

轻量级 K 匿名增量近邻查询位置隐私保护算法

陈赛特, 李卫海, 姚远志, 俞能海

(中国科学技术大学网络空间安全学院, 安徽 合肥 230001)

摘要: 基于位置的服务为人们生活带来便利的同时可能暴露用户的位置隐私。在 k 近邻查询问题中, 构造 K 匿名区来保护用户位置隐私的方法具有一定的安全性, 但会带来较大的通信开销。SpaceTwist 方案使用锚点代替真实位置进行 k 近邻查询, 实现简单且通信开销小, 不过其无法保证 K 匿名安全, 也没有给出锚点的具体选取方法。为克服 SpaceTwist 的缺点, 一些方案通过引入可信匿名服务器或者使用用户协作的方式计算用户的 K 匿名组, 进而加强算法的查询结束条件以实现 K 匿名安全; 一些方案结合地图中兴趣点的大致分布, 提出锚点的优选方法, 进一步减小了平均通信开销。基于完善 SpaceTwist 的考虑, 提出轻量级 K 匿名增量近邻 (LKINN, lightweight K -anonymity incremental nearest neighbor) 查询位置隐私保护算法, 借助凸包这一数学工具计算 K 匿名组的关键点集合, 并在此基础上给出一种锚点的选择方法, 能够以更低的响应时间和通信代价实现 K 匿名安全。此外, LKINN 基于混合式位置隐私保护架构, 对系统中的所有成员都只做半可信的安全性假设, 与现有的一些研究成果相比, 降低了安全性假设要求。仿真实验结果表明, LKINN 能够防止半可信用户对正常用户位置隐私的窃取, 保护地图中正常用户的位置隐私, 并且与一些现有方案相比, LKINN 拥有更短的查询响应时间以及更小的通信开销。

关键词: 基于位置的服务; 位置隐私保护; K 匿名; 凸包; 锚点

中图分类号: TP391

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023038

Location privacy protection method based on lightweight K -anonymity incremental nearest neighbor algorithm

CHEN Saite, LI Weihai, YAO Yuanzhi, YU Nenghai

School of Cyber Science and Technology, University of Science and Technology of China, Hefei 230001, China

Abstract: The use of location-based service brings convenience to people's daily lives, but it also raises concerns about users' location privacy. In the k -nearest neighbor query problem, constructing K -anonymizing spatial regions

收稿日期: 2022-05-07; 修回日期: 2022-06-10

通信作者: 李卫海, whli@ustc.edu.cn

基金项目: 国家自然科学基金 (61802357); 国家重点研发计划 (2018YFB0804102)

Foundation Items: The National Natural Science Foundation of China (61802357), The National Key R&D Program of China (2018YFB0804102)

引用格式: 陈赛特, 李卫海, 姚远志, 等. 轻量级 K 匿名增量近邻查询位置隐私保护算法[J]. 网络与信息安全学报, 2023, 9(3): 60-72.

Citation Format: CHEN S T, LI W H, YAO Y Z, et al. Location privacy protection method based on lightweight K -anonymity incremental nearest neighbor algorithm[J]. Chinese Journal of Network and Information Security, 2023, 9(3): 60-72.

is a method used to protect users' location privacy, but it results in a large waste of communication overhead. The SpaceTwist scheme is an alternative method that uses an anchor point instead of the real location to complete the k -nearest neighbor query, which is simple to implement and has less waste of communication overhead. However, it cannot guarantee K -anonymous security, and the specific selection method of the anchor point is not provided. To address these shortcomings in SpaceTwist, some schemes calculate the user's K -anonymity group by introducing a trusted anonymous server or using the way of user collaboration, and then enhance the end condition of the query algorithm to achieve K -anonymity security. Other schemes propose the anchor point optimization method based on the approximate distribution of interest points, which can further reduce the average communication overhead. A lightweight K -anonymity incremental nearest neighbor (LKINN) location privacy protection algorithm was proposed to improve SpaceTwist. LKINN used convex hull mathematical tool to calculate the key points of K -anonymity group, and proposed an anchor selection method based on it, achieving K -anonymity security with low computational and communication costs. LKINN was based on a hybrid location privacy protection architecture, making only semi-trusted security assumptions for all members of the system, which had lax security assumptions compared to some existing research schemes. Simulation results show that LKINN can prevent semi-trusted users from stealing the location privacy of normal users and has smaller query response time and communication overhead compared to some existing schemes.

Keywords: location-based service, location privacy preservation, K -anonymity, convex hull, anchor

0 引言

随着空间定位技术与移动通信技术的发展, 越来越多基于位置的服务 (LBS, location-based service) 出现在人们的日常生活中。常见的 LBS 包括天气推送、路线导航以及 k 近邻 (k NN, k -nearest neighbor) 查询等^[1]。 k NN 查询是 LBS 中最常见的一种形式, 用户首先利用定位技术确定自身的空间位置, 然后将自身的真实位置和查询内容发送给 LBS 提供商以查询最近的 k 个兴趣点 (POI, point of interest), 如“查询距离我当前位置最近的 3 家餐厅”。在获取 LBS 的同时, 用户需要以提供自身真实位置信息为代价, 这使得恶意的服务提供商或攻击者能够从中窃取用户的位置信息。研究表明, 通过收集分析用户的位置隐私信息, 能够推断出更多用户的敏感隐私信息, 如结合时间段信息推测用户的家庭地址以及工作单位^[2], 会带来严重的安全问题。如何在 LBS 中保护用户的隐私信息成为亟待解决的问题。

LBS 中的用户隐私包括查询隐私与位置隐私两类。在查询隐私保护方面, PIR 方案^[3]基于二次剩余定理实现了安全的隐私信息检索技术, 用户向 LBS 提供商提供加密后的查询内容并能够

获得相应的查询结果。由于查询内容与查询结果都是加密的, 只有用户能够解密获得信息, 所以 PIR 方案具有很高的安全性。但是, PIR 方案的计算复杂, 并且效率较低, 无法保证 LBS 中的查询质量。

构造 K 匿名区^[4-7]以及使用虚假位置^[8-18]等方法主要保护用户的位置隐私。构造 K 匿名区方法的主要思想是将真实的用户位置隐藏在包含至少 K 个用户的匿名区域 K -ASR (K -anonymizing spatial region) 中, 使得攻击者无法从中区分出真实的用户位置。LBS 提供商根据 K -ASR 计算所有可能结果并返回给用户, 随后用户从中筛选出精确的查询结果。构造 K 匿名区的方法能实现 K 匿名安全, 具有较高的安全性, 但是返回的查询结果中只有小部分是用户真正需要的, 会造成较大的通信开销。

不同于构造 K 匿名区的方法, 在使用虚假位置的方法中, 用户使用虚假位置作为查询起点来完成查询服务。Yiu 等^[8]提出一种名为 SpaceTwist 的位置隐私保护方案, 用户首先选取一个自身位置以外的其他随机位置作为锚点代替其真实位置向服务器发起查询服务请求, 然后通过增量近邻 (INN, incremental nearest neighbor) 查询算法实现安全的 k NN 查询。与构造 K 匿名区的方法相比,

SpaceTwist 的计算和通信成本都更小,但是会受到基于背景知识的攻击,无法保证 K 匿名安全性,并且没有给出具体的锚点选取方法。Meng 等^[9]基于分布式隐私保护架构提出了 Coprivacy 方案,介绍了一种公用锚点的选取方法,但无法保证 K 匿名安全且需要面临协作用户无法完全信任的问题。Gong 等^[10]基于中心式隐私保护架构提出 KAWCR (K -anonymity without cloaked region) 方案,能够实现 SpaceTwist 的 K 匿名安全,但缺少对锚点选取的讨论。还有一些方案^[15-18]针对 SpaceTwist 查询不均衡的问题提出了改进思路,提高了查询命中率,但通信开销有所增加。

本文基于 SpaceTwist 方案,提出轻量级 K 匿名增量近邻 (LKINN, lightweight K -anonymity incremental nearest neighbor) 查询位置隐私保护算法,在 k NN 查询问题中,能够以更小的响应时间和通信代价(轻量级)实现 K 匿名安全,同时给出一种锚点的选择方法,缓解了锚点的选取困难问题。LKINN 基于混合式位置隐私保护架构,系统由 4 个部分组成:① 用户 (User); ② 位置匿名服务提供商 (LASP, location anonymizer service provider); ③ 位置服务提供商 (LSP, location-based service provider); ④ 匿名协作用户 (AC, anonymous collaborator), 其中所有参与者都是诚实且好奇的,即半可信的。

本文的贡献如下。

1) 通过引入凸包这一数学工具,结合位置变换实现 K 匿名组关键点的计算,降低了查询结束条件,从而实现轻量化的查询,并基于关键点提出一种锚点的选取方法。

2) 基于混合式位置隐私保护架构设计 LKINN 位置隐私保护算法,分离了用户的位置隐私信息与查询隐私信息,其中 LASP 与匿名协作用户都只需要是半可信的,降低了系统的安全性假设。

1 相关工作

本节介绍位置隐私保护中 k NN 查询问题的常见保护方法,主要包括构造 K 匿名区的方法与基于 SpaceTwist 的方法。

1.1 构造 K 匿名区的方法

通过构造 K 匿名区实现 k NN 查询服务的方法

主要分为中心式系统架构和分布式系统架构两种类型。中心式系统架构中引入一个可信第三方作为匿名服务器,并为用户计算 K 匿名区。Gruteser 等^[4]首次将 K 匿名的概念引入位置隐私保护领域中,引入可信匿名服务器不断将地图空间进行 4 等分划分,满足用户 K 匿名需求的最小分块时,即所求 K 匿名区。Chow 等^[5]改进了上述方案,提出了 Casper 方案,服务器根据用户不同的匿名需求维护一个金字塔形数据结构,大大提升了 K 匿名区的计算效率,并且给出一种更精确的基于 K -ASR 的候选 POI 结果计算方法,有效减少了通信开销。

在分布式系统架构方面,Chow 等^[6]提出用户间采用 P2P 的通信方式实现 K 匿名区的计算。从请求用户开始,经过单跳或多跳通信与周围用户协商建立 K 匿名组,再计算得到最终的 K 匿名区。Niu 等^[7]指出上述基于 P2P 方式计算 K 匿名区的方法会遭受基于最小方差的攻击。攻击者可以计算每个用户到其他用户位置距离的方差,而具有最小方差的用户位置极有可能为真实的请求用户位置,为了防止此类攻击,本文提出了一种名为基于随机漫游的匿名组构造算法,真实用户位置不再作为协商过程中每一跳的初始点,而是由其他随机用户位置进行代替。

1.2 基于 SpaceTwist 的方法

增量近邻算法是 SpaceTwist 方案的主体,其查询始末如图 1 所示。

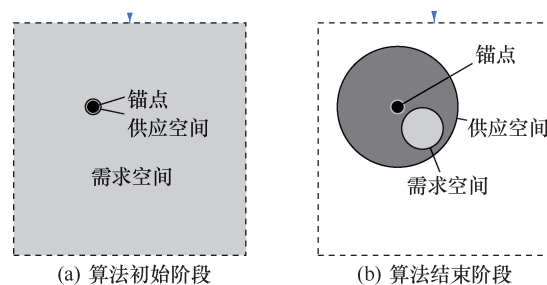


图 1 INN 算法查询始末示意
Figure 1 Initial and final stages of INN algorithm

在 INN 算法中,查询用户挑选随机位置作为锚点,如图 1(a)所示,发送锚点与目标 POI 信息给 LBS 服务器。服务器接收到查询请求后,计算距离锚点最近的几个 POI 查询结果并按照距离从小到大的顺序不断返回给查询用户。锚点与距离

其最远 POI 的距离记为供应空间半径大小, 最新的 k NN 查询结果中距离真实用户位置最远 POI 的距离记为需求空间半径大小, 当供应空间完全覆盖需求空间时, 如图 1(b)所示, 查询结束。选择的锚点、近邻查询数量 k 以及每轮查询返回的 POI 数目 α 都会影响 INN 算法的最终查询结果。

根据文献[11]对 SpaceTwist 方案的分析, 设满足 INN 算法查询过程的可能用户位置构成攻击区域 Ω , 实现 K 匿名安全的 INN 算法即要求攻击区域 Ω 内至少包含 K 个真实用户位置。CoPrivacy^[9]方案采用 P2P 的模式为请求用户构造 K 匿名组, 然后计算 K 匿名组的密度中心点作为该匿名组内所有用户的公共锚点, 保证了每个锚点都有至少 K 个用户在使用。然而, Coprivacy 并不能实现严格意义上的 K 匿名安全, 因为即使使用相同的锚点, 由于实际位置不同, 不同用户与 LBS 服务器之间完成 INN 算法查询后的查询结果可能是不同的, 攻击者能够借此对不同用户实现区分而使得 K 匿名失败。事实上, 如何保证攻击区域 Ω 内包含至少 K 个真实用户位置才是实现 INN 算法 K 匿名安全的关键。P2P 模式下组建 K 匿名组的方法还面临一个不可避免的问题: 协作用户并非完全可信的。针对这一问题, Chen 等^[13]基于博弈论的思想提出了一种非可信用户间的锚点协商方法。

KAWCR^[10]方案引入可信第三方作为匿名服务器为查询用户计算 K 匿名组, 并代替用户向 LBS 服务器发起查询请求, 当查询结果对 K 匿名组中的所有用户都满足 INN 算法结束条件时查询结束。KAWCR 能够实现 INN 算法的 K 匿名安全, 但当用户隐私需求增大时, 判断查询是否结束的计算代价会随之线性增加, 匿名服务器的计算压力大, 容易成为系统瓶颈。Zhou 等^[15]针对 SpaceTwist 方案在某些情况下查询不平衡的问题(查询结果主要围绕锚点), 提出均衡增量近邻(HINN)算法, 加强了算法的查询结束条件, 能够找到更多用户真实位置附近的兴趣点, 通信开销有所增加, 不过无法实现 K 匿名安全。Sun 等^[17]基于使用双锚点的思想提出 DATiwt 方案, 在 HINN 算法的基础上减小了一定的通信开销, 并实现了 K 匿名安全, Zhou 等^[16]在路网环境下完

成了相似了工作。除此之外, Liu 等^[14]结合路网分布密度, 提出一种结合锚点优选的 SpaceTwist 方案, 提高了查询准确度并且能够实现 K 匿名安全保护。Zhou 等^[12]通过引入可信匿名服务器, 将路网交叉口作为锚点, 实现了路网环境下的连续 k NN 查询。

2 预备知识

本节介绍凸包的相关知识以及凸包的计算方法——Graham 扫描法。

2.1 凸包与相关引理

定义 1 给定一个平面上的非线性点集 X , 凸包就是其最外围的点连接起来所构成的凸多边形, 它能包围点集中所有的点。

引理 1 对于一个平面上的非线性点集 $X = \{x_1, x_2, \dots, x_n\}$, 设其凸包为 $S = \{s_1, s_2, \dots, s_m\}$, 取该平面上任意两点 F_1, F_2 , 对任意 $x_i \in X$, 存在 $s_j \in S$, 使得构成的线段满足

$$F_1 x_i + F_2 x_i \leq F_1 s_j + F_2 s_j \quad (1)$$

证明 若 $x_i \in S$, 存在 $s_j = x_i$ 使得不等式(1)成立。若 $x_i \notin S$, 由定义 1 可得, 点 x_i 在凸包 S 所构成的多边形内部。对于不等式(1)左边, 记焦点为 F_1, F_2 , 到两焦点距离和为 $F_1 x_i + F_2 x_i$ 的椭圆为 E_i 。同样地, 对不等式(1)右边, 记焦点为 F_1, F_2 , 到两焦点距离和为 $F_1 s_j + F_2 s_j$ 的椭圆为 E_j 。假设 S 内的所有点 s_j 均不满足不等式(1), 可知 S 内的所有点均在椭圆 E_i 中(不包括边界)。因为 S 是 X 的凸包, 所以 X 内的所有点也都严格含于椭圆 E_i , 这与点 x_i 在椭圆 E_i 边界上这一条件矛盾! 故假设不成立, 即必然存在 $s_j \in S$, 使得不等式(1)成立。

综合得, 引理 1 成立。

引理 2 对于一个平面上的非线性点集 X , 设其凸包为 $S = \{s_1, s_2, \dots, s_m\}$, 记 X 的最小覆盖圆为 R , 圆心为点 o 。对 S 内所有点 s_i , 连接 os_i 并做延长线交圆 R 于点 t_i , 记点集 $T = \{t_1, t_2, \dots, t_m\}$, 则 $X \cup T$ 的凸包为 T 。

证明 由定义 1, $X \cup T$ 的凸包等价于 $S \cup T$ 的凸包, 即需证 $S \cup T$ 的凸包为 T 。

对任意与点 o 不构成线性关系的两点 s_i, s_j , 其与圆心 o 构成三角形 $\Delta os_i s_j$, 都存在对应两点 t_i, t_j , 使其与圆心 o 构成的三角形 $\Delta ot_i t_j$ 包含三角形 $\Delta os_i s_j$, 由定义 1, $S \cup T$ 的凸包即 T , 证毕。

引理 3 在一个平面上, 随机独立选取不同的 4 个点, 记其最小覆盖圆为圆 R , 那么这 4 个点都在圆 R 圆周上的概率为 0。

证明 若 4 个点中存在三点共线, 则仅有两个或 3 个点可能在圆周上, 引理 3 成立。

若 4 个点均不构成线性关系, 则必可找到其中 3 个点确定圆 R , 并使第 4 个点位于圆上或圆内。因为每个点都是随机独立选取的, 所以第 4 个点在圆 R 圆周上的概率为圆 R 的周长 / 平面上部分区域面积 = 0。

综合得, 引理 3 成立。

引理 4 在一个正方形平面区域中, 设有满足随机均匀分布的 n 个点, 它们构成点集 N , 则点集 N 的凸包所包含的点的个数期望为 $O(\ln n)$ 。

Sariel 在文献[19]中给出了详细证明。

2.2 Graham 扫描法

设 n 为待计算点集 $X = \{x_1, x_2, \dots, x_n\}$ 中点的个数, 则 Graham 扫描法的计算复杂度为 $O(n \ln n)$, 其具体步骤如下。

1) 找到点集中纵坐标最小的点, 不妨设为点 x_1 , 如果有多个, 则选择其中横坐标最小的点。点 x_1 一定是凸包上的点。

2) 以点 x_1 为原点重新计算其他点的直角坐标, 然后为各点计算相应的极坐标, 并按照极角的大小对各点进行逆时针排序, 具有相同极角的点中只保留具有最大极径的点。将排序后各点的直角坐标依次保存在列表 $list$ 中, 不妨记为 $x_2 - x_1$ 。点 x_2 也一定是凸包上的点。

3) 从 $list$ 中取出 x_2 , 并将点 x_1, x_2 入栈 $stack$ 。

4) 取出 $list$ 的首元素 x_i , 记 $stack$ 栈顶两元素为 a, b (a 在 b 上), 计算叉积 $c = (b - a)(x_i - a)$, 若 c 小于 0, 则将点 x_i 入栈; 否则, 将栈顶元素 a 出栈, 并把元素 x_i 重新放回 $list$ 首位。

5) 判断 $list$ 是否为空, 若不为空, 则进入步骤 4); 若为空, 则循环结束, 还原 $stack$ 内各元素的原直角坐标, 即所求凸包结果。

3 系统模型与符号定义

3.1 系统模型

LKINN 方案系统架构模型如图 2 所示, 系统由查询用户、匿名服务提供商、匿名协作用户和位置服务提供商构成。

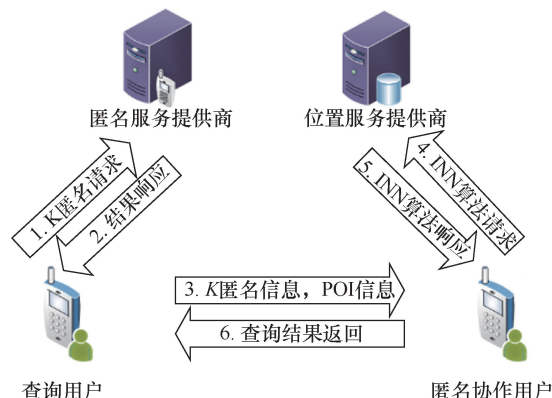


图 2 LKINN 方案系统架构模型
Figure 2 System architecture model of LKINN scheme

为实现 kNN 查询服务, 查询用户首先向 LASP 发送自身位置信息并发起 K 匿名请求, LASP 返回响应结果, 然后查询用户将 POI 信息以及接收到的 K 匿名信息等打包发送给随机的匿名协作用户, 匿名协作用户代替查询用户与位置服务提供商完成 INN 算法查询, 最后将所有查询结果返回给查询用户并由其完成最终的结果求精工作。

1) LASP 拥有所有用户的位置分布信息并且实时更新, 能够为查询用户实现最优 K 匿名组的计算, 返回的 K 匿名信息中仅包含关键点信息 I 而非全部 K 个用户的位置信息。LASP 不知道查询用户所查询的 POI 信息, 只知道查询用户的位置隐私信息。

2) 匿名协作用户接收查询用户的 K 匿名信息和 POI 信息等, 知道查询用户的查询隐私信息, 但无法确定查询用户的真实位置。关键点为真实用户位置的概率接近于零, 匿名协作用户仅能获得少量映射后的关键点信息 I 。

3) LSP 支持 INN 算法查询, 由于引入匿名协作用户作为查询代理, 能实现匿名查询, 有效防止关联攻击。即使 LSP 能实施基于背景知识的攻击, LKINN 算法也是具有 K 匿名安全性的。

系统内的所有成员都是半可信的，会诚实地完成自身任务，但同时对其他用户的隐私感兴趣。

3.2 符号定义

本文中相关符号定义如表 1 所示。

表 1 符号定义
Table 1 Symbol definition

符号表示	符号含义
q	用户位置
q'	锚点位置
$\text{dis}(a,b)$	a,b 两点距离
K	匿名需求数量
POI	兴趣点
k	近邻查询数量
uid	用户标志符
I	关键点集合
α	LSP 每次返回的查询结果数目

4 轻量级 K 匿名增量近邻查询算法

4.1 系统流程设计

LKINN 方案的具体流程如下。

1) 记查询内容为 $(\text{uid}, q, K, \text{POI}, k, \alpha)$ ，查询用户先发送 (uid, q, K) 到 LASP 进行 K 匿名请求。

2) LASP 接收到 (uid, q, K) 请求后，首先，根据 uid 和位置 q 更新用户的位置信息，如果是第一次收到 uid，则在用户列表中加入该 uid 并将位置 q 直接加入地图信息中，否则直接更新用户的位置信息；然后，LASP 根据匿名需求数量为其计算 K 匿名组，进而计算关键点集合 I 与锚点 q' ，最后将 (I, q') 返回给查询用户。

3) 查询用户接收到 (I, q') 后，通过 P2P 模式在网络中随机选择一位匿名协作用户，将 $(\text{uid}, I, q', \text{POI}, k, \alpha)$ 发送给该匿名协作用户。

4) 接收到上述消息之后，匿名协作用户发送 (q', POI, α) 给 LSP 发起 INN 算法查询请求。

5) LSP 响应 INN 算法查询，向匿名协作用户返回查询结果，直到匿名协作用户根据关键点集合 I 判定查询结束。

6) 匿名协作用户将所有候选查询结果打包返回给查询用户。

7) 查询用户根据自身真实位置从查询结果中筛选出精确的 k NN 查询结果。

4.2 关键点与锚点计算

本节介绍 LASP 端的关键点与锚点计算算法。关键点信息 I 能辅助实现 INN 算法的 K 匿名安全，锚点计算能够解决 INN 算法的输入参数选取问题。

上文提到，在 KAWCR 方案中，可信匿名服务器为查询用户计算用户 K 匿名组，并当 K 匿名组内的所有用户位置都满足 INN 算法查询结束条件时完成 k NN 查询，能够实现 INN 算法的 K 匿名化。为了减少查询结束条件的判断次数，减少相应的存储与计算代价，本文方案为用户 K 匿名组计算关键点信息 I ，只关键点满足查询结束条件即可保证 K 匿名组中的所有用户位置都满足查询结束条件。在关键点结果的基础上，给出锚点的计算算法，该锚点选取算法具有全局最优性。

首先，给出用户 K 匿名组的计算。LASP 拥有整体的用户位置分布信息，并且能够实时更新，为了提高空间数据的检索效率，引入 R 树^[20]为其构造数据结构。R 树是用来存储高维数据的平衡树，不必遍历所有数据即可获得答案，显著提高查询效率。此外，R 树能够实现动态插入、删除、更新等操作，便于用户动态管理。初始创建完成用户位置信息的 R 树后，后续只需维护更新现有 R 树即可。LASP 构造的 R 树更倾向于为查询用户计算 K 匿名组，并不关心具体的位置关系。因此，在构造 R 树的过程中，每个非叶子节点处将额外记录当前包含的所有用户位置数量 nums 。

算法 1 给出了 K 匿名组的具体计算过程。具体 K 匿名组计算中，LASP 根据查询用户的 uid 和位置 q 更新 R 树，回溯位置 q 的父节点。当某父节点的 nums 第一次不小于 K 时，回溯结束（第 3~5 行），该父节点包含至少 K 个用户位置。为了对抗最小方差攻击，从 nums 个用户位置中随机挑选包含位置 q 的 K 个用户位置（第 6 行），构成用户 K 匿名组，而非从中选取 $K-1$ 个距离位置 q 最近的用户位置。

算法 1 K 匿名组计算

输入 R 树，用户标识符 uid，用户位置 q ，匿名需求 K

输出 用户 K 匿名组 X

1) 根据 uid 和 q 更新 R 树

- 2) node ← q 所在叶子节点
- 3) **while** node.num < K
- 4) node ← node 的父节点;
- 5) **end while**
- 6) $X \leftarrow$ node 中包含 q 的随机 K 个用户位置
- 7) 返回 K 匿名组结果 X
- 8) **END**

得到 K 匿名组结果 X 后, LASP 继续计算 X 的凸包。不妨记 $X = \{x_1, x_2, \dots, x_K\}$, 记 X 的凸包为 $S = \{s_1, s_2, \dots, s_m\}$, 其中 $x_i (i=1, 2, \dots, K)$, $s_j (j=1, 2, \dots, m)$ 均表示用户的位置。在 INN 算法查询中, 记锚点为 q' , 对于任意用户位置 x_i , 在查询之前, kNN 结果可能为地图中的任意位置, 设其中距离 x_i 最远的 POI 位置为 p_i , 代入引理 1, 其中 F_1, F_2 对应 p_i, q' , 可知对 X 内的任意 x_i , 在 S 内都存在一个 s_j , 使得不等式

$$p_i x_i + q' x_i \leq p_i s_j + q' s_j \quad (2)$$

成立。设用户位置 s_j 的 kNN 结果中距离 s_j 最远的 POI 位置为 p_j , 同理有不等式

$$p_j x_i + q' x_i \leq p_j s_j + q' s_j \quad (3)$$

成立。

若 $p_i x_i \leq p_j x_i$, 由式(2)、式(3)可得

$$p_i x_i + q' x_i \leq p_j s_j + q' s_j \quad (4)$$

若 $p_i x_i > p_j x_i$, 则 p_j 在以点 x_i 为圆心, $\text{dis}(p_i, x_i)$ 为半径的圆内, 地图空间中存在一点 p_h , 使得 $p_h x_i = p_i x_i$ 且 $p_h s_j \leq p_j s_j$ 。再次使用引理 1 结论, 有不等式

$$p_h x_i + q' x_i \leq p_h s_j + q' s_j \quad (5)$$

成立。结合点 p_h 满足的条件, 可得不等式(4)成立。

综合可得, 不等式(4)成立。

设用户位置 s_j 查询结束时的供应空间半径为 r_j , 由 INN 算法性质, 供应空间大小完全覆盖需求空间大小, 可得不等式

$$p_j s_j + q' s_j \leq r_j \quad (6)$$

成立。结合不等式(4)有

$$p_i x_i + q' x_i \leq r_j \quad (7)$$

因为系统内成员都是半可信的, 除 LSP 外, 发起查询的用户以及匿名协作用户都可能对地图内的其他用户位置分布感兴趣。若 LASP 直接将 K 匿名组的凸包结果返回给查询用户, 或者查询用户将 K 匿名组的凸包结果发送给匿名协作用户, 由于凸包结果包含的都是真实的用户位置信息, 恶意的查询用户能够通过提交多次假位置查询的方式, 恶意的匿名协作用户能够通过多次匿名协作的方式, 不断地获取地图内其他用户的真实位置信息, 造成群体用户隐私泄露。为了避免此类安全问题, 需要继续处理上述凸包结果, 在保证实现算法 K 匿名的前提下, 不包含或尽量少包含真实的用户位置。

关键点与锚点计算如算法 2 所示。计算得到 K 匿名组的凸包结果后 (第 1~17 行), LASP 继续计算该凸包的最小覆盖圆, 将其表示为圆 R , 并设其圆心为点 o , 半径为 r 。对凸包内的每个点 s_j , 连接 os_j 并延长交圆 R 于点 t_j , 使得点 s_j 在圆心点 o 与点 t_j 之间, 若点 s_j 本身就在圆 R 圆周上, 则点 t_j 即点 s_j (第 18~24 行)。由引理 2 可知, 所有点 t_j 构成的凸多边形能够覆盖所有点 s_j , 使用点 t_j 代替点 s_j 依然能保证算法的 K 匿名安全, 称这些点 t_j 为关键点, 其构成的点集称为关键点集合, 记为 I 。至此, LASP 完成关键点与锚点的计算。

算法 2 关键点与锚点计算

输入 K 匿名组结果 $X = \{x_1, x_2, \dots, x_K\}$

输出 关键点集合 I , 锚点 q'

- 1) 找出 X 中纵坐标最小的点 x ;
- 2) **for** x_i in X
- 3) $x_i \leftarrow x_i - x$;
- 4) 计算 x_i 相应的极坐标;
- 5) **end for**
- 6) 根据对应极坐标极角的大小按逆时针对 X 中的 x_i 进行排序;
- 7) $\text{stack} \leftarrow \emptyset; d \leftarrow 3$;
- 8) x_1, x_2 入栈 stack ;
- 9) **while** $d < \text{size of } X$
- 10) $a, b \leftarrow$ 栈顶两元素;
- 11) $c \leftarrow (x_1 - a) \times (b - a)$ // 计算叉积

```

12) if  $c < 0$ 
13)  $x_1, x_2$  入栈 stack;
14)  $d \leftarrow d + 1$ ;
15) else  $a$  出栈;
16) end if
17) end while
18)  $R \leftarrow$  stack 的最小覆盖圆;
19)  $q' \leftarrow R$  的圆心;  $r \leftarrow R$  的半径;  $I \leftarrow \emptyset$ ;
20) for point $i$  in stack
21)  $t_i \leftarrow \text{point}_i * r / \text{dis}(q', \text{point}_i)$ ;
22)  $t_i \leftarrow t_i + x$ ;
23) 将  $t_i$  加入  $I$  中;
24) end for
25) 返回关键点集合  $I$ , 锚点  $q'$ ;
26) END

```

4.3 LKINN 算法

本节综合关键点与锚点计算算法以及 INN 算法, 给出完整的 LKINN 算法, 具体如算法 3 所示。

算法 3 LKINN 算法

输入 用户标志符 uid, 用户位置 q , 匿名需求 K , 兴趣点 POI, 近邻查询数量 k , LSP 每次返回的查询结果数目 α

输出 k NN 查询结果

```

1) LASP 根据 uid,  $q$ ,  $K$  计算  $I$ ,  $q'$ ;
2) 匿名协作用户  $u_r$  获取参数 uid,  $I$ ,  $q'$ , POI,  $k$ ,  $\alpha$ ;
3)  $u_r$  发送  $q'$ , POI,  $\alpha$  到 LSP 发起  $k$ NN 查询请求;
4) INN 算法阶段:
5) flag  $\leftarrow$  true; result  $\leftarrow \emptyset$ ;
6) while flag
7) LSP 以增量形式向  $u_r$  返回最新查询结果 POI;
8)  $u_r$  将 POI 加入 result 中;
9) flag  $\leftarrow$  false;
10)  $r_s \leftarrow$  供应空间半径;
11) for  $t_i$  in  $I$ 
12)  $r_d \leftarrow$  需求空间半径;
13) if  $r_d + \text{dis}(q', t_i) > r_s$ 
14) flag  $\leftarrow$  true; // 查询尚未结束

```

```

15) end if
16) end for
17) end while
18) INN 算法结束;
19)  $u_r$  根据 uid 将 result 返回给原查询用户;
20) 用户从 result 中计算得准确的  $k$  近邻结果;
21) END

```

4.4 算法分析

(1) 安全性分析

LASP 知晓用户的真实位置, 但无法获取用户的兴趣点信息, 无法基于仅知的用户位置信息对用户的其他隐私做出推断。匿名协作用户知道用户的兴趣点信息, 但无法获取用户的真实位置信息, 也不能关联用户的位置隐私及查询隐私从而实现隐私推断攻击。对于 LSP, 即使能够实施背景知识攻击 (知道地图中整体的用户位置分布), 在具体算法部分已经证明, LKINN 仍是 K 匿名安全的。并且, 由匿名协作用户代为查询, 使得查询具有匿名性, LSP 无法记录用户 id 进而实施关联攻击。因为匿名协作用户是查询用户随机挑选的, LASP 无法确定具体匿名协作用户实施共谋攻击; 与 LASP 共谋的匿名协作用户被查询用户选中的概率很低。若 LSP 与匿名协作用户共谋, LSP 能获取关键点集合 I , 下面分析 I 的安全性。

假设用户随机均匀分布在一个正方形的地图区域中, 根据引理 4, 得到 K 匿名组的凸包所包含的位置点数期望为 $O(\ln K)$ 。对引理 3 做简单推广, 则 K 匿名组凸包内的点在其最小覆盖圆圆周上的数目为 2 或 3, 设其期望为 w ($2 < w < 3$), 则关键点为用户位置点的期望概率可表示为 $w / O(\ln K)$ 。对于关键点集合 I , 其包含查询用户真实位置的概率仅为 w/K , 关键点为用户位置的概率为 $w / O(\ln K)$, 查询用户、匿名协作用户或者与之共谋的 LSP 都难以判断某关键点是否为用户位置, 因而不能获取地图中其他用户的位置分布信息。

(2) 锚点分析

当 INN 算法查询结束时, K 匿名组凸包内的任意点都满足供应空间覆盖需求空间这一条件。对凸包内的任意点, 需求空间是以该点为圆心的

圆, 因此最终供应空间会严格覆盖凸包内的所有点。又因为圆 R 为凸包的最小覆盖圆, 所以最终供应空间半径一定大于 r 。无法确定所有点最终的需求空间大小, 选择圆心点 o 作为 INN 算法中的锚点是合适的。虽然它可能不是某些查询中的最佳锚点, 但在全局查询中能够表现出计算的均衡性。

(3) 方案比较

将本文方法 (LKINN) 和其他方法的性能进行比较, 见表 2。

表 2 性能比较
Table 2 Performance comparisons

方法	保护架构	K 匿名保护	安全性假设
SpaceTwist ^[8]	独立式架构	无	无
KAWCR ^[10]	中心式架构	完全	可信匿名服务器
CoPrivacy ^[9]	分布式架构	不完全	可信协作用户
HINN ^[15]	中心式架构	无	可信匿名服务器
DATwist ^[17]	分布式架构	完全	可信协作用户
LKINN	混合式架构	完全	半可信 LASP 与 AC

与 SpaceTwist 和 HINN 相比, 本文方法能够实现 K 匿名保护。与 KAWCR、CoPrivacy 以及 DATwist 相比, 本文方法对匿名服务器与协作用户的安全性假设降低, 虽然同时引入了匿名服务器与协作用户, 但匿名功能由匿名服务器实现, 无须像 CoPrivacy 和 DATwist 一样进行烦琐的 P2P 通信, 在网络中选取一位匿名协作用户即可。

5 实验结果

本节将 SpaceTwist^[6]、KAWCR^[8]、Casper^[5] 以及 LKINN (本文方案) 分别在合成数据集 UI 以及真实地理数据集 CN 上进行了仿真实验比较。实验结果表明, LKINN 方案拥有更小的 INN 算法查询时间响应以及更小的通信代价。

5.1 实验设置

搭建实验设定混合式结构通信网络, 算法采用 Python 实现, 在 Intel 六核 2.90 GHz CPU、8GB 内存的 Windows10 平台上运行。实验采用的数据集为合成数据集 UI 与真实地理数据集 CN。

数据集 UI 由随机均匀分布的点构成, 数据集 CN 从中国地理位置信息中抽取获得, 包含

321 106 个 POI。地图大小设置为 10 000 m, 用户数量为 200 000 个, 用户位置分布满足随机均匀分布。参数默认值见表 3。

表 3 参数默认值
Table 3 Default values of parameters

参数	默认值
K	40
k	10
α	67
用户数量	200 000 个
POI 数量	1 000 000 个

根据文献[11]的分析, 网络上最大传输单元 (MTU) 的典型大小为 576 byte, 表示一个坐标位置信息需要 8 byte, 并且每个数据包都有一个 40 byte 的标头, 可以得到最佳的每次查询返回的 POI 数目取值 $\alpha = (576 - 40) / 8 = 67$ 。因为仅返回最近邻查询结果, 无法满足许多真实场景下的用户查询需求, 如用户想要查询距离自己最近的几个餐厅, 并从中挑选出评分最高的一家, k 的默认值取 10。对于匿名需求参数 K , 本文认为当 K 取值不超过 20 时, 用户的匿名隐私需求较小; 当 $K \in (20, 60]$ 时, 用户的匿名隐私需求一般; 当 $K \in (60, 100]$ 时, 用户的匿名隐私需求较强, 实验中 K 的默认参数选择为 40, 表示用户具有一般的匿名需求。

在每组对比结果中, 统一进行 50 次模拟实验并取其平均值。

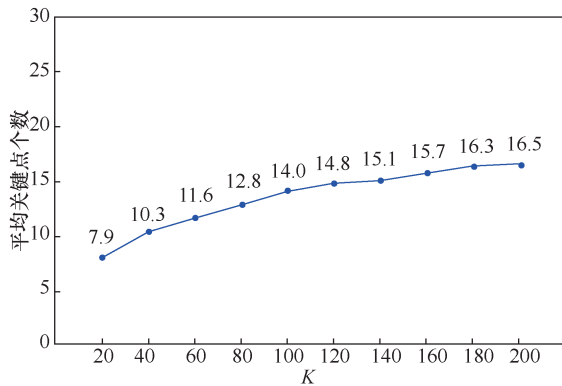
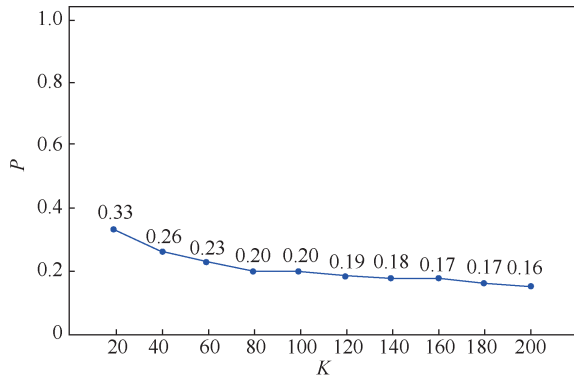
5.2 算法安全性

KAWCR 以及 Casper 都需要引入可信第三方作为匿名服务器, LKINN 只需引入半可信的 LASP, 安全性假设更为宽松。下面考察 K 值对关键点集合 I 安全性的影响。

平均关键点个数与 K 值的关系如图 3 所示, 其中 K 的取值为 20~200, 步长为 20。曲线走势与对数函数 $\ln K$ 类似, 上文给出了关键点个数的理论值为 $O(\ln K)$, 事实上, 关键点个数与 $\ln K$ 的比值关系取决于具体的用户位置分布关系。

相同 K 值范围下, 继续讨论关键点为用户位置的概率 P 与 K 的关系。如图 4 所示, P 随 K 的增加而不断减小, 这说明 K 越大对地图内

用户位置信息的保护效果越好。 P 的理论取值为 $w/O(\ln K)$ ，当 K 值超过一定值时（图中为 120）， P 值变化趋于平缓，此处的 K 值对应较强的查询用户匿名需求，能够较好地实现地图中其他真实用户的位置信息保护。当 K 值在区间 $[40, 60]$ 内时，对应一般的查询用户匿名需求，关键点为用户位置的概率为 23%~26%，仍能一定程度上保护其他用户的位置信息。

图3 平均关键点个数与 K 的关系Figure 3 The relationship between the average number of critical points and K 图4 P 与 K 的关系Figure 4 The relationship between P and K

5.3 响应时间对比

SpaceTwist、KAWCR 与本文方案都基于 INN 算法实现 k NN 结果查询，并且 KAWCR 和本文方案都引入匿名服务器实现匿名模块功能，本节比较 SpaceTwist、KAWCR 与本文方案在查询端的平均处理时间，即判断查询是否结束的响应时间。由于 SpaceTwist 与 KAWCR 均未给出锚点的具体选取方法，为了比较公平性，SpaceTwist 与 KAWCR 中选取的用户位置与锚点的距离 $\text{dis}(q, q')$ 与本文方案中的一致。

图 5(a)、图 5(b)分别展示在 UI 和 CN 数据集上，当 K 取值为 10~100，步长为 10，其他参数取默认值时，SpaceTwist、KAWCR 和本文方案在 INN 算法查询过程中响应时间的结果对比。可以看到，本文方案的响应时间比 KAWCR 更小，并且随着 K 值的增大，本文方案的响应时间增幅远小于 KAWCR。事实上，当 K 足够大时，本文方案的理论响应时间为 KAWCR 的 $O(\ln K)/K$ 。本方案的响应时间略大于 SpaceTwist 方案，这是实现 K 匿名保护而额外带来的时间响应开销。另外，在相同匿名需求参数下，CN 数据集上的响应时间大于 UI 数据集上的响应时间，这是由真实地理位置分布不均匀所导致的。

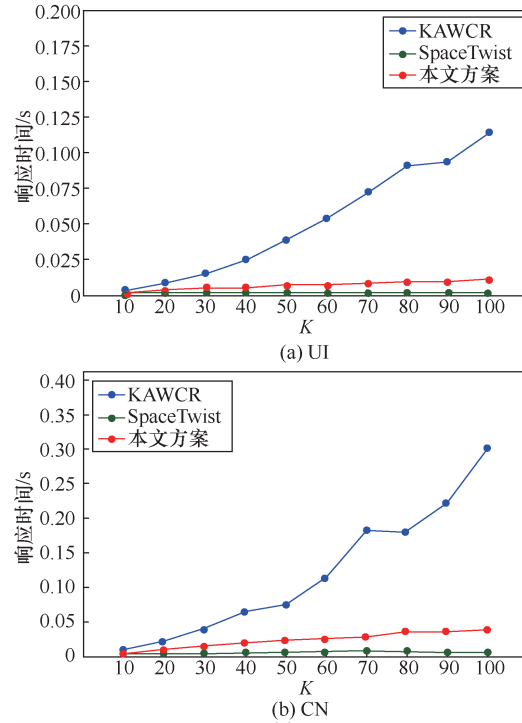
图5 不同 K 值下响应时间的比较 (step=10)Figure 5 Comparisons of response time at different K value (step=10)

图 6(a)、图 6(b)分别展示在 UI 和 CN 数据集上，当 K 取值为 2~20，步长为 2，其他参数取默认值时，SpaceTwist、KAWCR 和本文方案在 INN 算法查询过程中响应时间的结果对比。可以看到， K 值的变化对 SpaceTwist、KAWCR 以及本文方案的响应时间并无太大影响，并且本文方案的响应时间都小于 KAWCR。

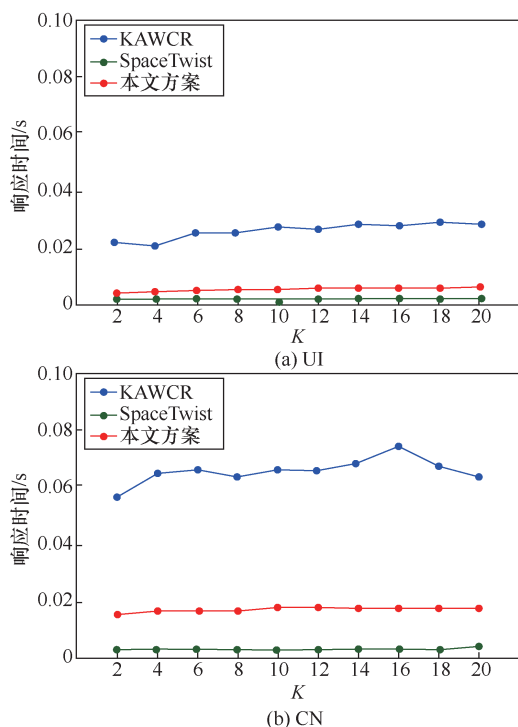


图6 不同 K 值下响应时间的比较 (step=2)
Figure 6 Comparisons of response time at different K value (step=2)

5.4 通信开销对比

通信开销定义为 LSP 向查询端返回的所有候选 POI 结果的数目。本文额外引入 Casper 方案进行实验比较, 与 KAWCR 以及本文方案相同, Casper 引入第三方服务器计算实现匿名模块功能, 不同的是, 其最终采用基于 K 匿名区的方法实现候选 POI 的计算。

本节实验对比 SpaceTwist、KAWCR、Casper 与本文方案的通信开销。由于 INN 算法的特殊性, 除 Casper 方案以外的其余方案中返回的最终数据包大小均为 α 的整数倍, 而对 Casper 的最终数据包大小规格不做要求。

图 7(a)、图 7(b) 分别展示在 UI 和 CN 数据集上, 当 K 取值为 10~100, 步长为 10, 其他参数取默认值时, SpaceTwist、KAWCR、Casper 和本文方案的通信开销对比。可以看到, 本文方案的通信开销比 Casper 更小, 这是因为采用 SpaceTwist 框架使得 kNN 查询过程具有反馈性, 查询端根据返回的最新数据包判断查询是否结束并反馈给 LSP, 减少了不必要的 POI 结果的传输, 因此通信开销更小。此外, 在 CN 数据集实验结果下, Casper 方案的波动性较大,

用户匿名需求的变化可能会带来不稳定的服务质量。

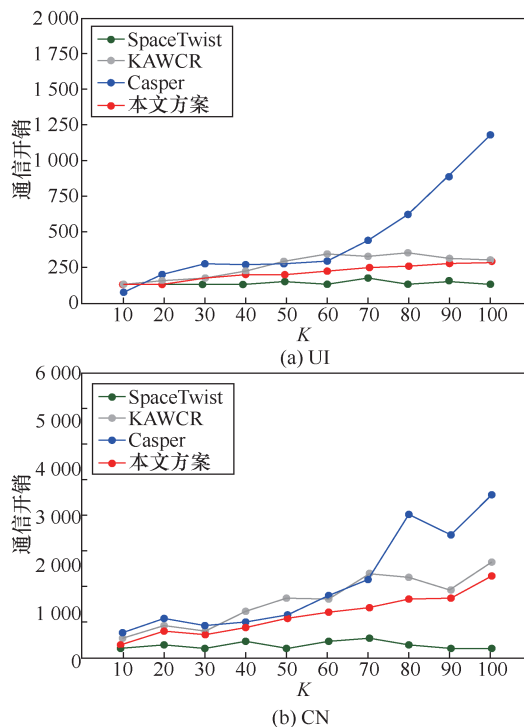


图7 不同 K 值下通信开销的比较
Figure 7 Comparison of communication overhead at different K value

与 KAWCR 相比, 本文方案的通信开销略小, 这是因为本文方案选取了具有全局计算均衡性的锚点, 能够在整体用户查询方面实现最优, 从而降低平均通信开销。

6 结束语

本文针对 SpaceTwist 无法实现 K 匿名以及基于构造 K 匿名区的方案通信开销大的缺点, 提出一种轻量级 K 匿名增量近邻查询算法。结合凸包这一数学概念, 在 KAWCR 方案基础上实现了轻量级的时间响应与略小的通信开销, 并且较构造 K 匿名区的方案具有更小的通信开销。此外, 本文方案基于混合式位置隐私保护架构, 降低了系统中所有参与者的安全性假设要求, 同时使得位置匿名服务器摆脱了系统瓶颈角色, 更好地利用了整个系统中参与者的算力。本文在不同数据集上对比了本文方案与 SpaceTwist、KAWCR 以及 Casper, 实验结果表明, 本文方案在降低安全性假设的条件下, 仍拥有更短的算法查询响应时间

以及更低的通信开销。

然而, 本文方案中的 K 匿名组计算算法仍有待完善, 如何构造高效的 R 树是主要难题。此外, INN 算法仅适用于欧几里得距离, 无法直接应用在现实场景的路网结构中, 结合路网信息改进本文方案亦是未来研究的方向。

参考文献:

- [1] 万盛, 牛犇, 李风华, 等. 位置隐私保护技术研究进展[J]. 通信学报, 2016, 37(12):124-141.
WAN S, NIU B, LI F H, et al. Research progress on location privacy-preserving techniques[J]. Journal on Communications, 2016, 37(12):124-141.
- [2] KRUMM J. Inference attacks on location tracks[C]//5th Springer International Conference on Pervasive Computing. 2007:127-143.
- [3] PAULET R, KAOSAR M G, YI X, et al. Privacy-preserving and content-protecting location-based queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(5):1200-1210.
- [4] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//1st ACM International Conference on Mobile Systems, Applications, and Services (MobiSys'03). 2003: 31-42.
- [5] CHOW C Y, MOKBEL M F, AREF W G. Casper*: query processing for location services without compromising privacy[J]. ACM Transactions on Database Systems, 2009, 34(4):1-48.
- [6] CHOW C Y, MOKBEL M F, LIU X. A peer-to-peer spatial cloaking algorithm for anonymous location-based services[C]//14th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS'06). 2006: 171-178.
- [7] NIU B, ZHU X Y, LI Q H, et al. A novel attack to spatial cloaking schemes in location-based services[J]. Future Generation Computer Systems, 2015, 49(1):125-132.
- [8] YIU M L, JENSEN C S, MØLLER J, et al. SpaceTwist: managing the trade-offs among location privacy, query performance, and query accuracy in mobile services[C]//Proceedings of the 24th International Conference on Management of Data. 2000: 121-132.
- [9] 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[J]. 计算机学报, 2011, 34(10): 1976-1985.
HUANG Y, HUO Z, MENG X F. CoPrivacy: a collaborative location privacy preserving method without cloaking region[J]. Chinese Journal of Computers, 2011, 34(10): 1976-1985.
- [10] GONG Z, SUN G, XIE X. Protecting privacy in location-based service using k -anonymity without cloaked region[C]// Proceeding of the International Conference on Mobile Data Management (MDM'10). 2010: 366-371.
- [11] YIU M L, JENSEN C S, MØLLER J, et al. Design and analysis of a ranking approach to private location-based services[J]. ACM Transactions on Database Systems, 2011, 36(2): 1-42.
- [12] 周长利, 陈永红, 田辉, 等. 保护位置隐私和查询内容隐私的路网 K 近邻查询方法[J]. 软件学报, 2020, 31(2): 471-492.
ZHOU C L, CHEN Y H, TIAN H, et al. Location privacy and query privacy preserving method for K -nearest neighbor query in road networks[J]. Journal of Software, 2020, 31(2): 471-492.
- [13] 陈玉凤, 刘学军, 李斌. 基于博弈论的用户相互协作的位置隐私保护方法[J]. 计算机科学, 2013, 40(10):92-97.
CHEN Y F, LIU X J, LI B. Collaborative position privacy protection method based on game theory[J]. Computer Science, 2013, 40(10): 92-97.
- [14] 刘振鹏, 赵璇, 董亚伟, 等. 结合锚点优选算法改进的 SpaceTwist 隐私保护方法[J]. 通信学报, 2017, 38(Z1): 32-38.
LIU Z P, ZHAO X, DONG Y W, et al. Improved SpaceTwist privacy protection method based on anchor optimization algorithm[J]. Journal on Communications, 2017, 38(Z1):32-38.
- [15] 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 on Improved SpaceTwist privacy protection method based on sensitive location diversity[J]. Journal on Communications, 2015, 36(4): 129-140.
- [16] 周宇航, 柳毅. 路网环境下基于双锚点的位置隐私保护方法[J]. 计算机应用研究, 2022, 37(11): 3392-3395+3400.
ZHOU Y H, LIU Y. Location privacy protection method based on double anchor point in road network environment[J]. Application Research of Computers, 2022, 37(11): 3392-3395+3400.
- [17] SUN R Z, XIANG G L, GAO Q W, et al. A study of location privacy protection about k -interesting points query based on double anchors[C]//Ubiquitous Positioning, Indoor Navigation and Location-Based Services. 2018.
- [18] LIU H Y, ZHANG S W, LI M L, et al. An effecton location privacy-preseving k -anonymity scheme in location based services[C]//IEEE International Conference on Electronic Technology, Communication and Information(ICETCI). 2021.
- [19] SARIEL H. On the expected complexity of random convex hulls[EB].
- [20] GUTTMAN A. R-trees: a dynamic index structure for spatial searching[R]. 1984, 14.

[作者简介]



陈赛特（1996—），男，浙江温州人，中国科学技术大学研究生，主要研究方向为网络安全与隐私保护。

李卫海（1975—），男，辽宁大连人，中国科学技术大学副教授，主要研究方向包括多媒体内容安全、数据安全。

姚远志（1989—），男，安徽望江人，博士，中国科学技术大学副研究员，主要研究方向为信息隐藏、数据安全和视频编码。



俞能海（1964—），男，安徽无为，人，中国科学技术大学教授、博士生导师，主要研究方向包括视频处理与多媒体通信、信息检索、媒体内容安全、数据安全。