



基于均衡增量近邻查询的位置隐私保护方法

庄礼金, 欧毓毅, 凌捷

(广东工业大学计算机学院, 广东 广州 510006)

摘要: 针对用户在获得位置服务的同时, 用户的位置隐私可能会被泄露的问题, 采用协作的分布式模型, 移动终端和 LBS 服务器这两端都使用 k -匿名机制, 提出基于 k -匿名的均衡增量近邻(KHINN)查询方法。该方法在移动终端构造匿名用户组时, 利用安全多方求和的技术计算锚点以保证用户隐私; 在处理查询结果时, 使用基于 k -匿名的 SpaceTwist 方法提高查询隐私度和准确度。经过性能分析和实验结果表明, 在考虑用户节点之间半可信或不可信的情况下, 可以解决 SpaceTwist 方法中的查询兴趣点围绕锚点分布不均衡的缺陷问题, 提高查询准确度。

关键词: LBS 服务器; k -匿名; 锚点; 查询兴趣点; 安全多方求和

中图分类号: TP309

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016205

Location privacy protection technology based on KHINN

ZHUANG Lijin, OU Yuyi, LING Jie

Computer College, Guangdong Industry University, Guangzhou 510006, China

Abstract: Aiming at the problem that the user's location privacy may be leaked while users obtain location-based services, mobile terminal and LBS server the ends use k -anonymity mechanisms using the collaborative distributed model. Balanced incremental nearest neighbor (KHINN) query method based on k -anonymity was proposed. When the method structuring anonymous user group in mobile terminal, secure multiparty summation was used to calculate the anchor point to ensure user privacy; when processing the query results, SpaceTwist method based on k -anonymity was used to improve the query privacy and accuracy. Performance analysis and experimental results show that undering the condition of semi trusted or untrusted users between nodes, the proposed method can solve defects of uneven distribution of discovering points around the anchor point in SpaceTwist method and improve the accuracy of the query.

Key words: LBS server, k -anonymous, anchor, query point of interest, secure sum

收稿日期: 2016-04-13; 修回日期: 2016-07-13

基金项目: 广东省重大科技专项基金资助项目 (No.2015B010128014, No.2015B010108002); 广东省教育部产学研合作资助项目 (No.2015B090906015, No.2014A010103029, No.2014B090908010, No.2014B090901053)

Foundation Items: Guangdong Provincial Major Scientific and Technological Special (No.2015B010128014, No.2015B010108002), Guangdong Provincial Education Department Research Cooperation Project (No.2015B090906015, No.2014A010103029, No.2014B090908010, No.2014B090901053)



1 引言

基于位置的服务(location based service,LBS)使传感器技术和移动通信设备飞速发展。LBS 将移动通信设备的位置和其他信息整合起来,为用户提供增值服务^[1],即用户将当前位置信息和查询请求内容发送给位置服务器,以获得查询位置服务,其中最主要的位置服务有广告分发服务、交通导航服务和信息娱乐服务等。攻击者能够根据位置数据的时空敏感特性推测出用户的个人信息、位置信息等。用户获取位置服务既有利又有弊,因而位置隐私问题亟待解决。

目前,许多研究者致力于在位置服务和位置隐私之间寻找一个平衡点。参考文献[2,3]是最早将 k -匿名应用到数据隐私的隐私保护方法。位置隐私保护的系统结构有中心服务器系统结构^[4]、客户服务器系统结构、移动 P2P 系统结构等。中心服务器结构是指移动终端用户提出 LBS 查询时,不是直接发送用户的真实位置和位置查询服务属性,而是采用匿名服务器收集相邻的 k 个用户并构造匿名区发送给 LBS 服务器。参考文献[5,6]提出了个性化的 k -匿名算法。因中心服务器会成为系统的性能瓶颈和唯一攻击点^[7,8]以及消耗额外的计算资源和通信代价,许多学者对无中心服务器的研究越来越深入^[9]。客户服务器系统结构是指查询用户向 LBS 服务器发送 k 个位置服务查询属性,其中包括查询用户的真实位置和 $k-1$ 个假位置,使得 LBS 服务器无法识别查询用户的真实位置^[10]。参考文献[11-13]摆脱了中心服务器,采用客户端服务器系统结构。参考文献[11]中,Yiu 等人提出 SpaceTwist 的查询方法是以锚点代替用户真实位置进行增量近邻查询为重点,以 LBS 服务器返回结果候选集为基础,根据隐私需求 k 计算出自己所需的 k 个近邻兴趣点。SpaceTwist 在移动终端没有使用匿名区域,其计算代价和通信代价较低,但是用户在使用 SpaceTwist 过程中没有协作,没有达到 k -匿名。参考文献[12]提出一种基于 SpaceTwist 的 k -匿名增量近邻查询位置隐私保护算法,该算法是根据查询用户的隐私需求以及路网环境构建用户端的匿名区,用户端以该匿名区进行 LBS 查询。参考文献[13]利用四叉树索引划分路网节点构建匿名区,请求路网和兴趣点位置信息,增加查询通信开销。客户服务器系统结构虽然容易实现,但是移动终端用户不能利用周边其他用户的位置信息,因为容易被攻击者识别,而且移动终端必须具备强大的计算和存储能力,造成共享资

源的浪费。移动 P2P 系统结构是通过多跳路由随机选择一个对等点作为代理,由对等点把用户的服务请求发送给服务器,服务器将查询结果返回给该对等点,该对等点对查询结果过滤后返回给查询用户。参考文献[14,15]解决单点脆弱性问题。参考文献[14]中的 CoPrivacy 方法先构建匿名用户组,再进行增量近邻查询。CoPrivacy 在构建匿名用户组时,假设用户彼此之间的协作是可信的,没有考虑到用户之间存在不可信的情况。参考文献[15]参考路网密度构建匿名区,将该匿名区作为代理用户进行 LBS 查询,使得通信开销变得可控。

基于无中心服务器结构的各种隐私方法,大多是假设用户节点足够以及用户之间的协作是完全可信的。若移动终端用户节点不足或存在一个用户不可信时,攻击者易推测用户的真实位置。参考文献[16]提出了一种均衡增量近邻(homogeneous incremental nearest neighbor,HINN)兴趣点查询算法,该算法对不同敏感位置进行定义和筛选,提高用户在敏感位置出现的多样性。当用户节点稀疏,不满足 k -匿名机制时,参考文献[17,18]构建虚拟节点,使用户依据用户隐私需求 k 达到 k -匿名机制,以达到更为理想的位置隐私保护效果。

本文提出基于 k -匿名的均衡增量近邻(K -homogeneous incremental nearest neighbor,KHINN)查询方法是采用移动 P2P 系统结构,研究用户在移动终端如何构建匿名用户组达到 k -匿名机制,如何引用安全多方和技术提高隐私保护度,如何利用均衡增量近邻查询方法提高位置查询的准确度。KHINN 主要包括以下 3 方面内容:如何找到足够的节点数满足匿名用户组的需求;如何在多个节点中锚点;如何解决 SpaceTwist 方法中兴趣点查询结果不准确的问题。

2 位置隐私保护方法

本文提出的方案是基于移动 P2P 系统结构,包括移动终端和 LBS 服务器两部分。该结构通过用户之间的协作完成,具有全局信息。

(1)攻击模型

假定攻击者拦截用户的位置查询服务属性和 LBS 服务器返回的查询结果,但是攻击不可以篡改任何内容。攻击者可根据事先已知的背景知识,由概率的大小推理出用户的敏感信息。

(2)用户模型

对于用户来说,要确保攻击者不能通过查询内容和查

询位置找到用户。将所有用户的查询属性都与锚点的坐标位置对应起来。用户的 LBS 查询被定义为一个 4 元组 $\langle U, t, loc, U_{poi} \rangle$, U 是用户 ID 地址, loc 是用户在 t 时刻提交查询内容的位置, U_{poi} 是位置查询服务属性。假设所有用户节点分别为 $U_1, U_2, U_3, \dots, U_n$, 用户节点对应的真实坐标位置分别为 $loc_1, loc_2, loc_3, \dots, loc_n: (x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$, 其中, 用户协作选取的锚点坐标为 $loc_0: (x_0, y_0)$ 。用户原来的查询请求为 $U_1: (loc_1, U_{poi1}), U_2: (loc_2, U_{poi2}), U_3: (loc_3, U_{poi3}), \dots, U_n: (loc_n, U_{poi n})$ 。对用户组进行锚点后, 每个用户的查询请求为: $U_1: (loc_0, U_{poi1}), U_2: (loc_0, U_{poi2}), U_3: (loc_0, U_{poi3}), \dots, U_n: (loc_0, U_{poi n})$ 。

其中, U_{poi} 定义如下。

定义 1 (查询 U_{poi}) 用户发出的查询 U_{poi} 可定义为: $U_{poi} = \{loc_0, t, CS_i, k_i\}$, 其中, loc_0 表示锚点坐标; t 表示用户提出查询请求的时刻; CS_i 表示第 i 个用户请求查询的内容; k_i 表示第 i 个用户隐私需求。 k_i 越高, 返回的兴趣点个数越多, 用户可以找到最优的兴趣点, 但是查询处理效率会降低, 这由用户进行权衡。

2.1 移动终端的匿名用户组构建

定义 2 (匿名用户组 k -GR) 可以形式化地表示为 $k-GR = \{gid, k, loc_0\}$, 其中, gid 表示匿名用户组的 ID 地址, k 表示用户隐私需求即匿名用户组内的用户节点个数, loc_0 表示匿名用户组的锚点, 每个用户都可使用该锚点提出 LBS 查询。

构建匿名用户组的过程如图 1 所示。用户隐私需求为 k_i , 假定用户 U_1 作为算法的发起者, 发起者生成新的组编号 gid 并广播请求消息, 请求消息内容 $\langle gid, ack \rangle$ 。发起者发出请求后, 等待邻居用户节点的响应, 把响应消息 $\langle res, k_i \rangle$ 的用户节点加入集合 S 中, 若集合 $|S| \geq k$, 查找用户节点结束。当节点搜索结束时, 若 $|S| < k$, 需要生成虚拟节点扩充集合 S 。

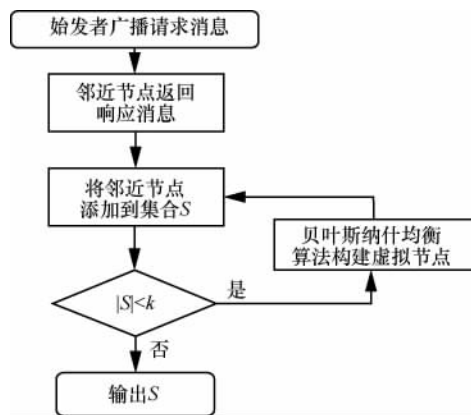


图 1 构建匿名用户组的过程

2.2 移动终端的锚点计算

本文通过 k -匿名机制使用户无法从匿名区域中区分某用户, 即区域的锚点代替查询用户进行 LBS 查询。引入安全多方求和技术计算锚点。这样无需知道每个节点的真实位置就可以得到笔者想要的结果, 并且增加攻击者推测出用户真实位置的难度。集合 S 中 k 个用户的坐标分别是 $U_1(x_1, y_1), U_2(x_2, y_2), U_3(x_3, y_3), \dots, U_k(x_k, y_k)$, 则锚点计算如下。

$x' = (\sum x_i + \sum r_i \times p) \bmod p = \sum x_i \bmod p = \sum x_i (r_i, t_i)$ 是每轮计算选取的随机整数, 假定 p 是大素数), 同理:

$$y' = (\sum y_i + \sum t_i \times p) \bmod p = \sum y_i \quad (1)$$

故:

$$x_0 = (x' \bmod p) / |S| = \sum x_i / k \quad (2)$$

$$y_0 = (y' \bmod p) / |S| = \sum y_i / k \quad (3)$$

即 (x_0, y_0) 为锚点的坐标。计算的简要过程如下。

假定有 5 个用户参与锚点计算, 这些用户可以根据安全多方求和策略, 秘密选择计算的次序为 $U_1 U_2 U_3 U_4 U_5$, 每个用户计算分别如下:

U_1 : 初始计算 (x', y') , $list = \{U_2, U_3, U_4, U_5\}$;

U_2 : 更新 (x', y') , $list = \{U_3, U_4, U_5\}$;

U_3 : 更新 (x', y') , $list = \{U_4, U_5\}$;

U_4 : 更新 (x', y') , $list = \{U_5\}$;

U_5 : 更新 (x', y') , $list = \emptyset$ 。

U_4 将结果返回给 U_1 , 便可得到锚点的坐标。

2.3 LBS 服务器的查询处理

引用 SpaceTwist 的方法进行查询处理, 解决 SpaceTwist 的查询兴趣点总是围绕锚点中心, 分布不平均的问题。查询兴趣点在锚点相反方向上分布可能会不均衡, 会导致兴趣点的查全率下降。在查询处理过程中, U_0 首先建立小顶堆存储距离 U_1 最近的兴趣点, 然后以 U_0 为中心向 LBS 服务器发起增量近邻查询请求。LBS 服务器逐步增量返回兴趣点候选集, 用户通过判断返回的兴趣点与真实位置之间的距离获取精确结果。

算法 1 查询处理算法

输入 用户 U_1 为发起者, 兴趣点集为 C , 锚点为 U_0 , 单次返回兴趣点个数 β , LBS 查询内容为 CS ;

输出 V_k 中前 k 个兴趣点。

以兴趣点 p_i 与用户位置 U_1 之间的距离建立小顶堆 V_k ;
小顶堆 V_k 插入 k 组 $\langle Null, \infty \rangle$;

$\tau \leftarrow 0$; // 供应空间半径赋初值



```

 $\gamma \leftarrow V_k$  堆底元素; //需求空间半径赋初值
锚点  $U_0$  向 LBS 服务器发起增量近邻查询;
While ( $\gamma + \text{dist}(U_1, U_0) > \tau$ )
{
    将  $\langle U_0, CS, \beta \rangle$  发送给 LBS 服务器;
     $C \leftarrow$  LBS 服务器中的兴趣点数据;
     $\tau \leftarrow \text{dist}(U_0, P_i)$  的最大值; //扩大供应空间,  $P_i$  为第  $i$  个
    查询兴趣点
    for (int  $i=0$ ;  $i \leq |C|$ ;  $i++$ )
    {
        if ( $\text{dist}(U_1, P_i) < \gamma$ )
            用  $P_i, \text{dist}(U_1, P_i)$  更新  $V_k$ ;
         $\gamma \leftarrow \text{dist}(U_1, P_i)$ ; //缩小需求空间
    }
}
 $\gamma \leftarrow \text{dist}(U_1, P_k)$ ; //扩大并固定需求空间大小,  $P_k$  为第  $k$  个
查询兴趣点
While ( $\gamma + \text{dist}(U_1, U_0) > \tau$ )
{
     $C \leftarrow$  LBS 服务器中的兴趣点数据;
     $\tau \leftarrow \text{dist}(U_0, P_h)$  的最大值; //继续扩大供应空间,  $P_h$  为
    第  $h$  个查询兴趣点
    if ( $\text{dist}(U_1, P_h) < \gamma$ )
        用  $P_h, \text{dist}(U_1, P_h)$  更新  $V_k$ ;
}
return  $V_k$  中的前  $k$  个兴趣点。

```

3 安全性分析

KHINN 方法是从隐私保护度衡量本文算法的性能。隐私保护度是指攻击者推断用户敏感信息大小来衡量。表1中的第一种方法需要一个可信第三方,而当第三方变得不可信时,用户的隐私将受到严重的威胁。KHINN 算法的前两个步骤都保证了隐私度:匿名用户组的构建以及锚点的计算。KHINN 方法在初始时,始发者先建立匿名用户组编号,然后查找另外 $k-1$ 个用户节点,构建匿名用户组。在某区域内节点不足的情况下,在使用贝叶斯纳什均衡产生虚拟节点的同时使用博弈论的方法权衡生成节点所需的代价,使其满足 k -匿名。当用户节点稀疏时,生成虚拟节点使匿名用户组达到 k -匿名机制,这样保证了算法的可行性。在用户半可信和完全不可信的情况下,引入博弈论的

思想,显然在这种情况下,匿名成功率比其他算法明显要高。CoPrivacy 方法是使用匿名用户组的密度代替锚点。当匿名用户组的密度高时,攻击者很容易找到其锚点。KHINN 引用安全多方求和技术计算锚点,这样就可以在不知道节点准确坐标的情况下,笔者可以得到想要的结果。锚点的计算可以防御伪装在用户之间的内部攻击者和外部攻击者,以此同时,确保敏感信息传输和位置的安全,保护各节点隐私。在匿名用户组中有 1 个敏感位置,用户出现在 1 个敏感位置的概率为 $P(x_i) = \frac{1}{l}$, 用户使用锚点查

询的信息熵为 $H(u_1) = \sum_{i=1}^l P(x_i) \lg \frac{1}{P(x_i)} = \sum_{i=1}^l \frac{1}{l} \lg l = \lg l$ 。

当用户节点分布均匀时,其信息熵最大,意味着攻击者的不确定性达到理论最高值。用户以锚点作为代理查询,攻击者很难将锚点信息与用户的真实位置联系在一起。

将本文的方法和其他方法进行详细比较,见表 1。

表 1 方法性能比较

方法	可信第三方	匿名成功率	安全性
位置 k -匿名方法	有	低	弱
SpaceTwist 方法	无	无	弱
HINN 方法	无	无	弱
CoPrivacy 方法	无	高	较强
本文方法	无	高	强

4 实验仿真分析

本文实验是在 CPU 2.6 GHz、内存为 4 GB 的电脑平台上运行,算法采用 Java 实现。实验数据是 Thomas Brinkhoff^[19] 路由器生成的模拟数据并且以城市 Oldenburg 的交通路网为输入。KHINN 与 CoPrivacy 方案进行实验比较,假定某区域是固定的,图 2 给出平均响应时间与匿名区域内用户数目 N 和匿名用户组锚点之间的关系。从图 1 可以看出,在某区域内,锚点成功率随着移动终端的用户数据增加而提高,与此同时,平均响应时间随着移动终端的用户数量增加而减少。这是因为区域内用户数目增加,节点密集度增加,不需要使用贝叶斯纳什均衡方法构造虚拟节点,便于匿名用户组迅速构建和锚点,减少了平均响应时间。

图 3 是在移动终端构建匿名用户组的过程。从 CoPrivacy 与本方案比较来看,在相同的实验环境下, KHINN 的匿名成功率总是接近 100%。因为在节点稀疏的情况下,用户节点不足,CoPrivacy 很难构建匿名用户组,

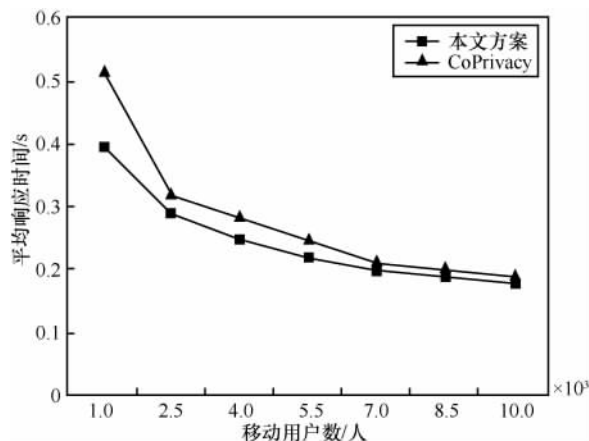


图2 平均响应时间

而KHINN可使用贝叶斯纳什均衡方法构造匿名用户组。此外，密集的用户生成的锚点更接近用户的真实位置，使得用户查询结果比较准确。

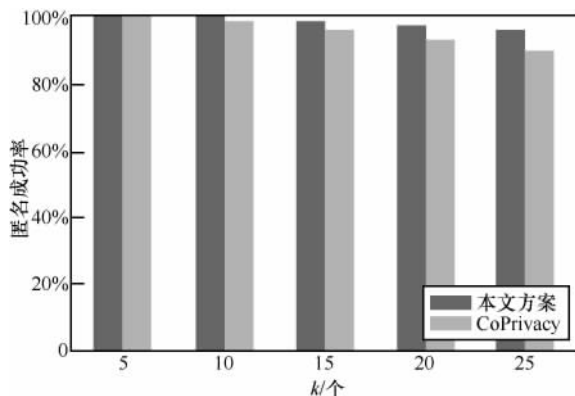


图3 匿名成功率

如图4所示，KHINN的查准率比CoPrivacy高。因为KHINN在移动终端和LBS服务器两端都使用了 k -匿名机制，特别是在增量近邻查询的过程中，解决查询兴趣点围绕锚点为中心分布导致查询结果不均衡的情况。

5 结束语

以往的位置隐私保护方法都是采用可信第三方的系统结构，可行第三方将含用户位置的匿名区域代替用户真实位置进行查询处理，这容易使可信第三方成为系统的性能瓶颈和集中攻击目标。KHINN方法是通过用户之间协作形成匿名用户组，当用户节点稀疏时，可利用贝叶斯纳什均衡方法构建虚拟节点，使用安全多方求和的方法进行锚点计算，利用均衡增量近邻查询算法进行查询处理。该方法考虑了用户节点半可信和不可信的情况，提高查准率。

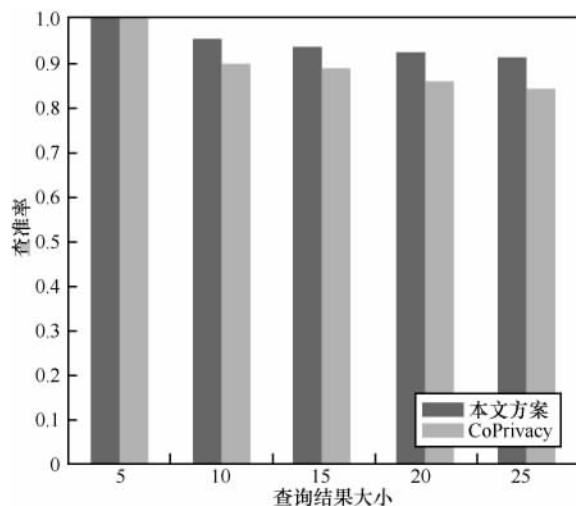


图4 CoPrivacy与本文方案的查准率比较

参考文献：

- [1] 潘晓,肖珍,孟小峰.位置隐私研究综述[J]. 计算机科学与探索, 2007, 1(3): 268-281.
PAN X, XIAO Z, MENG X F. Location privacy research review [J]. Journal of Computer Science and Exploration, 2007, 1 (3): 268-281.
- [2] SWEENEY L. K -anonymity: a model for protecting privacy [J]. International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.
- [3] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]// The 1st International Conference on Mobile Systems, Applications and Services, September 11-15, 2003, San Francisco, USA. New York: ACM Press, 2003: 31-42.
- [4] GEDIK B, LIU L. Protecting location privacy with personalized k -anonymity: architecture and algorithms [J]. Mobile Computing, 2008, 7(1): 1-18.
- [5] GEDIK B, LIU L. A customizable k -anonymity model for protecting location privacy[J]. Icdcs, 2004: 620-629.
- [6] GONG Z, SUN G Z, XIE X. Protecting privacy in location-based services using k -anonymity without cloaked region[C]//Eleventh International Conference on Mobile Data Management, MDM 2010, May 3-6, 2010, Kanas City, Missouri, USA. New Jersey: IEEE Press, 2010: 23-26.
- [7] 霍峥, 孟小峰, 黄毅. Private Check In: 一种移动社交网络中的轨迹隐私保护方法[J]. 计算机学报, 2013, 36(4): 716-726.
HUO Z, MENG X F, HUANG Y. Private check in: a method of trajectory of mobile social network privacy protection [J]. Journal of Computers, 2013, 36(4): 716-726.
- [8] MOKBEL M F, CHOW C Y, AREF W G. The new casper: query



- processing for location services without compromising privacy[C]//The International Conference on Very Large Data Bases(VLDB'06), May 12-14, 2006, Seoul, South Korea. New York: ACM Press, 2006: 763-774.
- [9] 张学军, 桂小林, 伍忠东. 位置服务隐私保护研究综述[J]. 计算机学报, 2015, 26(9): 2373-2395.
ZHANG X J, GUI X L, WU Z D. Location privacy protection and research review [J]. Journal of computers, 2015, 26(9): 2373-2395.
- [10] KIDO H, YANAGISAWA Y, SATOH T. An anonymous communication technique using dummies for location-based services[C]//International Conference on Pervasive Services, 2005 (ICPS'05), July 11-14, 2005, Santorini, Greece. New Jersey: IEEE Press, 2005: 88-97 .
- [11] YIU M L, JENSEN C S, HUANG X, et al. Spacetwist: managing the trade-offs among location privacy, query performance, and query accuracy in mobile services [C]//IEEE 24th International Conference on Data Engineering, April 7-12, 2008, Cancun, Mexico. New Jersey: IEEE Press, 2008: 366-375.
- [12] 胡德敏, 郑霞. 基于 SpaceTwist 的 k -匿名增量近邻查询位置隐私保护算法[J]. 计算机应用研究, 2015(33): 144-162.
HU D M, ZHENG X. K -anonymous incremental neighbor query based on SpaceTwist location privacy protection algorithm [J]. Computer Application Research, 2015(33): 144-162.
- [13] 周长利, 马春光, 李增鹏. 一种保护用户隐私的路网兴趣点 KNN 查询方法[J]. 计算机应用研究, 2016, 33(1): 262-265.
ZHOU C L, MA C G, LI Z P. A network point of interest to protect the privacy of users KNN query methods[J]. Computer Application Research, 2016, 33(1): 262-265.
- [14] 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 no anonymous user collaboration area of location privacy protection method [J]. Journal of Computers, 2011, 34(10): 1976-1985.
- [15] 王鹏飞, 李千目, 朱保平. 基于增量近邻查询的位置隐私保护方法[J]. 计算机科学, 2015, 42(3): 158-161.
WANG P F, LI Q M, ZHU B P. Location privacy protection method based on incremental neighbor query [J]. Journal of Computer Science, 2015, 42(3): 158-161.
- [16] 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): 125-136.
ZHOU C L, MA C G, YANG S T. LBS location privacy protection method based on position sensitive diversity study[J]. Journal of Communications, 2015, 36(4): 125-136.
- [17] LIU X X, LIU K K, GUO L K, et al. A game-theoretic approach for achieving k -anonymity in location based services [C]//INFOCOM'13, April 14-19, 2013, Turin, Italy. New Jersey: IEEE Press, 2013: 2985-2993.
- [18] MIURA K, SATO F. A hybrid method of user privacy protection for location based services[C]//CISIS'13, July 3-5, 2013, Taichung, China. New Jersey: IEEE Press, 2008: 55-62.
- [19] BRINKHOFF T. A framework for generating network based moving objects[J]. Geoinformatica, 2000, 6(2): 153-180.

[作者简介]



庄礼金(1990-),女,广东工业大学计算机学院硕士生,主要研究方向为网络信息安全技术。



欧毓毅(1974-),女,博士,广东工业大学计算机学院主任、副教授,主要研究方向为信息系统集成。



凌捷(1964-),男,博士,广东工业大学计算机学院院长、教授,主要研究方向为网络信息安全技术、计算机网络通信。