计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2017,26(8):267−272 [doi: 10.15888/j.cnki.csa.005885]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

一种位置服务隐私保护方法①

王木涵, 汪佳帧, 阳　杰, 迟焕醒, 徐九韵

(中国石油大学(华东) 计算机与通信工程学院, 青岛 266000)

摘　要: 近年来, 大多数的位置服务隐私保护转向了基于用户协作的 P2P 形式, 但是此模式最大的弊端在于协作用户存在不诚信的情况, 如果出现恶意的协作用户, 那么协作组用户信息可能会被泄露. 在此前提下, 本文通过添加第三方信任机构为移动用户进行网络行为异常检测的方式, 提出一种防范位置隐私泄露给恶意用户的方法, 并通过二叉树的形式扩展 P2P 模式下的寻找协作用户的范围. 通过实验验证, 可以有效地防范非诚信用户加入匿名组, 降低了信息泄露的可能.

关键词: 位置服务隐私保护; P2P; 第三方信任; 二叉树模型

引用格式: 王木涵,汪佳帧,阳杰,迟焕醒,徐九韵.一种位置服务隐私保护方法.计算机系统应用,2017,26(8):267–272. http://www.c-s-a.org.cn/1003-3254/5885.html

Location-Based Service Privacy Protection Method

WANG Mu-Han, WANG Jia-Zhen, YANG Jie, CHI Huan-Xing, XU Jiu-Yun

(School of Computer & Communication Engineering, China University of Petroleum, Qingdao 266580, China)

Abstract: In recent years, most location-based service privacy protection turns to peer-to-peer based on subscriber cooperation. However, the biggest drawback of that mode lies in the integrity of collaborative users. If there are vicious users in ambient collaborative users, the information of collaborative users may be accessed illegally. So this paper proposes a method of preventing the location privacy from being revealed to vicious users via using third-party trust institutions to detect the abnormal network behaviors for mobile users. In the meanwhile, we broaden the zone of collaborative users in the mode of peer to peer through binary tree. The experimental result shows the method can effectively prevent unfaithful users from joining in the anonymous group and reduce the probability of the information leakage.

Key words: location-based service privacy protection; P2P; the trusted institutions third party; binary tree

1　引言

基于位置的服务作为一项位置增值服务[1], 用户通过提交位置查询请求, 便可以轻松方便的获取位置服务, 如位置近邻查询, 路线查询等等. 随着数据挖掘技术的不断发展, 如果用户直接将真实的位置信息提交给位置服务提供商, 难免信息可能会遭到泄露, 所以很多专家学者也开始对位置隐私保护展开研究[2].

目前, 位置服务隐私保护研究主要集中在中心式及分布式的系统结构, 本文结合以上两种结构, 转变可信第三方的职能为信任检测机构的方式, 建立一种半中心式半分布式的系统结构, 这种方式具有以下优势: (1)能有效避免中心式结构中的系统瓶颈问题; (2)确保分布式结构中的用户诚信问题, 主要工作如下:

(1) 通过第三方信任检测机构来记录并检测用户是否存在行为异常, 将用户的不诚信度反馈给查询用户, 过滤掉不诚信用户, 不与之协作匿名.

① 收稿时间: 2016-12-03; 采用时间: 2016-12-19

(2) 采用二叉树的形式广播匿名需求, 用户较多时不至于协作用户过于集中, 匿名区域过小而造成查询位置泄露, 同时避免中心攻击.

本文第 2 节讲述相关工作, 第 3 节介绍本文系统结构, 第 4 节介绍位置隐私保护方法, 第 5 节讲述本文实验结果及分析, 最后第 6 节总结及展望.

2 相关工作

近年来, 基于位置服务的隐私保护问题得到了许多研究者的重视. 总结目前的研究主要包括以下的几个方面:

文献[3]基于对数据库信息的保护问题最早提出了 k-匿名保护模型, 2006 年 Mokbel[4]首次提出了将 k-匿名应用到 LBS 中, 其基本思想是在用户发起位置服务请求时, 用一个包含其他 k-1 个用户的匿名区域代替用户真实位置进行服务请求, 使得恶意用户无法判断究竟是哪个用户提出的请求, 以此达到保护用户位置隐私的效果. 同年, Mokbel[5]又提出 Casper 空间匿名模型, 该模型使用网格对匿名服务器所覆盖的范围进行划分, 以单个网格为单位进行 k-匿名构造, 此方式的好处在于可以提前将匿名服务器范围内具有查询需求的用户保存到网格中, 加快响应速度.

由于中心式结构存在系统瓶颈, 一招失失全盘的致命性缺点, 更多的研究转向了分布式无中心化的隐私保护结构, 2006 年 Chow[6]等提出了基于分布式的位置隐私保护结构, 此结构取消了中心式结构的匿名服务器, 由用户自行组织构成匿名组形成 k-匿名区域, 随机选取其中一个用户向位置服务器发出请求, 位置服务器将候选结果集反馈给该用户, 再由该用户将结果求精后返回给请求用户. 文献[7]提出了 SpaceTwist 的方法, 用户随机选取一个自己附近的点作为锚点, 并以此位置向位置服务器进行增量的请求信息, 但是此方法没有实现 k-匿名, 隐私效果较差. 孟小峰[8]等对此进行了改进提出了 Coprivacy 方法, 使用单跳或者多跳的形式构造匿名组, 并以匿名组密度中心作为锚点发起增量式的请求服务, 最后每个用户根据返回结果与自己的位置计算得到想要的结果. 由于短程通信约束(协作用户发现策略和通信范围)和用户的移动模式, 现有基于 P2P 模式遭受两个问题, 首先, 所选的协作用户很可能均匀的分布在真实用户的附近; 其次, 隐匿区可能不是足够大来满足每个移动用户的最低要求. 因此, 攻击者可以实现以较高的概率确定用户的实际位置[12]. Niu 等人提出了基于偏差的攻击并通过 R-cloak 方法减轻了改攻击. Chow 等[14]提出了共享对等信息及历史位置信息减小通信开销, 此等结构虽然排除了系统瓶颈的问题但是增加了移动端的开销, 而且周围用户的诚信程度也难于鉴别, 如果存在恶意的协作匿名用户, 那么其他用户的隐私也可能遭到泄露. 基于此, 陈玉凤[9]等提出了基于博弈论的用户相互协作的位置隐私保护方法, 并通过安全求和[10]的方法来计算锚点, 解决不诚信合作的问题. YANG[11]等通过数字签名技术, 用接收方的私钥加密位置信息, 实现诚信用户之间的验证, 避免了位置隐私的泄露. Yang 等[13]通过网络中的直接互惠机制引入到 LBS 隐私保护中, 利用网络中节点向其他节点提供服务直接获取利益. 这个机制依赖于 2008 年 Levin 及 LaCurts 提出的拍卖模型, 并且现在是唯一一个在 LBS 隐私保护中引入了激励机制的研究. Li 等[15]基于软计算的模糊逻辑提出一种基于信贷激励的机制, 为其他人提供帮助, 用户可以获得积累自己的信用, 当信用达到某个阈值的时候才可以帮助其他用户.

3 位置服务隐私保护系统框架

本文选取基于半中心式半分布式的系统结构, 包括智能移动终端、位置服务提供商及可信的第三方信任检测机构三部分, 系统框架如图 1 所示.

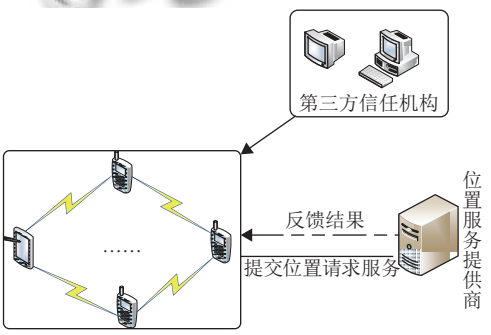

图 1 系统框架图

智能移动终端中 GPS 模块用来获取移动用户的位置信息, 通信模块要求移动终端支持 P2P 通信及无线通信, P2P 通信模块用户可用蓝牙及无线通讯通过一跳或者多跳的形式与其他用户通信形成匿名组, 无线通信模块主要通过运营商的 2G/3G/4G 数据流量或

者 WLAN 来向位置服务提供商请求位置服务.

匿名模块需要完成用户之间的协作组成 k-匿名, 移动用户可以自定义匿名需求 k 及用户不诚信度 unv.

第三方信任检测机构仅具有用户异常行为检测功能, 给用户反馈检测结果之后, 空间立即释放, 本身不具有任何用户隐私信息. 用户异常检测模型(图 2)主要分为训练阶段及检测阶段, 训练阶段通过对抓取的历史用户数据包进行数据的处理, 选择出我们所需要的字段, 形成训练样本数据集, 通过对访问频率的分析统计, 我们认为用户访问频率大概的服从高斯分布(公式 1), 当访问频率 f 的概率高于某个值的时候, 我们即认为该用户具有异常行为; 检测阶段通过之前预测样本的值进行直接检验, 省去复杂的检验步骤, 具有较快的速度检测用户是否存异常. 如果过去 30 天之内存在短时间内频繁变换访问请求, 且流量传输巨大, 则认为此用户行为异常, 不诚信值取值范围(0, 1), 不诚信值随时间衰减, 时间衰减函数如公式 2, 并将不诚信值 unv 发给查询用户, 由用户预先设定的不诚信值进行比较后决定是否与之协作.

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2}) \quad (1)$$

$$\text{unv}=e^{-k\cdot\Delta t} \quad (2)$$

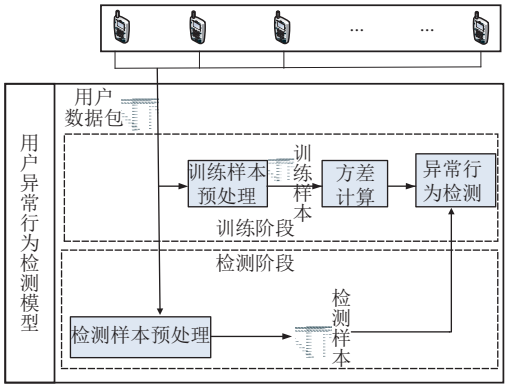

图 2 用户异常行为检测模型

4 位置隐私保护算法

4.1 预备知识

定义 1. 用户请求: 本文用户请求定义为集合 Q = {id,t,loc,con,k,unv}, 其中 id 为用户请求服务时用户的标识符, 主要用来判断用户是否加入匿名组; t 为用户发起请求时的时间; loc 为用户发起请求时的地理位置; con 为用户发起请求的具体内容; k 为用户的协作匿名参数. unv 为用户设置的协作用户不诚信参数值, 默认 0.01, 越小证明需要的安全性越高.

定义 2. k-匿名组: 协作匿名组 k-ag={fgId,k,n, anchorg}, 其中 gId 为匿名组的标识符; k 为用户发起的匿名参数; n 为当前加入匿名组的用户数; anchor 为根据匿名组内用户的位置坐标计算得出的密度中心点.

定义 3. 节点间距离:

$$\text{dis}\left(U_{(x_1,y_1)},U_{(x_2,y_2)}\right)=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2} \quad (3)$$

定义 4. 不诚信用户: 本文认为不诚信用户为短时间内频繁访问不同站点的具有异常行为的用户.

4.2 匿名算法

步骤 1. 发现节点. 用户 U_q 发起位置请求时, 首先建立匿名组生成组标识符 gId, 并置初始匿名组集合为 G{u}, u 为加入匿名组的用户, 已发现的用户数目为 n=|G|, 初始情况下即为 0. 首先用户 U_q 广播节点发现消息 FROM_GROUP={gId}, 在一跳范围内以二叉树的形式随机选取两位协作用户, 收到消息的用户 U_p 首先反馈给一个用户的唯一标识 f, 用户 U_q 使用该标识请求第三方验证该用户的诚信度, 第三方反馈用户 U_p 不诚信值, 如果满足 U_q 预先设置的不诚信值, 则将该用户纳入匿名组, 并将匿名参数广播给该用户, 记录 U_p 为子节点, U_p 记录其父节点, 用户数目 n+1, 否则剔除该用户不与之协作匿名. 检测匿名组内协作用户的匿名需求参数 k', 如果 k'>k, 则更新 k=k'. 当组内用户数 n<k 时, 以叶子节点为中心一跳范围内继续以二叉树的形式寻找协作用户, 如果增加叶子节点后仍不满足匿名需求且叶子节点不能找到子节点, 则叶子节点回退一级父节点, 重新寻找其他子节点, 如果直到返回根节点时仍无法满足匿名需求, 则认为匿名失败.

算法 1: 节点发现

输入: FROM_GROUP={gId}
输出: G{u}

1. //发现节点
2. 请求用户 U_q
3. 初始化匿名组 U_q.gId
4. WHILE(n<k-1)
5. 广播发现节点消息 FROM_GROUP(U_q.gId)
6. 收到广播的用户 U_p 反馈给标识 f 给 U_q

7. U_q 请求第三方信任机构该标识的诚信度
8. IF U_p.unv > U_q.unv
9. 抛弃该用户;
10. ELSE
11. 以二叉树的形式随机选择协作用户, 记录子节点标识, 子节点同时记录父节点标识
12. END IF
13. n=|G|;
14. IF(n<k-1&& U_p.leftchild==null&& U_p.rightchild==null)
15. 回溯父节点寻找其他子节点
16. ELSE
17. 匿名失败
18. END IF
19. END IF
20. END WHILE

作为加入匿名组的响应节点 U_p, 在收到发起用户的广播消息 FROM_GROUP(U_q.gId)后, 首先检查 U_p.gId 若为 null, 则反馈自己的唯一标识 f 给发起用户, 当 U_p.f<U_q.f 时, 用户更新 U_p.gId=U_q.gId, 同时如果响应节点 U_p 的匿名需求 k'>k, 则更新 k= k', 如果 n<k-1, 则证明相应节点需要继续广播该消息到其子节点; 当 U_p.gId==U_q.gId, 否则忽略此广播, 确保一个用户在一个匿名组中.

算法 2: 用户加入匿名组

1. 输入广播 FROM_GROUP(U_q.gId)
2. 自身节点为 U_p
3. IF(U_p.gId==null)
4. 反馈自己的唯一标识 f 给发起用户
5. IF(U_p.f< U_q.f)
6. U_p.gId= U_q.gId
7. n++
8. IF(k′ >k)
9. k= k′
10. END IF
11. END IF
12. IF(n<k-1)
13. 向下一级广播 FROM_GROUP(U_q.gId)
14. END IF
15. END IF

步骤 2. 计算锚点

计算集合 G 内用户位置的密度中心作为锚点 anchor(*anchor.x*,*anchor.y*), 公式如下:

anchor.x = $\frac{\sum_1^k x_i}{k}$, anchor.y = $\frac{\sum_1^k y_i}{k}$, x 为节点横坐标, *y* 为节点纵坐标, *i* 取值为 1, 2, 3……*k*.

步骤 3. 广播锚点

用户 U_q 计算得到锚点 anchor 后, 将锚点广播给自己所有的孩子节点, 用户以锚点位置作为自己位置发送位置服务请求.

算法 3: 广播锚点

1. 输入 U_q. anchor
2. while(u.leftchild!=null&& u.rightchild==null)
3. u anchor==U_q. anchor
4. end while
5. 替换位置点向位置服务器发送请求 *Q*=fid,t,anchor, con,k,unvg

5 实验

本文使用 IBM 服务器 CPU: Intel Xeon 2.2 GHz, 内存: 16 GB, 操作系统: Windows Server 2008, 使用 Java 编程语言进行模拟实验. 实验参数采取文献[6]的实验数据, 移动用户之间通过一个具有 2 Mbps 带宽及 250 米的传播范围的半双工无线信道进行交流, 移动用户与位置服务提供商之间采取 10 Mbps 带宽进行通信. 实验默认参数值如表 1 所示.

表 1 变量取值

参数名称	默认值
移动用户数量	400
区域范围	1000 m^2
通信半径	250 m
k值	5
不诚信用户数	1

5.1 第三方诚信检测

移动用户向网络发送服务请求, 网络传播请求数据必然要通过 http 请求, 第三方信任检测机构负责监视并记录移动用户访问网络的记录, 通过对网络上传输数据包的检测分析得出网络行为异常的用户, 获取

的数据包主要包括请求时间, 源地址, 目标地址, 信息流量等.

通过 6112 组用户网络数据包进行分析统计, 如图 3 所示, 可以看出访问频率服大概从高斯分布, 于频次等于 15 次时到峰值, 而在高斯分布中$|X-\mu|\geqslant 3\sigma$是一个小概率事件, 如果 x 是 X 的一个观察值, 通常我们认为当$|X-\mu|\geqslant 3\sigma$时 x 是一个异常数据, 故而根据我们的实验数据, 当用户访问频次超过 26 次时认为此用户存在异常行为, 将此用户不诚信度置为 1.

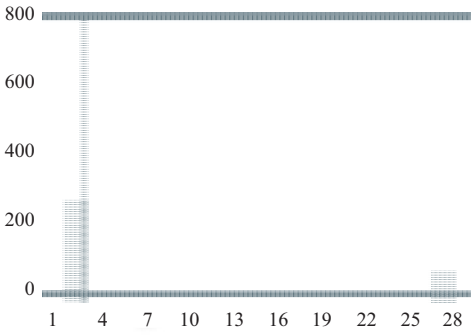

图 3 用户数据包分析

本文假设不诚信影响程度在 30 天左右, 30 天以后影响基本为 0, 由图 4 我们看到当 k=0.2(虚线)时, days=20, untrustworthy 变的很小与 days=30 基本差不多, 与需求相差较大; 而当 k=0.1(点线)时, days=30, untrustworthy≈0.05 依旧相对较大, 与需求相差也较大; k=0.15(实线)时, days 与 untrustworthy 约为 0.01 的变化趋势正符合我们的要求。因此, 本文将时间衰变参数 k 设置为 0.15. 图 4 中, 实线为 k=0.15 时, 即反馈信息用户不诚信度随时间衰减情况. 当在不诚信值衰减的过程中, 如果用户出现请求访问量再度超过预设值, 则重新按照衰减函数计算不诚信值.

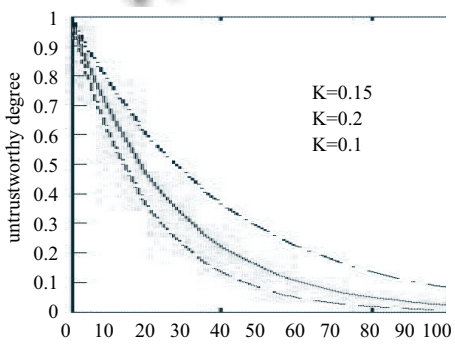

图 4 时间衰减函数

5.2 二叉树模型匿名实验

实验的可扩展性通过增加移动用户数量, k 值以及不诚信用户的数量对实验结果进行数据的统计.

在默认参数条件下本文通过增加区域内移动用户的数量进行通信量的实验, 实验统计结果如图 5 所示. 可以看出随着用户数量的增多, 通信数量逐渐增长, 这是因为随着用户数量的增多, 在用户可通信的范围内的移动用户数量有所增加.

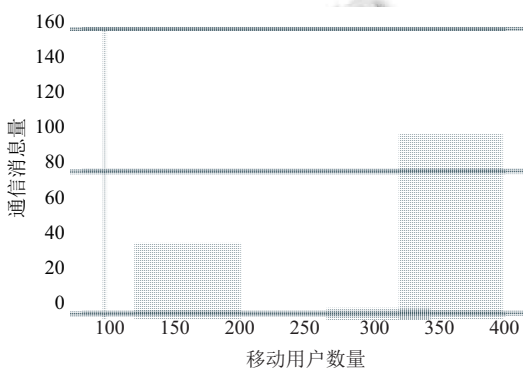

图 5 通信消息量

图 6 显示随 k 值的增加, 响应时间的变化情况. 响应时间是指用户发起匿名协作请求到发送位置请求服务的时间差. 由于 k 值的增大, 用户请求协作用户数量增多, 节点间的通信增多, 响应时间随之增加. 当不诚信用户数目增多时, 响应时间有所增加, 但是没有出现大幅度的变化.

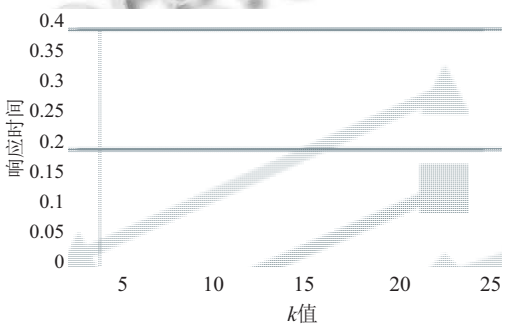

图 6 响应时间

图 7 展示的是当移动用户数量一定的情况下, 随着 k 值的不断变化, 用户协作匿名组成功率的变化情况. 匿名成功率指用户发起的匿名协作请求成功的数量与总的请求数量的比值. 可以看出匿名成功率总体维持较高的水平, 随着不诚信用户数量的增多, 对匿名成功率略有影响, 但是影响的效果并不十分明显, 从一定角度可以看出本文算法的优越性.

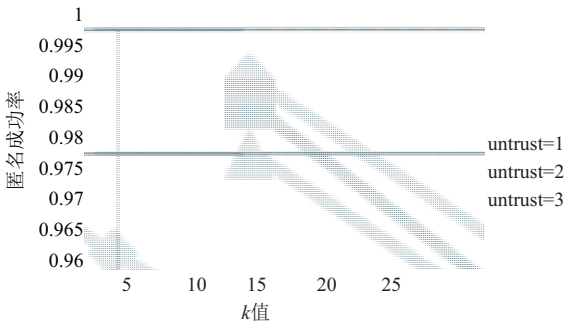

图 7 匿名成功率

6 总结

本文考虑到在现实环境中存在不诚信协作用户的情况, 提出基于异常用户检测的二叉树模型匿名保护方法, 通过向第三方请求协作用户可信情况来判断是否与之协作匿名, 并通过使用二叉树的形式寻找协作用户, 在成功实现杜绝不诚信用户加入匿名组的同时, 扩展了协作用户的范围, 通过实验对响应时间及成功率进行了统计分析, 验证算法的有效性.

参考文献

1 Mokbel MF. Privacy in location-based services: State-of-the-art and research directions. Proc. of 2007 International Conference on Mobile Data Management. Mannheim, Germany. 2007. 228–228.

2 张学军, 桂小林, 伍忠东. 位置服务隐私保护研究综述. 软件学报, 2015, 26(9): 2373–2395.

3 Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking. Proc. of the 1st International Conference on Mobile Systems, Applications and Services. San Francisco, California, USA. 2003. 31–42.

4 Mokbel MF. Towards privacy-aware location-based database servers. Proc. of the 22nd International Conference on Data Engineering Workshops. Atlanta, GA, USA. 2006. 93.

5 Mokbel MF, Chow CY, Aref WG. The new Casper: Query processing for location services without compromising privacy. Proc. of the 32nd International Conference on Very Large Data Bases. Seoul, Korea. 2006. 763–774.

6 Chow CY, Mokbel MF, Liu X. A peer-to-peer spatial cloaking algorithm for anonymous location-based service. Proc. of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems. Arlington, Virginia, USA. 2006. 171–178.

7 Yiu ML, Jensen CS, Huang XG, *et al*. SpaceTwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services. Proc. of IEEE the 24th International Conference on Data Engineering. Cancun, Mexico. 2008. 366–375

8 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.

9 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. 计算机科学, 2013, 40(10): 92–97. [doi: 10.3969/j.issn.1002-137X.2013.10.019]

10 张国荣, 印鉴. 基于博弈论的安全多方求和方法. 计算机应用研究, 2009, 26(4): 1497–1499, 1502.

11 Yang YQ, Yuan JB. Research on incredible users cooperate constructing anonymous region in LBS. Computer Engineering and Applications, 2014, 50(14): 82–87.

12 Niu B, Zhu XY, Li QH, *et al*. A novel attack to spatial cloaking schemes in location-based services. Future Generation Computer Systems, 2015, 49: 125–132. [doi: 10.1016/j.future.2014.10.026]

13 Yang DJ, Fang X, Xue GL. Truthful incentive mechanisms for k-anonymity location privacy. Proc. of the 32nd IEEE International Conference on Computer Communications. Turin, Italy. 2013. 783–796.

14 Chow CY, Mokbel MF, Liu X. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments. Geoinformatica, 2011, 15(2): 351–380. [doi: 10.1007/s10707-009-0099-y]

15 Li XH, Miao MX, Liu H, *et al*. An incentive mechanism for *K*-anonymity in LBS privacy protection based on credit mechanism. Soft Computing, 2016, doi: 10.1007/s00500-016-2040-2. (in Press)