一种基于兴趣点分布的匿名框 *K*NN 查询方法

朱顺痣[1],黄 亮[2],周长利[3],马 樱[1]
(1. 厦门理工学院计算机与信息工程学院,福建厦门 361024;2. 国家计算机网络应急技术处理协调中心,北京 100029;
3. 华侨大学计算机科学与技术学院,福建厦门 361021)

摘 要: 针对利用匿名框实现的兴趣点 *K* 近邻(*K*NN)查询带来的通信开销大、时延长等问题,提出了基于单一兴趣点 Voronoi 图划分和四叉树层次化组织的 *K*NN 查询方法. 该方法根据兴趣点层次信息有针对性的构造查询匿名框用来获取详细查询信息,在保护位置隐私的同时,降低了查询通信开销,同时注入虚假查询保护了用户的真实查询内容隐私. 最后分别采用模拟地理数据和真实地理数据进行理论分析和有效性验证.

关键词: 位置隐私; 基于位置的服务; 匿名框; *K* 近邻查询
中图分类号: TP311 **文献标识码:** A **文章编号:** 0372-2112 (2016)10-2423-09
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2016.10.021

A Privacy-Preserving Method Based on PoIs Distribution Using Cloaking Region for *K* Nearest Neighbor Query

ZHU Shun-zhi[1], HUANG Liang[2], ZHOU Chang-li[3], MA Ying[1]
(1. *School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian* 361024, *China*;
2. *National Computer Network Emergency Response Technical Team Coordination Center of China, Beijing* 100029, *China*;
3. *School of Computer Science and Technology, Huaqiao University, Xiamen, Fujian* 361021, *China*)

Abstract: Achieving *K*NN query with traditional cloaking region brings higher communication cost and delay caused by useless points of interest (PoI) returned, a new *K*NN query method is proposed. Based on Voronoi diagram division of PoIs and hierarchical index quadtree structure, cloaking region can be constructed purposefully. Due to the targeted query request, the communication cost is decreasing compared with traditional cloaking region methods. And injecting fake query requests makes the query content privacy preserving work. We have verified the effectiveness of our proposal by analysis and experiments.

Key words: location privacy; location based service; cloaking region; *K* nearest neighbor query

1 引言

基于位置的服务(Location-Based Service, LBS)推动了移动智能终端各类型应用的快速发展. 基于位置的查询服务是目前广泛被用户使用的服务类型之一,用户利用携带的智能终端获取所需的查询服务,典型的两种查询方式分别为 *K* 近邻(*K* Nearest Neighbor, *K*NN)查询和 *R* 范围查询[1~3]. 然而,用户获取查询结果前必须提供自身位置信息,位置越精确查询结果质量越高,但用户的隐私也越容易暴露. 移动用户隐私可以分为位置隐私和查询内容隐私两类[4~8],在查询过程中,理想查询状态是既不让其它任何实体知道其所在位置,也不让其它实体知道他的查询内容. 位置 k 匿名[9~13]是实现位置隐私保护的有效方法之一,该方法构造包含 k 个不同用户位置的匿名框,用匿名框代替全部用户实际位置发起查询. 对于查询内容隐私,用户利用匿名框查询时也需保证查询请求内容不少于 l 种[14],避免查询内容隐私直接泄露给如 LBS 服务器等其它实体.

然而,匿名框查询需要返回匿名框内及其周围的多个兴趣点作为候选集合供用户选择,这会增大数据通信开销. 这是因为上述传统的构造匿名框查询方法并没有考虑兴趣点的分布情况. 如图 1 所示,用户构造

收稿日期:2015-04-14;修回日期:2015-11-28;责任编辑:孙瑶
基金项目:国家自然科学基金(No. 61373147, No. 61502404);福建省自然科学基金(No. 2016Y0079, No. 2015J05132);福建省教育厅 A 类项目(No. JA14234)

了一个1NN查询匿名框(虚线矩形框),根据该匿名框LBS返回给用户 u_k 的兴趣点不仅包括 P_3,还有 P_1、P_2、P_4 和 P_5,这些兴趣点都可能是用户查询的目标.

Voronoi图[15]是一种平面空间目标点之间的中垂线分割方法,如图1实线划分所示,利用该图可以表示空间目标位置远近关系. 位于兴趣点 P_3 划分单元(实线凸多边形)内的用户 u_k 一定距离 P_3 最近. 因此,在上述1NN查询中,如果用户知道兴趣点的Voronoi划分情况,则将目标兴趣点 P_3 置于匿名框内,并要求LBS服务器只查询返回匿名框内兴趣点描述信息,因此无需返回传统匿名框查询要返回的其他兴趣点 P_1、P_2、P_4 和 P_5 的描述信息.

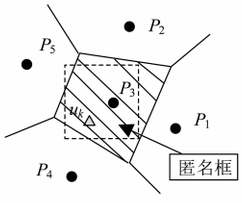


图1 查询匿名框构造

文献[16]提出了能够解决上述问题的一种方案2PASS,然而该方法在扩展到*K*NN查询过程中,由于将同一地图中多类不同兴趣点共同进行Voronoi图划分,使得用户需要对比更多的其他类型兴趣点来找到最近的 K 个目标兴趣点. 因此,本文提出单一兴趣点Voronoi划分方法,并利用四叉树组织该图,便于用户快速找到*K*NN目标兴趣点并构造查询匿名框,该方法能够在降低通信开销的同时,保护用户位置隐私和查询内容隐私.

另一方面,用户在提交查询请求时为了保证身份信息与位置隐私的不可关联,通常使用假名来替代用户真实身份. 史敏仪等[17]针对文献[18]中存在的假名失效的缺陷提出了敏感区域移动用户的位置隐私保护方法. Palanisamy B[19]和 Freudiger J[20]分别针对文献[18]中存在的缺陷提出了相应的假名使用方法. Mano[21]等提出了一种动态假名更换方案,用来保护移动用户的轨迹隐私. Yu等[22]提出了一种扩展假名更换区域MixGroup,显著提高了假名在保护用户位置隐私时的效能. Lin[23]在其著作中详细论述了基于密码技术的假名更换方法,用以有效保护用户位置隐私.

2 预备知识

给定 n 个兴趣点,可以利用Voronoi图将其划分成 n 个单元,邻近单元间没有重叠区域,位于划分单元内的用户与该兴趣点位置最近. 如图2所示,我们给每个顶点 P_i 赋予权值 (L_i, C_i),L_i 代表位置坐标,C_i 代表兴趣点的类型,如医院、加油站等.

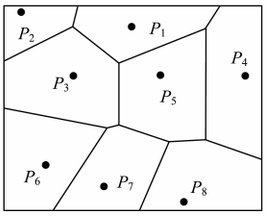


图2 Voronoi空间划分

本文方法采用"用户-服务器"架构[20],该架构用户获取LBS服务流程如图3所示的四步.

如图3所示,用户首先请求轻量化兴趣点分布信息,然后根据返回的信息构造合适的匿名框,将要查询的目标兴趣点包含在匿名框内发起查询请求. 用户通过第3步控制匿名框内兴趣点数量,从而控制第4步返回消息数量.

图3 系统架构及流程

3 兴趣点 *K*NN 查询

对于兴趣点的1NN查询,用户只需要根据兴趣点分布信息即可快速构建匿名框,然而对于兴趣点的*K*NN查询则用户需要比对较多的其他兴趣点,K 值越大比对的额外兴趣点数量就会越多. 解决该问题的方法就是构造同类兴趣点Voronoi图划分,并组织成轻量化、层次化的结构发送给用户,便于用户快速查找到 K 个目标兴趣点及其分布情况,这个工作是由LBS服务器来完成的.

3.1 四叉索引树

对地理区域进行网格划分是描述兴趣点位置的有效方法之一,其中金字塔形的组织形式能够体现平面空间区域间的层次化结构[24]. 如图4所示,从全局地图出发,将地理空间每次划分成四个网格,如此迭代划分下去得到 4^{h-1} 个网格(h 是划分深度),直至满足一定的粒度,并建立层次化索引表. 图4在全局地图上将同类兴趣点(如加油站)利用Voronoi图划分为4×4网格.

由此,每个兴趣点唯一属于一个网格,在网格划分粒度不同时,一个网格可能会包含多个兴趣点. 通常情况下,在粒度较小时每个兴趣点的划分单元可能覆盖多个网格,这意味着处于不同网格内的用户,可能处于同一个Voronoi划分单元内,即在不同网格内的用户可

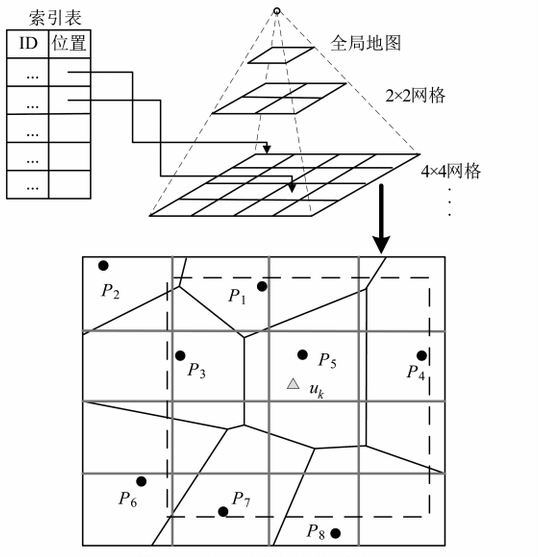


图4　金字塔层次划分示意

能距离同一个兴趣点最近.

这样,每个网格都可以用和该网格有重叠部分的 Voronoi 划分单元来表示,以便处于该网格的用户能够知道距离自己最近的兴趣点是哪个.

由此,以金字塔形状组织的兴趣点可以表示成是一个四叉树型的索引结构,每层网格以顺时针方向为编号顺序,如图 5 描述了图 4 的四叉树结构,树的叶子结点就是最底层网格,每个底层网格用与之有重叠区域的 Voronoi 划分单元的兴趣点来描述,便于处在该网格内用户掌握兴趣点分布情况后构造查询匿名框. 上述每类兴趣点四叉索引树可以用 Tree_{C_i} 表示,其中 C_i 表示兴趣点类型.

每一层的每个网格都用其左上角坐标(x_{lt},y_{lt})和右下角坐标(x_{rb},y_{rb})表示,用户确定自己所在最底层网格(即所在叶子结点),则执行如下算法 1,搜索四叉树.

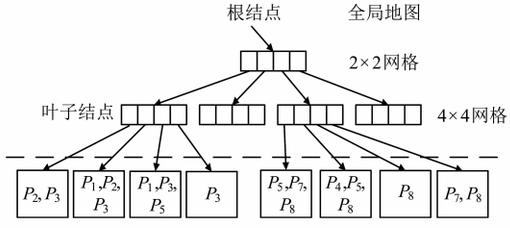


图5　兴趣点分布四叉树

算法 1　用户所在结点查询算法(用户端)

输入:以 T_i 为根结点的地图四叉树 Tree_{C_i},用户位置坐标 loc_{u_k},叶子结点网格面积 S_{leaf}

输出:四叉树 Tree_{C_i} 的某个结点 n_i

1. Procedure begins:
2. 获取根结点 T 的每个孩子结点 n_i
3. if 孩子结点 n_i 的面积 $> S_{\mathrm{leaf}}$ then
4. for 每个孩子结点 $t_i \in T$ do
5. 由孩子结点范围坐标(x_{lt},y_{lt})和(x_{rb},y_{rb})找到 loc_{u_k} 所在网格及表示该网格的结点 n_i
6. $\mathrm{Tree}_{C_i} \leftarrow n_i$ 为根结点的子树
7. 调用算法 1,参数为$(\mathrm{Tree}_{C_i},\mathrm{loc}_{u_k},S_{\mathrm{leaf}})$
8. return n_i
9. End Procedure

该算法以迭代的方式不断缩小用户所在区域网格的面积,确定用户所在最底层网格,即四叉树的叶子结点,算法采用自调用方式来实现,当全局地图面积为 n 时,设算法的执行时间为 $T(n)$,算法 1 第 6 行语句的执行时间为 $O(1)$,而算法 1 第 7 行递归调用语句的执行时间为 $T(\frac{n}{4})+C$,则 $T(n)=T(\frac{n}{4})+C=T(\frac{n}{16})+2C=\cdots$,其中 C 为常数,由迭代法解递归方程可知,直到 $\frac{n}{4^i}(i\in N)$ 的面积小于等于 S_{leaf} 时结束,由此可得递归方程的解为 $T(n)=O(n)$,即该递归算法的时间复杂度为 $O(n)$. 同理可得利用递归工作栈实现算法 1 的空间复杂度为 $O(n)$.

3.2　*K*NN 查询流程

对于每一类兴趣点,LBS 服务器都为之构造一个上一小节描述的兴趣点分布四叉索引树,用户在第 1 步请求周围兴趣点分布信息时,在真实目标兴趣点请求中注入虚假兴趣点查询请求,比如用户目标兴趣点是查询周围最近的 K 个加油站,同时注入餐厅、医院兴趣点分布情况查询作为虚假请求. 表 1 详细说明了图 3 中 KNN 查询完整流程. 请求格式为$(u_{k1}\parallel CT\parallel \mathrm{sig}_{\mathrm{key}_u}(CT)\parallel E_{\mathrm{key}_{\mathrm{lbs}}}(Q_{u_k}))$,其中 u_{k1} 是用户的一个假名,CT 为当前时间,$\mathrm{sig}_{\mathrm{key}_u}(CT)$ 表示用户利用私钥 key_u 做的签名,以便 LBS 服务器来验证其合法性,$E_{\mathrm{key}_{\mathrm{lbs}}}(Q_{u_k})$ 表示用户以 LBS 服务器的公钥 $\mathrm{key}_{\mathrm{lbs}}$ 加密查询请求 Q_{u_k} 发送给 LBS 服务器,$Q_{u_k}=(C_1,C_2,\cdots,C_3)$ 主要为查询兴趣点类型 C_i,Q_{u_k} 中有一个是用户真正要查询的目标兴趣点.

LBS 收到用户请求后,验证签名 $\mathrm{sig}_{\mathrm{key}_u}(CT)$ 的合法性,据此解密出查询请求 Q_{u_k},并返回 Q_{u_k} 中所有兴趣点类型 C_i 的四叉索引树,返回信息可表示为$(\mathrm{LBS}_{\mathrm{id}}\parallel \mathrm{Tree}_{C_1},\mathrm{Tree}_{C_2},\cdots,\mathrm{Tree}_{C_n})$,其中 Tree_{C_i} 为用户目标兴趣点的四叉索引树.

用户根据 LBS 服务器返回信息,找到自己目标兴趣点的四叉索引树 Tree_{C_i},执行算法 1 逐层搜索直至找到自己所在的网格,即叶子结点,根据叶子结点中存储的兴趣点权值 $P_i:(L_i,C_i)$ 信息找到距离自己最近的兴趣点 P_k,并以此为出发点根据 3.3 小节中定义 1 和定理

1,找到距离 P_k 最近的其他 $K-1$ 个兴趣点,并构造包含这 K 个兴趣点的匿名框.

最后,用户构造包含这 K 个兴趣点的匿名框,并更换假名 u_{k2} 后将满足用户需求的匿名框等信息($u_{k2} \| CT \| \mathrm{sig}_{\mathrm{key}_u}(CT) \| E_{\mathrm{key}_{\mathrm{lbs}}}(CR_{u_k} \| Q'_{u_k})$)发送给 LBS 服务器,其中 CR 为匿名矩形框,Q'_{u_k} 为 CR 内被查询兴趣点类型. LBS 服务器根据 Q_{u_k} 返回匿名框内兴趣点的描述信息 $R=((C_1,R_1),(C_2,R_2),\cdots,(C_n,R_n))$ 作为查询结果候选集,其中 R_1 为 C_1 类型兴趣点的具体描述信息,用户最终获得 K 个目标兴趣点的详细描述信息.

表 1 KNN 查询流程表

User(u_k)	LBS Server
1. 请求周围兴趣点信息阶段 选择假名 u_{k1} $(u_k \| CT \| \mathrm{sig}_{\mathrm{key}_u}(CT) \| E_{\mathrm{key}_{\mathrm{lbs}}}(Q_{u_k}))$ →	
	2. 返回兴趣点描述信息阶段 验证 $\mathrm{sig}_{\mathrm{key}_u}(CT)$ 的合法性 解密 $E_{\mathrm{key}_{\mathrm{lbs}}}(Q_{u_k})$ $(\mathrm{LBS}_{\mathrm{id}} \| \mathrm{Tree}_{C_1},\mathrm{Tree}_{C_2},\cdots,\mathrm{Tree}_{C_n})$ ←
3. 构造 KNN 查询匿名框阶段 找到目标兴趣点四叉树 Tree_{C_i} 执行算法 1 找到用户所在网格 n_i 找到网格内最近邻兴趣点 P_i 以 P_i 为起点,找到 KNN 兴趣点 执行算法 2,构造匿名框 CR_{u_k} $(u_{k2} \| CT \| \mathrm{sig}_{\mathrm{key}_u}(CT) \| E_{\mathrm{key}_{\mathrm{lbs}}}(CR_{u_k} \| Q_{u_k}))$ →	
	4. 匿名框内兴趣点描述信息 找到 CR_{u_k} 内 Q_{u_k} 指定类型兴趣点描述信息集 R $R=((C_1,R_1),(C_2,R_2),\cdots,(C_n,R_n))$ ←
5. 获取 K 个目标兴趣点的详细描述信息 在 R 中找到目标兴趣点的描述信息	

用户根据算法 1 返回的用户所在网格结点 n_i 来构造匿名框,算法如算法 2 所示.

算法 2 K 近邻兴趣点查找及匿名框生成算法(用户端)

输入:算法 1 返回的用户所在网格结点 n_i,KNN 查询请求
输出:包含 K 个目标兴趣点在内的匿名框 CR_{u_k}

```
1.  Procedure begins;
2.  for 每个 P_i ∈ n_i do  //依据图 7 找到 n_i 内兴趣点
3.      Array1←计算 dist(u_k, P_i)
4.  p←min{Array}  //用户在最近 P_i 的划分单元内
5.  heap←AN_1(p)// 参见定义 1;  小顶堆 heap
6.  Array2←AN_1(p)
7.  if |heap| < k then
8.      Array3←Array2
9.      for 每个 P_j ∈ Array3 do  //下一阶兴趣点
10.         heap←AN_1(P_j)
11.         Array4←AN_1(P_j)
12.     清空 Array2 //保存新找到的下一阶兴趣点
13.     Array2←Array4
14. neighbor←在 heap 中找到前 K 个兴趣点
15. 构造包含 neighbor 中全部兴趣点位置的矩形匿名框 CR_{u_k}
16. return CR_{u_k}
17. End Procedure
```

算法 2 的主要思想是以用户所在 Voronoi 图划分单元的兴趣点为起始点逐阶搜索下一级兴趣点,直至找到距离用户最近的 K 个兴趣点. 图 4 中的虚线框为根据算法 2 构造的 5NN 匿名框. 用户在提交查询时,不仅给出要查询的匿名框内目标兴趣点,同时还要求返回几类虚假查询兴趣点的描述信息,保护用户查询内容隐私. 在算法 2 的执行过程中,执行频度最高的语句为第 3、10 和 11 行,其时间复杂度为 $O(n)$. 显然,空间复杂度也取决于问题规模 n,为 $O(n)$.

3.3 性能分析

(1)查询准确性

用户根据自身位置确定所在网格(即算法 1),然后根据所在网格找到所在兴趣点 Voronoi 图划分单元(即找到距离自己最近的兴趣点),并由此划分单元逐步扩展到邻近划分单元确定最近的 K 个兴趣点(即算法 2)的过程中,需要保证两个算法的正确性.

首先,在算法 1 中,用户根据图 5 的 LBS 服务器返回的网格划分信息及轻量兴趣点分布信息,结合自身位置不断缩小自己所在网格,直到满足阈值 S_{leaf} 要求. 由于在该过程中,用户通过某一四叉树的每个网格左上角坐标(x_{lt},y_{lt})和右下角坐标(x_{rb},y_{rb})运算,算法 1 第 5 行可以精确计算出用户所在四个网格中的一个,然后以该网格为出发点迭代继续确定所在下一级网格,此过程相对简单.

然后,在算法 2 中,用户根据图 5 所示的网格和兴趣点 Voronoi 图层叠关系找到所在网格 n_i 中的所有兴趣点,确定距离自己最近的目标兴趣点,并以此兴趣点为生成元的 Voronoi 图划分单元 $V(P_i)$ 为出发点,找到其他 $K-1$ 个目标兴趣点. 在此过程中用户根据如下定义 1 和定理 1 可以确定周围 1 阶邻近兴趣点集合,如果不足 K 个,则在 2 阶邻近兴趣点集合继续找,直到找到 K 个兴趣点.

定义 1 一般地,给定 Voronoi 图内任意兴趣点 P_i,如果其 1 阶近邻兴趣点集合可以表示为:$AN_1(P_i)=\{P_k \mid V(P_i)$ 与 $V(P_k)$ 有公共边$\}$,那么 P_i 的 $n(n\geqslant 2)$ 阶

近邻兴趣点集为:$AN_n(P_i)=\{P_k \mid V(P_j)$ 与 $V(P_k)$ 有公共边,$P_j \in AN_{n-1}(P_i)\}$.

定理 1[25]　Voronoi 图内任意划分单元 $V(P_i)$ 内任意点 q 的第 K 个近邻兴趣点一定包含在集合 $AN_1(P_i) \cup AN_1(P_i) \cup \cdots \cup AN_{k-1}(P_i)$ 中.

定理 2[25]　Voronoi 图内,每个划分单元平均有 6 个 1 阶邻近划分单元.

这个过程还存在一个问题,根据定理 1,用户在找 K 个兴趣点的过程中,需要把 $V(P_i)$ 的 K 阶邻近兴趣点都找到(算法 2 第 5 ~ 14 行),这会带来一定时延. 通常情况下,由于多数用户设置 KNN 兴趣点查询时,K 通常取值为 1 ~ 30,由定理 2 可知,当 $V(P_i)$ 的邻近单元及下一阶邻近单元数均约为 6 个时,则满足 $6^n \geqslant 30$ 的最小正整数 n 为 2,即查找用户最近的 K 个目标兴趣点,最少可在其所在划分单元 $V(P_i)$ 的 2 阶近邻兴趣点集合中找到.

综上,用户在可以根据算法 1 和算法 2 确定自己所在网格和所在兴趣点划分单元,并在下限值为 2 阶的邻近划分单元中找到 K 个目标兴趣点,由于是逐层搜索邻近兴趣点,因此可以在保证较少通信量的同时获得正确的 K 个兴趣点查询结果. 然后构造包含这 K 个兴趣点的匿名框,发送给 LBS 服务器,LBS 服务器返回匿名框中兴趣点的详细描述信息,因此查询是有目标的,有效降低通信量.

(2)隐私安全性

和传统匿名框方法实现的 k 匿名性不同,本方法用户不必考虑匿名框中用户数量. 这是因为用户在构造匿名框时,可能不在匿名框内. 当 $K=3$ 时,图 6 中的虚线匿名框可能是处于匿名框外用户 u_k 构造的,因为匿名框只需覆盖距离 P_5 最近的其他 2 个兴趣点 P_1、P_3 即可,不必考虑用户位置.

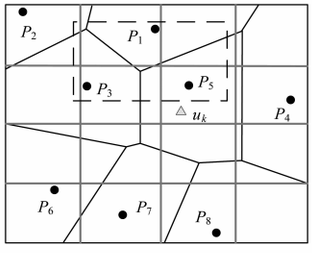


图6　匿名框

因此本方法的 k 匿名性取决于匿名框覆盖目标兴趣点 Voronoi 图划分单元内的用户数量,假设在没有注入虚假查询的情况下,则查询 Q_i 来自该匿名框覆盖 CR_k 的某个划分单元 $V(P_i) \in CR_k$ 的概率可以表示为 $p_{Q_i \to V(P_i)}$,则构造该匿名框发起一次查询 Q_i 的信息熵可以用式(1)表示:

$$H(Q_i) = -\sum_{V(P_i) \in CR_k} p_{Q_i \to V(P_i)} \times \log_2(p_{Q_i \to V(P_i)}) \quad (1)$$

则处于划分单元 $V(P_i)$ 内的用户 $u_k \in V(P_i)$ 是提出该 KNN 查询 Q_i 的用户的联合信息熵可以用式(2)表示:

$$H(Q_i,u_k) = -\sum_{u_k \in V(P_i)} \sum_{V(P_i) \in CR_k} \left(p_{Q_i \to V(P_i)} \times \log_2(p_{Q_i \to V(P_i)}) \right) \quad (2)$$

在现有文献中通常认为背景知识包括用户分布情况、用户提交的兴趣点查询请求集合及各类兴趣点四叉索引树. 由式(1)和(2)可知,即使攻击者在掌握一定的背景知识情况下,能够根据构造的匿名框推断出用户真实查询内容,也很难将查询内容对应到划分单元内某个具体用户身份上.

4　实验

本实验在 Windows 7 平台采用 JAVA 语言实现上述算法,硬件环境为 3.2GHz Intel Core i5 处理器,内存大小为 4GB. 实验地图采用美国地名委员会提供的地理数据集 GDS①,同时采用 Thomas Brinkhoff 路网移动节点数据生成器生成的模拟数据集 TBS②,数据集中包含 15 万条路网环境下的用户轨迹数据,用户均匀分布在每条路网上,每个用户的起点和终点均随机选取,行进速度受路段限制. LBS 服务器与用户采用 3G 网络通信带宽为 2Mbps,实际每次返回单个数据包为 128 字节,若每个兴趣点信息占用 8 个字节,除去 40 字节的包头,每个数据包包含兴趣点个数为 (128 - 40)/8 = 11 个. 当某一实验参数变化时,其他参数配置均采用默认值配置. 本实验参数取值范围及默认值如表 2 所示.

表 2　实验默认参数配置

参数名	取值范围	默认值
兴趣点多样性要求 l_{u_k}	$5 \leqslant l_{u_k} \leqslant 30$	15 个
区域内 PoI 总数 N	$2.5 \leqslant N \leqslant 20$	20 万
兴趣点查询数量 K	$5 \leqslant K \leqslant 30$	30 个
用户匿名需求 k	$5 \leqslant k \leqslant 20$	10 人
全局移动用户总数 U	$2.5 \leqslant U \leqslant 20$	15 万

4.1　匿名框构造成功率

本文首先考察匿名框构造成功率,即成功构造匿名框用户数量与用户总数量之比. 本文考察两个参数对匿名框构造成功率的影响:兴趣点多样性 l_{u_k} 和区域内 PoI 总数 N.

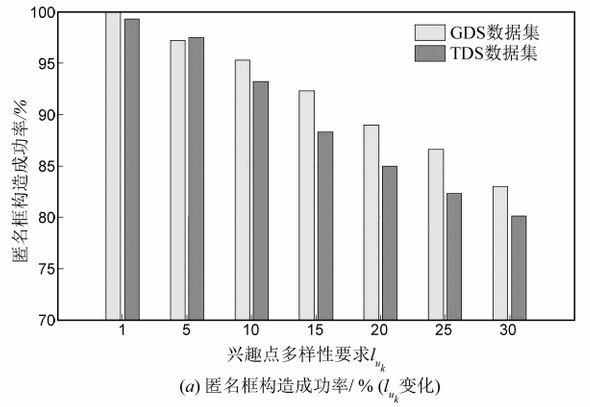
如图 7(a) ~ (b)所示,当用户为提高查询内容隐私保护强度而使兴趣点多样性参数 l_{u_k} 不断增大时,匿名框需要覆盖更多类型的兴趣点,因此匿名框构造成

① http://geonames.usgs.gov/index.html

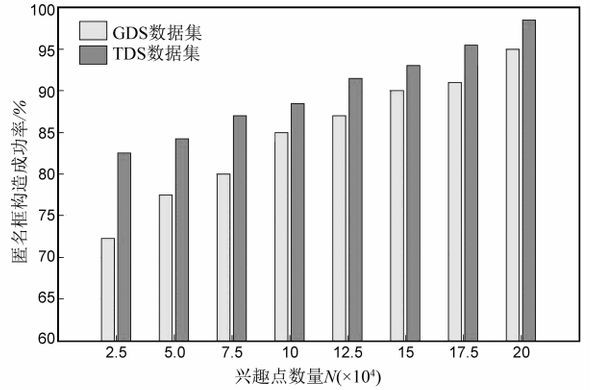
② http://iapg.jade-hs.de/personen/brinkhoff/generator/

功率随之降低,但成功率通常维持在80%以上,而 l_{u_k} 越大,用户查询内容隐私保护效果越好. 因此用户在提出较高查询内容隐私保护强度时,本方法能够保证较高的匿名框构造成功率.

同理,当区域内兴趣点分布较为密集时,用户更容易构造满足兴趣点多样性需求的匿名框. 因此,当区域内兴趣点数量 N 不断增加时,匿名成功率不断提高,而TDS数据集的成功率总是高于GDS数据集,这是由于GDS数据集兴趣点分布不均匀,使得用户在特定区域内兴趣点类型不足造成的.



(*a*) 匿名框构造成功率/ % (l_{u_k}变化)



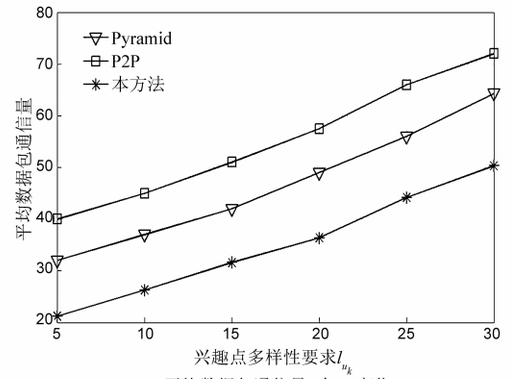
(*b*) 匿名框构造成功率/ % (N变化)

图7 匿名框构造成功率

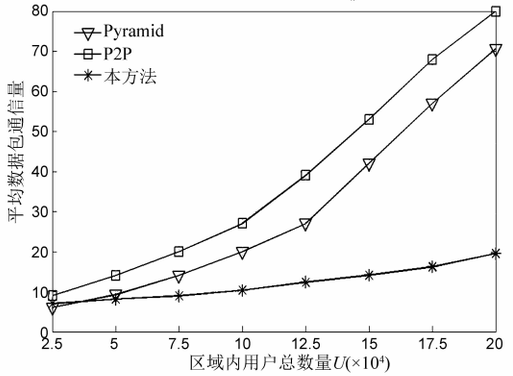
同时,匿名框构造成功率直接影响查询效率,这是由于LBS服务器需要根据用户提交的匿名框为用户查询目标兴趣点的详细信息,本文所提出的两个客户端算法能否正确、快速运行是快速构造高质量匿名框的关键. 通过实验我们可以看出,匿名框构造成功率较高,在LBS服务器根据匿名框返回查询结果平均时间相同的情况下,本方法的查询效率同样较高.

4.2 平均数据通信量

数据通信量是指LBS返回查询结果的信息量,本节将本文所提出的方法与LBS中的隐私保护经典方法数据通信量进行比较. 如图8(*a*)~(*b*)所示. 当兴趣点多样性 l_{u_k} 要求增大时,Pyramid方法[24]和P2P方法[26]都需要找到更多的兴趣点,并且由于没有考虑兴趣点的分布情况,查询过程中返回的兴趣点是盲目的,因此带来了通信数据包量较高. 本方法在考虑了兴趣点分布情况后,使得查询目的性更强,虽然通信量也随着 l_{u_k} 增长而增高,但本方法通信量明显低于传统方法.



(*a*) 平均数据包通信量/ 个(l_{u_k}变化)



(*b*) 平均数据包通信量/ 个(U变化)

图8 平均数据包通信量

同理,当用户数量快速增加时,传统方法数据通信量在数量增加后期快速增长,而本方法的平均数据通信量变化并不明显,其原因是本方法返回的兴趣点描述信息是固定的,数据通信量主要由定义的 l_{u_k} 和 K 影响. 此外,我们还对比了类似方法2PASS[16],如图9所示,本方法与其取得了类似的通信量变化,比2PASS方

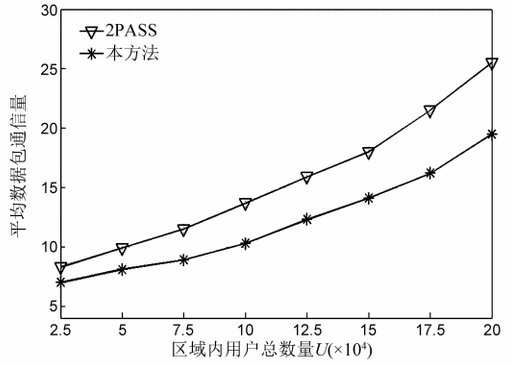


图9 平均数据包通信量/ 个 (与2PASS比较)

法稍低,这是由于本方法采用单一兴趣点 Voronoi 图构造方法,匿名框中虚假查询兴趣点类型用户数是可以掌控的,使通信量相对较低.

4.3　查询准确性

如前 3.3 小节所述,由于用户的算法 1 和算法 2 的查找方式为基于 Voronoi 图划分单元的逐层递进查找,当用户查找 K 近邻兴趣点时,最大会查找某个划分单元的 K 阶邻近单元,这会带来通信量的极大上升,进而使有效时间内获取的准确查询结果数量下降. 但通过分析可以发现,由于 Voronoi 图的邻近单元平均数量特点,使得在实际查找近邻兴趣点过程中通常可以不超过 2 阶近邻单元就可以找到 K 个目标兴趣点.

如图 10 所示,当 K 值不断增大时,表明用户需要一次找到更多的邻近兴趣点,不同数量用户分布在地图空间时,查询的准确率并没有明显的下降,而是稳定在 80% 以上,这表明在 K 值不断增大的查询过程中,所有查询均没有查找 K 阶邻近单元,而是在 2 阶邻近单元内找到了 K 个目标兴趣点,从而使有效时间内的查询准确率保持稳定,符合分析的结果.

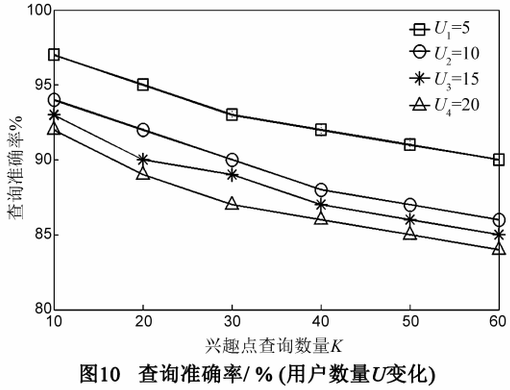


图10　查询准确率/ % (用户数量U变化)

通过类似方法,我们将本文所提出的方法与经典方法 2PASS、新近提出的类似查询方法 KAWCR[27] 和 Singoes[28] 进行了比较,如图 11 所示,发现本文方法具有同等较高的查询准确率,且略高于经典算法 2PASS.

4.4　隐私安全性

信息熵是离散事件集合的平均信息量,通常,处在划分单元 $V(P_i)$ 内的用户,当构造 $K=3$ 的匿名框发起查询时,如图 6 所示,由划分单元面积占总面积的比计算可得,真实发起查询的用户处于匿名框 P_1,P_3,P_5 的概率分别约为 0.4,0.3,0.3,则由式(1)计算可知其信息熵值约为 $H(Q_i)=H(0.4,0.3,0.3)\approx 1.57$bit,由此进一步可知,当用户查询的兴趣点数量 K 增大时,其信息熵的变化如图 12 所示.

由图 12 可知,本方法构造匿名框查询的信息熵随用户查询兴趣点数量增大而稳步增大,攻击者对于匿

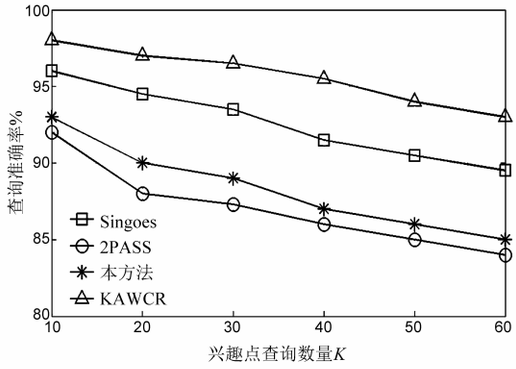


图11　查询准确率/ % (与其他方法对比)

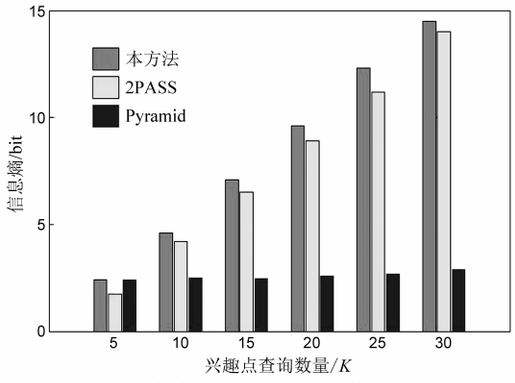


图12　信息熵/ bit (兴趣点查询数量K变化)

名框查询的不确定性随之升高,隐私保护效果越好. 导致信息熵升高的主要原因是,当 K 增大时,匿名框需要覆盖的划分单元数量增多,因此离散事件数量增大,并且更为重要的是,每个离散事件的概率分布相对均匀,使得攻击者的不确定性不断增加. 而传统构造匿名框的方法 Pyramid 其信息熵基本保持稳定,这是由于该方法未考虑兴趣点分布,只构造包含用户在内的匿名框造成的.

联合信息熵是攻击者确定处于某个划分单元内的用户可能性的度量. 由 3.3 节隐私安全性分析可知,从攻击者角度看,要进一步确定某个单元内用户是否为真实发起查询的用户,其联合信息熵值会进一步增加. 这是由于处于某个划分单元内的用户有多个,因此进一步增大了攻击者的不确定性,使隐私保护性能更好. 如图 13 所示:本方法联合信息熵值随兴趣点查询数量增加而稳步增加,符合性能分析的结果;Pyramid 的联合信息熵也略有增加,但是在 K 增长过程中依然保持稳定,是此类方法没有考虑兴趣点分布所造成的缺陷.

5　结论及展望

为了提高 KNN 兴趣点查询效率,本文提出单一兴趣点 Voronoi 图划分方法,该方法首先将同类兴趣点划

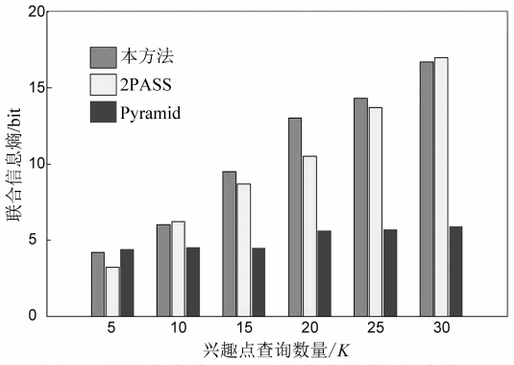


图13 联合信息熵/ bit (兴趣点查询数量*K*变化)

分成 Voronoi 单元,然后利用四叉树进行层次化组织,当用户需要该信息时,将层次化信息发送给用户,以便用户快速找到 *K*NN 兴趣点并构造匿名框. 最终,用户向 LBS 服务器提交匿名框,并只返回匿名框内几类兴趣点详细描述信息. 该方法一方面能够有效降低兴趣点返回数据包量、提高匿名框构造速度,同时能够保护用户的位置隐私和查询内容隐私,性能分析和对比实验表明,本方法具有良好的工作效率.

参考文献

[1] Palanisamy B, Liu L, Lee K, et al. Anonymizing continuous queries with delay-tolerant mix-zones over road networks [J]. Distributed and Parallel Databases, 2014, 32(1): 91 – 118.

[2] 周傲英,杨彬,金澈清,等. 基于位置的服务:架构与进展[J]. 计算机学报,2011,34(7):1155 – 1171.
Zhou Aoying, Yang Bin, Jin Cheqing, et al. Location-based services: Architecture and progress[J]. Chinese Journal of Computers, 2011, 34(7): 1155 – 1171. (in Chinese)

[3] Ghinita G. Privacy for location-based services[J]. Synthesis Lectures on Information Security, Privacy, & Trust, 2013, 4(1): 1 – 85.

[4] 叶阿勇,林少聪,马建峰,等. 一种主动扩散式的位置隐私保护方法[J]. 电子学报,2015,43(7):1362 – 1368.
Ye Ayong, Lin Shaocong, Ma Jianfeng, et al. An active diffusion based location privacy protection method[J]. Acta Electronica Sinica, 2015, 43(7): 1362 – 1368. (in Chinese)

[5] 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[J]. 计算机研究与发展,2010,47(1):121 – 129.
Pan Xiao, Hao Xing, Meng Xiaofeng. Privacy preserving towards continuous query in location-based services[J]. Journal of Computer Research and Development, 2010, 47(1): 121 – 129. (in Chinese)

[6] 刘华玲,郑建国,孙辞海. 基于贪心扰动的社交网络隐私保护研究[J]. 电子学报,2013,41(8):1586 – 1591.
Liu Hualing, Zheng Jianguo, Sun Cihai. Privacy preserving in social networks based on greedy perturbation[J]. Acta Electronica Sinica, 2013, 41(8): 1586 – 1591. (in Chinese)

[7] 唐科萍,许方恒,沈才樑,等. 基于位置服务的研究综述[J]. 计算机应用研究,2012,29(12):4432 – 4436.
Tang Keping, Xu Fangheng, Shen Caiheng, et al. Survey on location based service[J]. Application Research on Computer, 2012, 29(12): 4432 – 4436. (in Chinese)

[8] 陈丹伟,邵菊,樊晓唯,等. 基于 MAH-ABE 的云计算隐私保护访问控制[J]. 电子学报,2014,42(4):821 – 827.
Chen Danwei, Shao Ju, Fan Xiaowei, et al. MAH-ABE based privacy access control in cloud computing[J]. Acta Electronica Sinica, 2014, 42(4): 821 – 827. (in Chinese)

[9] Hashem T, Kulik L, Zhang R. Countering overlapping rectangle privacy attack for moving kNN queries[J]. Information Systems, 2013, 38(3): 430 – 453.

[10] Ngo C N, Dang T K. On efficient processing of complicated cloaked region for location privacy aware nearest-neighbor queries [A]. Proceedings of Information and Communication Technology [C]. Berlin Heidelberg: Springer, 2013. 101 – 110.

[11] Chow C Y, Mokbel M F, Liu X. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments[J]. GeoInformatica, 2011, 15(2): 351 – 380.

[12] 王丽娜,彭瑞卿,赵雨辰,等. 个人移动数据收集中的多维轨迹匿名方法[J]. 电子学报,2013,41(8):1653 – 1659.
Wang Lina, Peng Ruiqing, Zhao Yuchen, et al. Multi-dimensional trajectory anonymity in collecting personal mobility data[J]. Acta Electronica Sinica, 2013, 41(8): 1653 – 1659. (in Chinese)

[13] 徐建,徐明,林欣,等. 路网限制环境中基于匿名蜂窝的位置隐私保护[J]. 浙江大学学报(工学版),2011,45(3):429 – 439.
Xu Jian, Xu Ming, Lin Xin, et al. Location privacy protection through anonymous cells in road network[J]. Journal of Zhejiang University (Engineering Science), 2011, 34(5): 865 – 878. (in Chinese)

[14] Pingley A, Zhang N, Fu X, et al. Protection of query privacy for continuous location based services[A]. Proceedings of IEEE INFOCOM[C]. USA: IEEE PRESS, 2011. 1710 – 1718.

[15] 朱良,孙未未,荆一楠,等. 基于 Voronoi 图的路网 k 聚集最近邻居节点查询方法[J]. 计算机研究与发展,2011,48(z2):533 – 540.
Zhu Liang, Jing Yinan, Sun Weiwei, et al. Voronoi-based aggregate nearest neighbor query processing in road networks[J]. Journal of Computer Research and Development, 2011, 48(z2): 533 – 540. (in Chinese)

[16] Hu H, Xu J. 2PASS: Bandwidth-optimized location cloa-

king for anonymous location-based services[J]. IEEE Transactions on Parallel and Distributed Systems,2010,21(10):1458 – 1472.

[17] 史敏仪,李玲娟. 移动用户位置隐私保护方案研究[J]. 计算机技术与发展,2014,51(10):151 – 154.
Shi Minyi,Li Lingjuan. Study on location privacy protection scheme for moving objects[J]. Computer Technology and Development,2014,51(10):151 – 154. (in Chinese)

[18] Huang Leping,Yamane L,Mastsuura K,et al. Silent cascade:enhancing location privacy without communication QoS degradation[A]. Proceedings of the 3rd International Conference on Security in Pervasive Computing[C]. Berlin Heidelberg:Springer,2006. 165 – 180.

[19] Palanisamy B,Liu L. Mobimix:Protecting location privacy with mix-zones over road networks[A]. Proceeding of IEEE 27th International Conference on Data Engineering (ICDE)[C]. Hannover:IEEE PRESS,2011. 494 – 505.

[20] Freudiger J,Shokri R,Hubaux J P. On the optimal placement of mix zones[A]. Privacy Enhancing Technologies[C]. Berlin Heidelberg:Springer,2009. 216 – 234.

[21] Mano K,Minami K,Maruyama H. Privacy-preserving publishing of pseudonym-based trajectory location data set[A]. The Eighth International Conference on Availability,Reliability and Security (ARES)[C]. USA:IEEE PRESS,2013. 615 – 624.

[22] Yu R,Kang J,Huang X,et al. MixGroup:Accumulative pseudonym exchanging for location privacy preservation in vehicular social networks[J]. IEEE Transactions on Dependable and Secure Computing,2015,PP(99):1 – 12.

[23] Xiaodong L,Rongxing L. Vehicular Ad Hoc Network Security and Privacy[M]. Wiley:IEEE Press,2015. 71 – 89.

[24] Mokbel M F,Chow C Y,Aref W G. The new Casper:query processing for location services without compromising privacy[A]. Proceedings of the 32nd International Conference on Very Large Data Bases[C]. USA:VLDB Endowment,2006. 763 – 774.

[25] 郝忠孝. 时空数据库新理论[M]. 北京:科学出版社,2011.

[26] Chow C Y,Mokbel M F,Liu X. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments[J]. GeoInformatica,2011,15(2):351 – 380.

[27] Gong Z,Sun G Z,Xie X. Protecting privacy in location-based services using k-anonymity without cloaked region[A]. The Eleventh International Conference on Mobile Data Management (MDM)[C]. USA:IEEE Press,2010. 366 – 371.

[28] Ma C,Zhou C,Yang S. A voronoi-based location privacy-preserving method for continuous query in LBS[J/OL]. International Journal of Distributed Sensor Networks,2015,(2015):Article ID 326953. http://dx. doi. org/10. 1155/2015/326953.

作者简介

朱顺痣　男,1973 年 7 月出生,福建东山人,厦门大学博士,厦门理工学院计算机与信息工程学院教授、院长. 主要研究领域为数据挖掘、信息推荐、隐私保护和系统工程.
E-mail:szzhu@ xmut. edu. cn

黄　亮　男,1982 年6 月出生,湖南隆回人,中国科学院计算技术研究所博士. 目前工作于国家计算机网络应急技术处理协调中心,主要研究领域为无线资源管理、异构网络、集中式无线接入网络、移动互联网.

周长利　男,1985 年 3 月出生,黑龙江省哈尔滨人,哈尔滨工程大学博士,华侨大学计算机科学与技术学院讲师. 主要研究领域为位置隐私保护、网络与信息安全.
E-mail:zhouchangli666@ 163. com

马　樱　男,1982 年1 月出生,湖南邵阳人,电子科技大学计算机系统结构专业博士,厦门理工学院计算机与信息工程学院讲师. 主要研究方向为基于挖掘的软件工程、大数据、云计算.