿名模型,通过在每轮查询生成的混淆区域中保持稳定对象,实现对跟踪用户查询轨迹攻击的预防;文献[31]在 $k$ 匿名位置隐私模型的基础上引入 $m$-不变性约束,避免在两轮查询的匿名集相交时,攻击者利用交集位置区域及对象分布特征发起对查询者位置的逆推等.

**Table 1** Research on location privacy aware snap-shot query

**表 1** 保护位置隐私快照查询研究

| 序号 | 名称 | 算法思想 | 服务器端计算量 | 可信第三方 | 支持隐私偏好能力 | 文献 |
|---|---|---|---|---|---|---|
| 1 | Cliquecloak | 混淆区域,个性 $k$-匿名 | 重 | 在线第三方 | 较弱 | [11] |
| 2 | Casper | 混淆区域,$k$-匿名 | 重 | 在线第三方 | 弱 | [12] |
| 3 | DCQR | Hilbert 数据变换 | 较少 | 离线第三方 | 无 | [13] |
| 4 | SpaceTwist | 位置干扰 | 适中 | 不需要 | 无 | [14] |
| 5 | AHG | PIR | 适中 | 不需要 | 无 | [15] |
| 6 | Transfer | 同态加密,PIR | 适中 | 离线第三方 | 无 | [17] |
| 7 | PRQ | 借助 PEB 索引提供附加隐私保护 | 较重 | 由基本查询机制决定 | 由查询机制决定 | [18] |

**Table 2** Research on location privacy aware continuous query

**表 2** 保护位置隐私连续查询研究

| 序号 | 攻击形式 | 隐私模型 | 主要技术方法 | 在线可信第三方 | 文献 |
|---|---|---|---|---|---|
| 1 | 非等概率攻击 | AD-匿名 | 熵、混淆区域 | 是 | [5] |
| 2 | 查询采样攻击、查询轨迹攻击 | $k$-共享区域 | 混淆区域、记忆结构 | 是 | [6] |
| 3 | 基于运动速度的匿名集相交攻击 | $(k,A,dt)$模型 | 质量模型、区域混淆 | 是 | [7,19] |
| 4 | Timing,Transition 攻击 | Mix-Zone 模型 | 混淆区域、熵 | 是 | [8] |
| 5 | 跟踪用户查询轨迹 | 历史 $k$-匿名 | 混淆区域 | 是 | [30] |
| 6 | 匿名集相交攻击 | $k$-匿名、$m$-不变性 | 混淆区域 | 是 | [31] |

目前,保护位置隐私快照查询侧重于查询精度与隐私保护强度的兼顾,依赖 $k$-匿名思想进行空间混淆的处理策略,普遍存在通信及服务器开销大的缺陷;保护位置隐私连续查询通常可以看作由一系列快照查询组成,多数采用空间混淆技术实现,其隐私模型主要依赖于 $k$-匿名及其扩展思想,侧重于连续混淆区域间位置关联性造成的隐私泄露的预防,继承了 $k$-匿名思想对隐私偏好支持能力弱的特点.

此外,当前,保护位置隐私近邻查询服务质量方面的研究还较少,其服务质量主要体现在性能与位置隐私安全性两方面(如图 3 所示).其中,性能度量已有较成熟的标准:效率通常用完成每轮操作时耗(及通信量)来衡量;查询机制的可扩展性通常采用单位时间完成的查询次数来度量.攻击者可能掌握背景知识的不确定和位置隐私保护技术的多样性,使得位置隐私保护机制所提供的保护强度的度量变得困难,已有的保护机制通常结合所假设攻击模型和采用的隐藏技术特点对各自所提供的位置隐私安全强度进行定制分析,缺少统一的量化度量机制.

由此可见:虽然目前国内外在保护位置隐私近邻查询方面进行了大量卓有成效的工作,但针对隐私保护位置服务中支持隐私偏好的隐藏与查询技术还较缺乏.目前,支持隐私偏好的保护位置隐私近邻查询研究仍处于起步阶段.

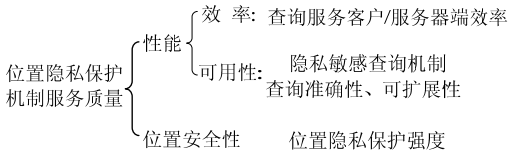



Fig.3 Service quality of location privacy protection schema

图 3 位置隐私保护机制服务质量

## 2 位置隐藏与查询处理技术

本节讨论保护位置隐私近邻查询所采用的主要位置隐藏及查询处理技术.保护位置隐私近邻查询的基本策略是:对查询者的准确位置进行隐藏,将隐藏后位置信息发送给 LBS 服务器(位置服务提供方)处理并返回查询结果.目前,常见的位置隐私保护技术可以归纳为空间混淆、空间变换以及位置干扰这 3 类.

### 2.1 空间混淆技术

空间混淆的基本思想是:用一个包含查询者当前位置的特殊平面区域替代查询者的准确位置,并向 LBS 服务器发起关于该平面区域的近邻查询请求.该平面区域需要满足查询者保护位置隐私的需求,查询者保护位置隐私需求通常用隐私模型描述.例如,$(k,A_{\min})$模型要求混淆区域覆盖至少 $k$–1 个同类移动对象,以保证查询者被辨识的概率低于 1/$k$.空间混淆通常依赖于局部区域内的移动对象分布信息,这些信息需要由可信第三方采集,因此空间混淆技术通常需要可信第三方的介入.LBS 服务器获取包含查询者位置的混淆区域后,执行关于混淆区域的近邻查询处理,获取混淆区域内任意位置可能的近邻点集的并集,将并集作为候选解集返回可信第三方,可信第三方从候选解集中筛选出查询者的近邻 POI,返回查询者.目前,研究者在基于空间混淆的保护位置隐私查询方面进行了大量的研究工作,取得了一系列成果[5−12,18−21].

基于空间混淆的保护位置隐私近邻查询需要在 LBS 服务器端部署特殊的区间近邻查询处理机制.以矩形混淆区域的最近邻查询为例(如图 4 所示),$q_1$,$q_2$,$q_3$,$q_4$ 为给定 POI 点集内到矩形 4 个顶点最近的 POI 点,$e_1$ 为 $q_1q_2$ 的中垂线与 $ab$ 的交点,对其余各边进行类似处理,则矩形内任意位置在 POI 点集内的最近邻一定位于图中以矩形顶点和 4 个交点为圆心(半径如图中标示)的圆区域内.

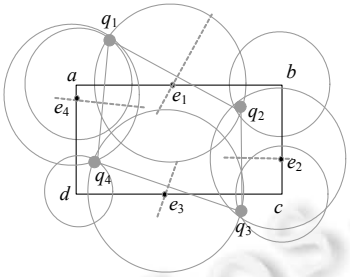



Fig.4 Illustration of nearest neighbor region query

图 4 区间近邻查询示意图

基于空间混淆的位置隐藏中,位置隐私保护的强度可以通过混淆区域的规模(面积)、区域内所包含的移动对象的分布结构来体现,混淆区域的面积越大,区域内包含的移动对象越多、移动对象特征与查询者特征的一致性越高,查询者位置隐私保护的强度也越高.该模式下,候选解区域受混淆区域的规模、位置以及服务器端 POI 分布等诸多因素的影响,其与混淆区域间缺少内在的制约关系.尽管查询者可以通过可信第三方调整目标混淆区域的规模以实现对隐私保护强度的调节,但候选解区域与混淆区域间的这种弱耦合,使得偏好调节变得不可控.同时,基于空间混淆技术的保护位置隐私查询还存在以下不足.

(1) 可信第三方容易成为系统的瓶颈.空间混淆的实质是:用特定区域替代查询者的准确位置进行查询,该区域需要包含足够多的与查询者具备共性特征的移动对象.对大量移动用户位置、特征信息的获取依赖于可信第三方,并且查询请求及候选解集的接收都由可信第三方完成,导致可信第三方成为系统的瓶颈;

(2) LBS 服务器端处理开销大.空间混淆技术将查询者的准确位置泛化为特殊区域(通常为矩形或圆形),LBS 服务器端需要进行区域近邻查询处理,区域近邻查询的复杂度远比常规查找某位置点的近邻的复杂度要高,在大量移动对象访问的情况下,服务器端处理开销较大.

### 2.2 空间变换技术

空间变换的基本思想是:设计数据映射机制,将查询者位置及原平面坐标下的 POI 位置坐标映射到一个新的数据空间,籍此实现对查询者原始位置的隐藏.为了兼顾查询的准确性,通常要求所设计的空间映射机制尽可能地保持变换前后位置间距离关系不变.类似空间混淆技术,空间变化操作亦需要由第三方执行.但不同于空间混淆技术依赖在线可信第三方全程介入位置隐藏与查询处理过程,执行空间变换的第三方可以采用离线方式进行.例如,文献[13]采用 Hilbert 编码机制将 LBS 服务器端 POI 坐标点集映射为一维 Hilbert 编码,将二维平面查询简化为一维范围查找.通过在客户端嵌入具有 Hilbert 编码/解码功能的模块,查询者将自身的平面位置坐标编码为对应的 Hilbert 单元值,并将 Hilbert 编码值传输给服务器,利用 Hilbert 曲线函数的内聚性质:若两个平面坐标的 Hilbert 编码值相近,则这两个平面点位置亦相近.服务器端将距离查询者位置的 Hilbert 编码最近的 POI 的 Hilbert 编码返还查询者,查询者接收该 POI 编码,并解码得到原始 POI 坐标,作为距查询者位置最近的 POI.

基于空间变换的位置隐藏及保护位置隐私查询技术具有位置隐藏与查询处理效率高的优点,LBS 服务器通常不需要进行复杂的处理.但其查询准确性取决于所采用的空间变换机制的保距性,即,变换机制能否严格保证变换前后位置间的距离关系不改变.然而,严格保距的数据变换往往容易被攻击者逆推,即,攻击者根据已掌握的变换方法及部分原始位置数据,可以由变换后的新位置拟逆推出其变换前的原位置信息,导致查询者位置隐私的泄露.以隐私保护数据发布中经典的保距隐藏算法 RBT 为例,RBT 算法能保证隐藏前后任意 3 个数据点(不局限于二维平面坐标)间的距离关系不变.但文献[37]已给出证明:任意两条记录隐藏前后对应关系的泄露,将导致整个隐藏算法的失效.目前,采用空间变换的保护位置隐私近邻查询研究多数采用非严格保距变换方法,以提高位置隐私的安全性,研究重点在于对非保距变换可能导致的查询结果不准确问题的解决.例如,文献[13]中采用将查询者位置与 POI 位置转换为统一 Hilbert 编码的方法实现保护位置隐私最近邻查询.然而,位置相邻的两个平面坐标点,其 Hilbert 编码未必相近,从而导致返回的 POI 编码可能不是查询者的真实最近邻 POI.为了解决这一问题,文献[13]提出了双向 Hilbert 曲线编码方案,这无形中增加了离线编码的工作量,同时也增加了客户端及 LBS 服务器端的计算量.

基于空间变换的位置隐藏所提供的位置隐私保护强度取决于空间变换方法的逆推复杂度,通常具有较高的安全性.但对于某个保护位置隐私查询系统而言,所选择的变换机制是固定的,其所提供的位置隐私保护模式和强度亦固定不变.因此,基于空间变换的保护位置隐私查询机制所提供的位置隐私保护强度通常不具备可调节性.

近几年,基于隐私信息检索(PIR)[36]的保护位置隐私查询技术得到了研究者的关注[15].隐私信息检索理论最早被应用于访问网络中的外包数据(outsourced data),用户可以检索一个不可信服务器上的任意数据项而不暴露用户检索的数据项信息.实现 PIR 的方法可以按照隐私保护的强弱分为基于信息论的 PIR 方法和基于计算能力的 PIR 方法:基于信息论的 PIR 方法保证攻击者无论拥有怎样的计算能力,都不能区分用户对不同数据项的访问;基于计算能力的 PIR 方法假设攻击者不具有计算求解某个难题的能力,从而保证攻击者不能区分用户对不同数据项的访问.基于信息论的 PIR 方法有且只有一个平凡的解法:即将全部信息都发送给客户端[36].这需要的传输代价是 $O(n)$,其中,$n$ 为数据库的规模.已有的基于 PIR 的保护位置隐私查询研究通常采用基于计算能力的 PIR 方法.基于隐私信息检索的方法与加密方法类似,能够提供高效、安全的位置隐私保护,但其位置保护的强度依赖于固化的数据映射机制,同样难以提供保护强度动态可调节的功能.

### 2.3 位置干扰技术

不同于前两种技术,位置干扰技术并不对查询者位置进行显式的隐藏,其基本思想是:查询者选取假位置,并向 LBS 服务器发起关于假位置的同类查询;获取假位置的近邻 POI 后,查询者进一步结合真实位置、假位置以及假位置的近邻 POI 分布间的位置关系,判断是否满足查询终止条件(通常,终止条件是满足查询者对查询结果准确性的要求),若不满足,则选取新的假位置并发起新一轮查询.如此迭代,直到返回的 POI 信息满足其查询准确性要求为止.

相较前两种技术,基于位置干扰的保护位置隐私查询具有的显著优点是位置隐藏与查询过程无需可信第三方介入.此外,查询者对查询过程具有调控能力,当查询者判断某轮迭代查询返回的结果满足自身对查询准确性的要求时,可以选择停止查询过程.

位置干扰技术的主要不足体现在以下几个方面.

(1) 查询准确性难以保证

基于位置干扰的保护位置隐私近邻查询采用迭代试探策略,不断发起关于假位置的同类查询,查询者检测返回结果是否满足准确性要求,需要经过多少轮迭代方可获取准确查询结果,存在很大的未知性,往往最终不得不牺牲查询准确性以兼顾查询效率.

(2) 查询代价较大

基于位置干扰的保护位置隐私查询由多轮关于不同假位置的同类迭代查询组成,并且迭代次数不可预判.在现实应用中,查询者与 LBS 服务器间的多轮通信不仅降低了查询效率,不可预知迭代次数的通信也极大地增加了通信开销,提高了经济成本.

(3) 存在位置隐私泄露隐患

基于位置干扰的保护位置隐私查询难以提供对查询者位置隐私的可量化保护(例如最小逆推区域面积阈值等),并且查询者位置存在被逆推的可能,文献[14]已证明,查询者位置存在被逆推限定在一个不规则平面区域内的可能.尽管能够通过增加单次通信包内包含数据坐标数目的方法缓解这种风险,但这种逆推风险无法避免.

相较空间混淆与空间变换技术,位置干扰为查询者提供了查询处理过程中实时调控干预处理效果的能力,例如提前终止迭代,以查询准确性为代价换取查询效率的提高.这种实时调控干预能力与隐私偏好所要求的查询者对查询效果的动态调控思想相符.就这个意义而言,利用位置干扰技术解决隐私偏好问题具有一定的可行性.但是位置干扰技术固有的迭代轮次未知、所提供的位置隐私安全性难以量化问题,以及假位置与查询准确性、查询效率、位置隐私安全性间的关联制约关系复杂、难以建模,这些问题都严重影响了位置干扰技术对隐私偏好约束的支持,特别是当查询者不愿意牺牲查询准确性,只愿意在查询效率与位置隐私安全性间进行取舍时,位置干扰技术更将完全失效.

## 3 隐私偏好约束解析

目前,保护位置隐私近邻查询研究通常在查询者位置安全、查询效率和查询准确性间寻求折衷,难以提供有效的隐私偏好支持.

### 3.1 位置隐私偏好与隐私保护查询效能

从保护位置隐私近邻查询的内在机理角度分析,对查询者位置隐私保护的强度越高,LBS 服务器用于处理隐藏后位置的时间(通信)开销也越大,相应的近邻查询效率也越低;而近邻查询的准确性取决于位置隐藏机制对平面坐标点间距离关系维持的准确性,某个位置隐藏方法能够严格地维持隐藏前后位置间的距离关系,通常能够有效地兼顾保护位置隐私查询结果的准确性.然而事物均有两面性,位置隐藏方法能够准确维持位置点隐藏前后的距离关系,也意味着攻击者可以获取的背景知识的增加,这又将加大查询者位置隐私泄露的风险,因此从这个角度分析,查询准确性与位置保护强度间亦存在矛盾.可见,位置隐私安全性、查询效率以及查询结果准确性三者间彼此关联,又存在矛盾.目前,已有的研究大多难以同时兼顾这 3 个方面,而是在查询效率、查询准确

性和位置保护隐私强度这 3 个方面各有侧重(见表 3).

**Table 3** Comparison among obfuscation technologies

**表 3** 主要隐藏技术性能分析比较

| | 空间混淆 | 空间变换 | 位置干扰 |
|---|---|---|---|
| 保护位置隐私强度 | 中 | 高 | 较低 |
| 查询效率 | 中 | 高 | 低 |
| 查询结果准确性 | 准确 | 不确定 | 不确定 |
| 依赖可信第三方 | 是 | 是 | 否 |
| 支持隐私偏好能力 | 较弱 | 弱 | 较弱 |

对隐私偏好的支持,要求保护位置隐私查询机制为查询者提供以下功能:根据其对位置保护强度、查询效率以及查询准确性方面的偏好要求动态调节查询效果,即:在满足查询者关于位置保护强度、查询效率与查询准确性偏好需求的基础上,实现三者的带约束的兼顾和动态调节.

### 3.2 位置隐私偏好约束机理分析

在保护位置隐私近邻查询中,查询者需要获取的 POI 信息存储在 LBS 服务器端;而 LBS 服务器不知道查询者的位置,需要结合 LBS 服务器及查询者两端的信息才能完成查询过程.在保护位置隐私不泄露的前提下,从查询发起时查询者是否掌握服务器端 POI 信息的角度,可以将保护位置隐私近邻查询分为有指导查询与无指导查询两种模式.

- 无指导查询:查询者不掌握服务器端的 POI 分布信息,查询者将其位置坐标隐藏处理为粒度更粗糙(保证精确位置不泄露)的位置信息提交给 LBS 服务器,LBS 服务器结合隐藏后的位置信息进行定制查询分析,生成候选解集(包含查询目标对象的 POI 集合)返回查询者,供查询者筛选目标查询结果;
- 有指导查询:查询者预先获取服务器端的部分 POI 信息,借助已获取的 POI 信息,确定查询者当前位置的近邻 POI 的可能分布范围,进而向 LBS 服务器发起特定范围的定向查询,获取目标查询结果.

基于空间混淆和数据变换的保护位置隐私查询属于无指导查询,基于位置干扰的查询属于有指导查询.从位置隐藏与查询处理过程是否有指导的角度分析隐私偏好支持问题,对无指导查询,位置隐藏与查询过程中查询者并不掌握服务器端的 POI 分布信息,而对隐私偏好的支持要求查询者能够根据自身对位置隐私保护强度、查询效率或查询结果准确性的需求动态调节隐藏与查询过程,这种动态调节能力需要额外掌握的服务器端的 POI 信息作为依据,显然,无指导查询模式难以有效地实现隐私偏好支持.例如,尽管基于空间混淆技术的查询机制能够采用查询者指定混淆区域范围及区域内包含移动对象规模的方法来体现查询者对位置保护的强度需求,但这种不掌握服务器端数据分布的单向调节,不可避免地带来服务器端处理代价的激增,难以兼顾调节的动态性与查询效率.从这一角度而言,基于位置干扰的有指导隐藏与查询模式为支持隐私偏好提供了便利,查询者对服务器端部分 POI 分布信息的获取,为其构建查询者动态调控机制创造了条件.但位置干扰技术存在位置隐私保护安全性较弱和查询过程迭代轮次不可预知的固有缺陷,严重影响其对隐私偏好的支持.

支持隐私偏好与现有的位置隐藏以及近邻查询处理方法的主要冲突体现在以下几个方面.

(1) 偏好强调个性与位置隐私模型侧重共性存在矛盾

隐私偏好约束通常体现的是个体查询者在某个时刻(位置场合)对隐私保护强度、查询准确性及查询效率的动态调控需求,具有较强的实时化和个性化特征;而目前,位置隐私模型侧重从隐藏机制所提供的位置保护强度的通用性角度提取外在共性因素构建隐私模型,存在偏好彰显个性与模型侧重共性的冲突.

(2) 偏好对查询中间结果动态可控依赖与查询简化中间结果的思想相抵触

目前,保护位置隐私查询方法多数都显式或隐式地依赖于查询中间结果,实现对目标查询结果及查询者位置的隐藏.查询者关于位置保护强度、查询准确性及查询效率的偏好调控,势必需要查询中间结果亦动态可控,从查询中间结果的角度分析,中间结果规模越大、结构越复杂,所提供的位置保护强度越高;然而从提高查询效率的角度看,简化查询中间结果是获取较高查询效率的有效保障.

(3) 连续查询中支持隐私偏好存在基于候选解集攻击的风险

源于服务器端返回候选解集结构的稳定性,已有的保护位置隐私连续查询多数着眼于从查询者的运动模式(例如速度、方向等)以及查询者提交 LBS 服务器的信息(通常为隐藏后位置信息)角度防止可能存在的对查询者当前或历史查询位置的逆推攻击.引入隐私偏好约束后,连续查询中,查询者每次隐私偏好的差异将可能导致 LBS 服务器返回的候选解集结构的个性化差异,存在基于连续返回的候选解集发起攻击的风险.

## 4 支持隐私偏好面临的主要问题

目前,已有的位置隐藏与查询处理方法大多难以提供有效的隐私偏好支持,保护位置隐私近邻查询中支持隐私偏好,主要面临下述问题.

### 4.1 隐私偏好与隐私模型构建角度选取问题

隐私偏好是用户对隐私保护的一种个性化要求,隐私偏好的实现以所定义的隐私模型为基础.因此,构建便于支持隐私偏好的隐私模型必须考虑以下两方面因素.

(1) 模型共性与偏好个性的兼顾;

(2) 模型现实描述能力与实现难度的兼顾.

目前,已有的研究大多从移动对象位置分布的角度构建隐私模型,这存在以下不足.

(1) 隐藏方法严重依赖于移动对象的实时位置分布,这些信息需要可信第三方采集处理,使得可信第三方成为系统的瓶颈,造成模型实现困难和系统性能较差;

(2) 数据分布描述个体数据的共性特征,隐私模型对个体对象分布的依赖过度彰显了模型的共性,加剧了兼顾模型共性与偏好个性的难度.

需要选取合适的角度构建隐私偏好和隐私模型,选取的角度应能兼顾上述要求.

### 4.2 隐私偏好与查询候选解集间制约机理剖析

保护位置隐私查询的基本思想是:通过位置隐匿机制向 LBS 服务器提交查询发起者隐匿后的位置结构信息,服务器返回包含目标查询结果的候选解集(candidate answer set),由查询发起者/可信第三方从中筛选出查询结果.候选解集中的非目标解对目标查询结果起到隐匿作用.候选解集的存在,不可避免地增加了查询代价.同时,候选解集也是攻击者籍以逆推查询者位置的重要线索.

(1) 候选解集与隐私偏好间存在紧密关联

支持隐私偏好的保护位置隐私查询技术需实现以下目标.

- 首先,对查询者隐私偏好的支持不影响查询结果的准确性;
- 其次,支持隐私偏好对查询性能的影响应尽量小.

这两个目标的实现与查询候选解集的结构密切相关:一方面,候选解集是攻击者可获取的重要背景知识,查询者可通过生成合适的候选解集,在兼顾查询准确性的同时保护其位置不泄露,其隐私偏好的支持同样需要对候选解集的组成与结构的控制来实现;另一方面,候选解集的生成、传输和处理是保护位置隐私查询所需额外计算和通信开销的重要组成部分,隐私偏好对查询性能的影响源于其对候选解集结构与规模的影响.从这个角度考虑,候选解集与隐私偏好之间表现出关联一致性.如图 5 所示:若查询者希望提高位置隐私保护的强度,通常将导致查询性能和准确性方面的损失,隐私保护强度与查询效率或准确性间将失衡,这时就需要支持隐私偏好的“补偿机制”平抑这种损失,达到新的平衡.支持隐私偏好,即要求查询机制能在查询效率、准确性和所提供的位置隐私保护强度间实现一种动态平衡.

(2) 候选解集与隐私偏好在实现机理上存在对立

隐私偏好描述查询发起者对自身位置的个性化保护需求,偏好的描述与定义均需在客户端完成;而候选解集必须包括目标查询结果,严重依赖于 LBS 服务器端所存储的 POI 信息(points of interest,用户感兴趣的目标位置等信息).从便于实现的角度考虑,适宜在服务器端完成.于是,出现了二者分别适于在客户端和服务器端实现

的矛盾.

要实现快照查询中查询发起者的隐私偏好与查询服务质量的兼顾,对隐私偏好与候选解集间的关联和制约机理进行分析建模显得至为关键.通过设计有效的调控机制实现查询者的隐私偏好要求,同时尽可能地平抑由此带来的候选解集规模的扩大.

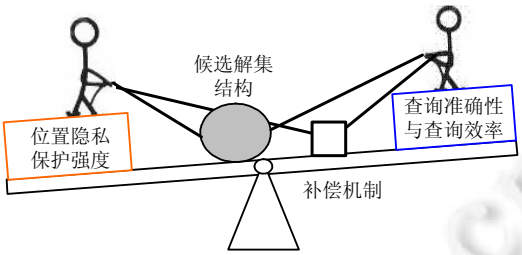



Fig.5 Relation among location protection strength, query accuracy and structure of candidate answer set

图 5 位置隐私保护强度、查询准确性、查询性能与候选解集的关系示意

### 4.3 连续查询中“候选解集拆分攻击”的预防

目前,保护位置隐私连续查询主要关注查询者连续提交的匿名组之间依存关系导致的位置信息泄露的预防,缺少对候选解集可能导致的隐私泄露的关注.连续查询中,候选解集同样需要在服务器与客户端之间传输,存在被攻击者截取发起逆推攻击的风险.

如图 6 所示:假设查询者在 $t_1$ 时刻发起“距其最近加油站”的查询,在向目标移动途中的 $t_2$,$t_3$ 时刻继续发起同样的查询,查询目标结果显然不变,3 次查询对应的候选解集分别表示为 $CAS[t_1]$,$CAS[t_2]$和 $CAS[t_3]$(假设候选解集对应的区域为圆形),即便每次查询均未泄露查询者的位置和查询结果,将 3 个候选解集进行级联分析,攻击者显然能够将目标查询结果锁定在阴影区域内,从而推断出查询发起者的查询目标;进而可以借助查询结果推测查询者的位置及身份信息(例如,借助目标结果与查询者经匿名处理后的位置结构间的集合关系,在目标查询结果处监控等).我们称这种依赖连续候选解间的共性约束发起的攻击为“候选解集拆分攻击”.

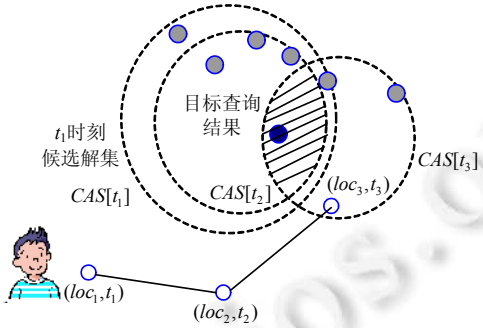



Fig.6 Attacking of intersection inferring to candidate answer sets

图 6 候选解集拆分攻击

在引入隐私偏好的连续查询场景下,候选解集表现出更显著的个性特征,查询者发起同一查询,其候选解集的结构与规模将因偏好而异,这种差异更容易为攻击者发起候选解集拆分攻击提供线索.需要分析查询发起者提交服务器信息结构、候选解集结构、隐私偏好以及拆分攻击成因间的内在制约机理,避免候选解集拆分攻击的发生.

### 4.4 位置隐私保护模式多元性与位置隐私安全度量一元性冲突

位置隐私保护模式的多元性主要表现在保护技术和攻击模式多样化,各种位置隐私保护机制所采用的技术在保护效果上各有偏重,所针对的攻击模式场景迥异.这种多元性导致目前位置隐私安全度量方法的弱通用性.每种保护机制通常结合攻击场景与所采用技术的特点定制其保护强度的衡量方法.具有良好通用性的位置隐私安全度量机制势必要求对各种保护机制提供可比的统一量化度量及解释,表现出一元性.隐私偏好约束的一个重要表现形式是:查询者能够根据自身所处的环境动态设置其对位置隐私保护强度的要求,并选择不同强度的保护机制,这势必要求对不同位置隐藏与查询方法所提供的位置隐私保护效果能够进行量化比较.

需要兼顾不同的攻击场景,分析各类位置保护技术实现位置隐藏方面的共性因素.基于这些共性因素,选取合适的角度构建位置隐私安全量化度量机制.

## 5 研究展望

本节结合保护位置隐私近邻查询中支持隐私偏好面临的主要问题,对隐私偏好及隐私模型建模、位置隐私保护强度量化度量及位置隐藏与查询处理等方面的解决技术和方法进行展望.

### 5.1 隐私偏好与隐私模型构建

查询者的隐私偏好表现为其对位置隐私保护强度、查询效率及查询结果准确性约束的动态调控.首先,隐私偏好模型在形式上应方便查询者表示上述约束;其次,隐私偏好模型的组成元素对位置保护强度、查询效率及查询结果准确性间的内在制约机制有较直接的调控效果.

最小逆推区域(minimum inferred region)是指攻击者借助已掌握的背景知识,能够推测出查询者所在的最小区域范围.对应的区域越大,查询发起者的位置隐私越安全;反之,隐私泄露的可能性越大.同时,攻击者往往借助候选解集的结构分析最小逆推区域,两者间有内在关联,而候选解集规模直接影响到查询服务质量,从最小逆推区域与候选解集的角度描述查询发起者的隐私偏好,具有直观、便于兼顾查询服务质量的优点.目前,保护位置隐私查询领域,最小逆推区域的概念在基于空间混淆技术中应用较多,可以从基于位置干扰、数据变换技术实现位置隐藏内在机理的角度建立最小逆推区域及候选解集模型,从最小逆推区域和候选解集规模的角度建立隐私偏好及位置隐私模型.从最小逆推区域与候选解集的角度建立隐私偏好与位置隐私模型具有如下优点.

(1) 从最小逆推区域与候选解区域的角度描述隐私偏好符合生活惯例,直观,易表示;
(2) 避免对移动对象实时分布信息的依赖,为兼顾隐私偏好与查询性能创造条件;
(3) 便于查询发起者描述其对性能与位置隐私安全性的偏重.通常,候选解的区域越大,查询开销往往越大;最小逆推的区域越大,查询发起者的位置隐私安全性越高.

### 5.2 位置隐私安全强度量化度量机制

位置隐私保护机制需要保护用户的位置信息不被泄露,目前的隐藏技术的主要思想是:对精确位置进行“模糊化”处理,从精确位置到“模糊”位置可以视作将个体元素以某种概率映射到某个集合的过程.“个体元素”与“集合”中某些元素映射的概率越高,个体元素泄露的可能性越大,对应的位置隐私安全性越差;反之,安全性越高.从而可以将位置隐私保护的安全性问题转化为该“集合”的系统稳定问题;另一方面,位置隐私安全不仅决定于所采用隐藏技术的特点,还受攻击场景的影响(如攻击者掌握部分用户的位置分布信息、空间变换参数等),这些成为影响“集合”稳定的条件,可以借助信息熵概念构建基于条件熵的位置隐私安全量化度量机制.

首先限定位置区域范围,以基于空间混淆的隐藏为参照标准,设置合适的单位混淆面积 $s$,确定混淆操作对应集合的域(称为混淆集合域,记作 $M$),进而设置其他各类位置隐藏操作理论映射参数,使其个体元素到“集合”的映射度亦为 $M$(例如,通过合理设置基于 Hilbert 编码的空间变换曲线的 5 个基本参数,使得区域内 Hilbert 格数目为 $M$ 等),以此作为不同隐藏技术保护强度量化可比的基础.进一步引入基本条件熵矩阵模型(如图 7 所示),其中,$O[i]$对应于各类位置隐藏的基本操作(如基于 $s$ 的混淆、数据变换等);$C[j]$对应于攻击者可能掌握的各类背景条件(如局部用户分布等);基本条件熵 $BCE[i,j]$对应于隐藏操作 $O[i]$完成单位位置模糊处理后,基于攻击条件

$C[j]$确定该位置映射实例所需的信息量.

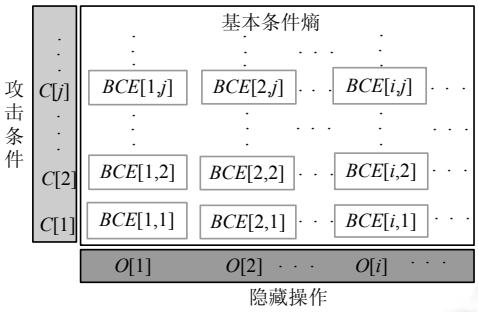



Fig.7 Basic conditional entropy based matrix model

图 7 基本条件熵矩阵模型

两个位置隐藏机制所提供的位置隐私安全性量化度量流程如图 8 所示.

① 设置各类位置隐藏操作的理论映射参数,确保基本条件熵矩阵中其个体元素到“集合”的映射度相同;

② 解析各位置隐私保护机制中的基本隐藏操作,并分析操作间的依赖和组合关系;

③ 分析各隐藏操作实际映射参数与理论映射参数关于隐藏强度的制约关系,生成相应的基本隐藏操作的实际条件熵;

④ 构建各隐藏机制基于实际条件熵的组合条件熵,通过最终条件熵间的对比,实现各位置隐藏机制位置隐私保护强度的量化比较.

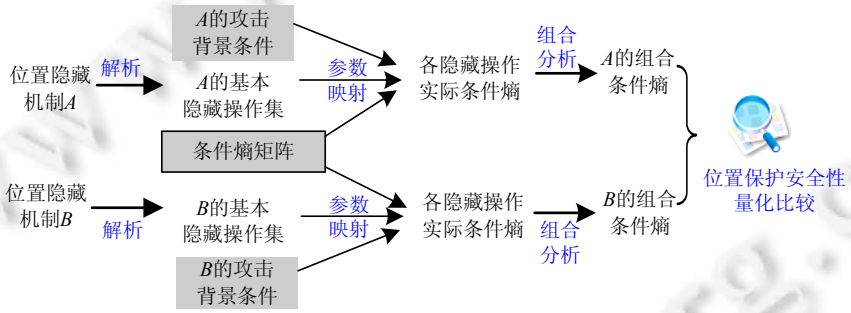



Fig.8 Conditional entropy based quantitative comparison process of location protection strength

图 8 基于条件熵的位置隐私保护机制位置保护强度量化比较流程

### 5.3 支持隐私偏好的保护位置隐私近邻查询技术

在保护位置隐私快照查询领域,查询处理策略和候选解集的结构是攻击者掌握的常见背景知识,攻击者通过分析候选解集的结构与查询处理策略推测查询者的位置范围,查询者指定阈值的最小逆推区域亦需通过特殊查询处理策略和可控候选解集的结构来实现.要在兼顾候选解集与最小逆推区域约束的同时实现查询者的隐私偏好要求,必须从最小逆推区域的逆推原理、候选解集的生成机理角度分析两者的内在联系和制约机理.

已有的保护位置隐私查询方法多数基于前述空间混淆、空间变换、位置干扰中的某一种技术.事实上,3 种技术各有自身的优点及不足,其在位置隐私保护的安全性、位置保护强度的可调控性和效率方面的比较如图 9 和图 10 所示.如图 9 所示:位置干扰技术具有较高的效率,但其所提供的位置隐私保护强度较弱;数据变换技术能够提供较强的位置隐私保护强度,但其效率相对较低;空间混淆技术在处理效率以及位置隐私保护强度方面介于两者之间.在保护强度的可调控性方面,如图 10 所示:空间混淆技术具有较好的位置隐私强度的可调性,但其效率较低且需要在线可信第三方介入隐藏与查询处理过程,严重影响了可用性;数据变换技术通常提供固定

不变的位置隐私保护强度,因此其保护强度的可调控性最弱;位置干扰在保护强度的可调控与查询处理效率方面介于空间混淆与数据变换之间.

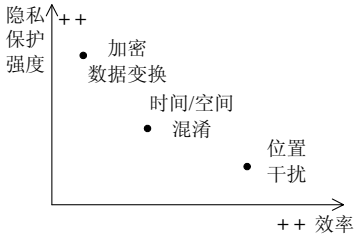



Fig.9 Protection strength/efficiency comparison

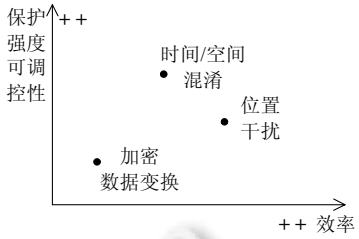



Fig.10 Controllability of protection strength/ efficiency comparison

图 9 保护强度/效率比较

图 10 保护强度可调节性/效率比较

隐私偏好约束要求查询者能够灵活地对位置保护强度、查询效率以及查询准确性的需求进行调控,并且这种调控具有非排它性.例如,查询者要求在兼顾较高的查询效率与查询准确性的前提下,有动态调控位置保护强度的能力.显然,上述 3 种技术中的任何一种都无法单独满足这一要求.结合偏好需求、分析各种技术的特点、综合采用两种或 3 种技术的混合式位置隐藏与查询处理,是一种可行的解决方法,能够充分发挥各种技术的优点,规避存在的缺陷.事实上,3 种技术原理上亦不是完全对立的.例如,空间混淆技术采用查询处理前预先确定目标最小逆推区域的方法实现查询者的位置隐私保护,即,在查询处理前将查询者的位置隐私保护要求固化实现为某个匿名区域,再将该固化区域提交服务器处理;位置干扰技术通过发起假位置查询来避免查询者的位置泄露,查询者的位置仍然可以进行范围界定,只是这种范围较难形式化界定,且属于查询结束之后的后验界定,查询者难以在查询前或查询过程中对后验界定的结果进行预见性干预.从对查询者的位置隐私施加约束的时机角度分析,空间混淆与位置干扰恰恰是两种极端情况,这也是空间混淆与位置干扰技术都难以有效提供隐私偏好支持的内在原因.若能将空间混淆的鲁莽式隐藏介入到位置干扰的过程中,通过对位置干扰过程的假位置选取策略、干预时机以及迭代查询终止条件施加约束,变位置干扰不可预知的无限迭代为有指导、可调控的有限迭代,应当能够在保持查询效能与查询者实施偏好调控的灵活性上找到折衷的调控点.

例如,文献[4]将位置干扰与区域混淆技术相结合,查询者通过向 LBS 服务器发起一轮关于假位置的近邻查询请求,获取服务器端的若干个 POI 位置信息,通过对自身位置与服务器端返回的 POI 位置关系以及查询者对最小逆推区域的约束分析,构建候选解区域模型,并将候选解区域模型提交 LBS 服务器,服务器返回模型区域内的 POI 作为候选解给查询者,供其筛选查询结果,避免了位置干扰技术多轮迭代导致的查询者难以控制查询中间过程以及单纯空间混淆技术固化泛化区域、割裂其与候选解集内在关联性导致的查询处理效能方面的不足.

## 6 总 结

位置服务中的隐私保护是近年来学术界的研究热点之一,本文对保护位置隐私近邻查询中的隐私偏好问题进行综述讨论.

- 首先,对保护位置隐私近邻查询中存在的隐私偏好问题进行了描述;
- 在此基础上,对已有的位置隐藏及近邻查询技术特点进行了介绍,并对现有的位置隐藏与查询策略支持隐私偏好能力进行了分析论述;
- 进一步地,从支持隐私偏好与保护位置隐私查询内在制约机理的角度,分析保护位置隐私近邻查询中支持隐私偏好约束需解决的主要问题;
- 最后,对所归纳问题的可能解决方法进行了展望.

基于位置的近邻查询作为位置服务中的基础性应用,具有广泛的应用前景,对基于位置的近邻查询中隐私

偏好问题的解决,有助于推进位置服务应用的继续深入和服务的安全化、个性化.因此,不论是在理论研究还是在实际应用领域,对位置服务中支持隐私偏好的位置隐藏与近邻查询处理技术进行研究,都具有非常重要的意义.

**References**:


[1] Jiang B, Yao XB. Location-Based services and GIS in perspective. Computers, Environment and Urban Systems, 2006,30(6): 712−725. [doi: 10.1016/j.compenvurbsys.2006.02.003]

[2] Zhou AY, Yang B, Jin CQ, Ma Q. Location based services: Architecture and progress. Chinese Journal of Computers, 2011,34(7): 1155−1171 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2011.01155]

[3] Wang L, Meng XF. Location privacy preservation in big data era: A survey. Ruan Jian Xue Bao/Journal of Software, 2014,25(4): 693−712 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4551.htm [doi: 10.13328/j.cnki.jos.004551]

[4] Ni WW, Zhen JW, Chong ZH. HilAnchor: Location privacy protection in the presence of users' preferences. Lecture Notes in Computer Science, 2011,6897(2):340−352.

[5] Lin X, Li SP, Yang CH. Attacking algorithms against continuous queries in LBS and anonymity measurement. Ruan Jian Xue Bao/ Journal of Software, 2009,20(4):1058−1068 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3428.htm [doi: 10.3724/SP.J.1001.2009.03428]

[6] Chow CY, Mokbel MF. Enabling private continuous queries for revealed user locations. In: Papadias D, Zhang DH, Kollios G, eds. Proc. of the 10th Int'l Symp. on Advances in Spatial and Temporal Databases (SSTD 2007). Berlin, Heidelberg: Springer-Verlag, 2007. 258−275. [doi: 10.1007/978-3-540-73540-3_15]

[7] Pan X, Hao X, Meng XF. Privacy preserving towards continuous query in location based services. Journal of Computer Research and Development, 2010,47(1):121−129 (in Chinese with English abstract).

[8] Palanisamy B, Liu L. Mobimix: Protecting location privacy with mix-zones over road networks. In: Abiteboul S, Böhm K, Koch C, Tan KL, eds. Proc. of the 27th Int'l Conf. on Data Engineering (ICDE 2011). Los Alamitos: IEEE Computer Society, 2011. 494−505. [doi: 10.1109/ICDE.2011.5767898]

[9] Xue J, Liu XY, Yang XC, Wang B. A privacy preserving approach on road network. Chinese Journal of Computers, 2011,34(5): 865−878 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2011.00865]

[10] 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]

[11] Gedik B, Liu L. Protecting location privacy with personalized *k*-anonymity: Architecture and algorithms. IEEE Trans. on Mobile Computing, 2008,7(1):1−18. [doi: 10.1109/TMC.2007.1062]

[12] Chow CY, Mokbel MF, Aref WG. Casper*: Query processing for location services without compromising privacy. ACM Trans. on Database Systems, 2009,34(4):1−45. [doi: 10.1145/1620585.1620591]

[13] Khoshgozaran A, Shahabi C. Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy. In: Papadias D, Zhang DH, Kollios G, eds. Proc. of the 10th Int'l Symp. on Advances in Spatial and Temporal Databases (SSTD 2007). Berlin, Heidelberg: Springer-Verlag, 2007. 239−257. [doi: 10.1007/978-3-540-73540-3_14]

[14] Yiu ML, Jensen CS, Huang XG, Lu H. SpaceTwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services. In: Alonso G, Blakeley JA, Chen ALP, eds. Proc. of the 24th Int'l Conf. on Data Engineering (ICDE 2008). Los Alamitos: IEEE Computer Society, 2008. 366−375. [doi: 10.1109/ICDE.2008.4497445]

[15] Papadopoulos S, Bakiras S, Papadias D. Nearest neighbor search with strong location privacy. Proc. of the VLDB Endowment, 2010,3(1-2):619−629. [doi: 10.14778/1920841.1920920]

[16] Ghinita G, Kalnis P, Khoshgozaran A, Shahabi C, Tan KL. Private queries in location based services: Anonymizers are not necessary. In: Wang JTL, ed. Proc. of the 2008 ACM SIGMOD Int'l Conf. on Management of Data. New York: ACM Press, 2008. 121−132. [doi: 10.1145/1376616.1376631]

[17] Paulet R, Kaosar MG, Yi X, Bertino E. Privacy-Preserving and content-protecting location based queries. In: Kementsietsidis A, Salles MAV, eds. Proc. of the IEEE 28th Int'l Conf. on Data Engineering (ICDE 2012). Los Alamitos: IEEE Computer Society, 2012. 44−53. [doi: 10.1109/ICDE.2012.95]

[18] Lin D, Jensen CS, Zhang R, Xiao L, Lu JH. A moving object index for efficient query processing with peer-wise location privacy. Proc. of the VLDB Endowment, 2011,5(1):37−48. [doi: 10.14778/2047485.2047489]

[19] Pan X, Xu JL, Meng XF. Protecting location privacy against location-dependent attacks in mobile services. IEEE Trans. on Knowledge and Data Engineering, 2012,24(8):1506−1519. [doi: 10.1109/TKDE.2011.105]

[20] Wang T, Liu L. Privacy-Aware mobile services over road networks. Proc. of the VLDB Endowment, 2009,2(1):1042−1053. [doi: 10.14778/1687627.1687745]

[21] Huang Y, Huo Z, Meng XF. CoPrivacy: A collaborative location-preserving method without cloaking region. Chinese Journal of Computers, 2011,34(10):1976−1985 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2011.01976]

[22] Yuan MX, Chen L, Yu PS. Personalized privacy protection in social networks. Proc. of the VLDB Endowment, 2010,4(2):141−150. [doi: 10.14778/1921071.1921080]

[23] Freni D, Vicente CR, Mascetti S, Bettini C, Jensen CS. Preserving location and absence privacy in geo-social networks. In: Huang J, Koudas N, Jones GJF, Wu XD, Collins-Thompson K, An AJ, eds. Proc. of the 19th ACM Conf. on Information and Knowledge Management (CIKM 2010). New York: ACM Press, 2010. 309−318. [doi: 10.1145/1871437.1871480]

[24] Hashem T, Kulik L, Zhang R. Countering overlapping rectangle privacy attack for moving *k*NN queries. Information Systems, 2013,38(3):430−453. [doi: 10.1016/j.is.2012.07.001]

[25] Xu J, Tang X, Hu H, Du J. Privacy-Conscious location-based queries in mobile environments. IEEE Trans. on Parallel and Distributed Systems, 2010,21(3):313−326. [doi: 10.1109/TPDS.2009.65]

[26] Bamba B, Liu L, Pesti P, Wang T. Supporting anonymous location queries in mobile environments with privacy grid. In: Huai J, ed. Proc. of the 17th Int'l Conf. on World Wide Web. New York: ACM Press, 2008. 237−246. [doi: 10.1145/1367497.1367531]

[27] Kido H, Yanagisawa Y, Satoh T. Protection of location privacy using dummies for location-based services. In: Aberer K, Franklin MJ, Nishio S, eds .Proc. of the 21st Int'l Conf. on Data Engineering. Los Alamitos: IEEE Computer Society, 2005. 1248. [doi: 10.1109/ICDE.2005.269]

[28] Yao B, Li FF, Xiao XK. Secure nearest neighbor revisited. In: Jensen CS, Jermaine CM, Zhou XF, eds. Proc. of the 29th IEEE Int'l Conf. on Data Engineering (ICDE 2013). Los Alamitos: IEEE Computer Society, 2013. 733−744. [doi: 10.1109/ICDE.2013.6544870]

[29] Yi X, Paulet R, Bertino E, Varadharajan V. Practical *k* nearest neighbor queries with location privacy. In: Cruz IF, Ferrari E, Tao YF, Bertino E, Trajcevski G, eds. Proc. of the IEEE 30th Int'l Conf. on Data Engineering (ICDE 2014). Los Alamitos: IEEE Computer Society, 2014. 640−651. [doi: 10.1109/ICDE.2014.6816688]

[30] Mascetti S, Bettini C, Wang XS, Freni D, Jajodia S. ProvidentHider: An algorithm to preserve historical *k*-anonymity in LBS. In: Huang JL, ed. Proc. of the 10th Int'l Conf. on Mobile Data Management (MDM 2009). Los Alamitos: IEEE Computer Society, 2009. 172−181. [doi: 10.1109/MDM.2009.28]

[31] Dewri R, Ray I, Ray I, Whitley D. Query *m*-invariance: Preventing query disclosures in continuous location-based services. In: Hara T, Jensen CS, Kumar V, Madria S, Zeinalipour-Yazti D, eds. Proc. of the 11th Int'l Conf. on Mobile Data Management (MDM 2010). Los Alamitos: IEEE Computer Society, 2010. 95−104. [doi: 10.1109/MDM.2010.52]

[32] Elmehdwi Y, Samanthula BK, Jiang W. Secure *k*-nearest neighbor query over encrypted data in outsourced environments. In: Cruz IF, Ferrari E, Tao YF, Bertino E, Trajcevski G, eds. Proc. of the IEEE 30th Int'l Conf. on Data Engineering (ICDE 2014). Los Alamitos: IEEE Computer Society, 2014. 664−675. [doi: 10.1109/ICDE.2014.6816690]

[33] Zhu Q, Zhao T, Wang S. Privacy preservation algorithm for service-oriented information search. Chinese Journal of Computers, 2011,33(8):1315−1323 (in Chinese with English abstract).

[34] Ali ME, Tanin E, Zhang R, Ramamohanarao K. Probabilistic voronoi diagrams for probabilistic moving nearest neighbor queries. Data & Knowledge Engineering, 2012,75(2):1−33. [doi: 10.1016/j.datak.2012.02.001]

[35] Hu HB, Xu JL, Chen Q, Yang ZW. Authenticating location-based services without compromising location privacy. In: Candan KS, Chen Y, Snodgrass RT, Gravano L, Fuxman A, eds. Proc. of the ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD 2012). New York: ACM Press, 2012. 301−312. [doi: 10.1145/2213836.2213871]

[36] Chor B, Goldreich O, Kushilevitz E, Sudan M. Private information retrieval. Journal of the ACM, 1998,45(6):965−981. [doi: 10.1145/293347.293350]

[37] Ni WW, Zhang Y, Huang MF, Chong ZH, Huo YZ. A vector equivalent replacing based privacy-preserving perturbing method. Ruan Jian Xue Bao/Journal of Software, 2012,23(12):3198−3208 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4286.htm [doi: 10.3724/SP.J.1001.2012.04286]

**附中文参考文献**:

[2] 周傲英,杨斌,金澈清,马强.基于位置的服务:架构与进展.计算机学报,2011,34(7):1155−1171. [doi: 10.3724/SP.J.1016.2011.01155]

[3] 王璐,孟小峰.位置大数据隐私保护研究综述.软件学报,2014,25(4):693−712. http://www.jos.org.cn/1000-9825/4551.htm [doi: 10.13328/j.cnki.jos.004551]

[5] 林欣,李善平,杨朝晖.LBS 中连续查询攻击算法及匿名性度量.软件学报,2009,20(4):1058−1068. http://www.jos.org.cn/1000-9825/3428.htm [doi: 10.3724/SP.J.1001.2009.03428]

[7] 潘晓,郝兴,孟小峰.基于位置服务中的连续查询隐私保护研究.计算机研究与发展,2010,47(1):121−129.

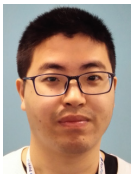
[9] 薛娇,刘向宇,杨晓春,王斌.一种面向公路网的位置隐私保护方法.计算机学报,2011,34(5):865−878. [doi: 10.3724/SP.J.1016.2011.00865]

[21] 黄毅,霍峥,孟小峰.CoPrivacy——一种用户协作无匿名区域的位置隐私保护方法.计算机学报,2011,34(10):1976−1985. [doi: 10.3724/SP.J.1016.2011.01976]

[33] 朱青,赵桐,王珊.面向查询服务的数据隐私保护算法.计算机学报,2010,33(8):1315−1323.

[37] 倪巍伟,张勇,黄茂峰,崇志宏,贺玉芝.一种向量等价置换隐私保护数据干扰方法.软件学报,2012,23(12):3198−3208. http://www.jos.org.cn/1000-9825/4286.htm [doi: 10.3724/SP.J.1001.2012.04286]

**倪巍伟**(1979－),男,江苏淮安人,博士,教授,博士生导师,CCF 会员,主要研究领域为复杂数据管理,数据隐私安全保护.



**陈萧**(1990－),男,硕士生,主要研究领域为数据隐私安全保护.