# 不影响隐私的位置服务查询处理模型

雷建云,姚 瑶
(中南民族大学 计算机科学学院,武汉 430074)

**摘 要** 基于位置服务的应用中,针对没有可信任的服务器人们的隐私信息将受到严重威胁的问题,提出了一个位置服务查询处理模型. 该模型是移动和固定用户在不显示其位置信息的情况下使用基于位置服务的新框架. 实验结果显示: 该模型位置匿名器采用的金字塔结构较其它算法有一定的优越性,用户数可达到 50000 个或更多,且用户数越多,位置匿名器的性能越高; 隐私感知查询处理器使用过滤算法可大幅减少查询处理时间. 该模型具有一定的理论价值和实用价值.

**关键词** 基于位置的服务; K-匿名; 位置隐私
**中图分类号** TP393 **文献标识码** A **文章编号** 1672-4321(2017)04-0121-05

# Query Processing Model for Location Services without Privacy Compromising

*Lei Jianyun*, *Yao Yao*
(College of Computer Science, South-Central University for Nationalities, Wuhan 430074, China)

**Abstract** In location based service applications, people's privacy may be compromised without trusted servers. To address this problem, a new model with a framework in which mobile and stationary users can use location-based services without revealing their location information is presented. The experimental results indicate that the model using pyramid structure can contain an amount of users more than 50000, and more users brings better performance, privacy query processor uses filter algorithms to drastically reduce query processing time. The new model has certain theoretical value and practical value.

**Keywords** location-based service; K-anonymity; location privacy

随着移动通信和传感器设备等位置感知技术的快速发展,基于位置的服务(LBS)应用越来越广泛[1]. 例如,查询兴趣点(结合当前位置信息查询最近的医院、加油站、宾馆以及娱乐场所等); 分享社交网络位置(签到、大富翁游戏等). 为了获取准确的结果,用户需不断地向服务器发送自己的位置信息,这些位置数据不仅直接包含用户的隐私信息,还可以挖掘出用户的健康状况、社会地位等敏感信息. 若位置信息被不正当使用或被第三方恶意攻击,会给用户的隐私带来严重的威胁.

最早 Samarati P 和 Sweeney L 提出 K-匿名技术[2],并用于保护用户的隐私. 2003 年,Gruteser 等人[3]将 K-匿名技术的思想引入到位置服务 LBS 中,通过模糊用户的空间位置信息以达到隐私保护的目的[4]. 2014 年,王璐,孟小峰根据位置隐私的保护程度,把现有方法总结为基于启发式隐私度量、概率推测和隐私信息检索的位置大数据隐私保护技术[5]. 文献[6]利用服务查询结果的相似性来辅助匿名区域,提出具有较高平衡性的 K-匿名位置隐私保护方法. 本文提出的保护隐私模型的位置匿名器具有以下优点: (1) 模型为每个用户提供可定制的隐私简档: 包括 $K$ 值,最小隐藏区域 $A_{\min}$; (2) 很好地衡量大量具有任意隐私简档的移动用户; (3) 不能逆向获取关于用户确切位置的任何信息.

**收稿日期** 2017-07-26
**作者简介** 雷建云(1972-),男,教授,研究方向: 信息安全,E-mail: leijianyun@ mail. scuec. edu. cn
**基金项目** 湖北省自然科学基金资助项目(2014CFB445)

## 1 系统框架

移动用户在注册使用系统时,可通过用户隐私简档自定义其隐私要求. 用户隐私简档为二元组($K$, $A_{\min}$),其中 $K$ 表示用户所需的匿名数,$A_{\min}$ 是最小匿名区域. 在用户密集区域内 $A_{\min}$ 特别有用,因为在密集区域内,即使 $K$ 很大也达不到用户的高隐私要求,此时可通过设置 $A_{\min}$ 满足用户需求. 图1描述的是系统构架的两个主要组件: 位置匿名器和隐私感知查询处理器.

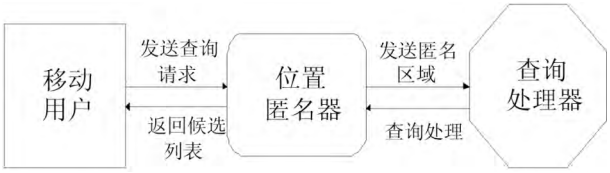



图1 系统框架图

Fig. 1 The system architecture

位置匿名器不断地获取用户更新的位置信息,模糊更新的位置信息并隐藏在满足用户隐私简档($K$, $A_{\min}$)的匿名区域内,发送匿名区域给基于位置的数据库服务器. 隐私感知查询处理器嵌入在基于位置的数据库服务器内,以匿名方式查询处理用户需求并将结果返回候选列表,用户根据自己的实际情况选择满意的结果. 候选列表的大小很大程度上是由用户设置的隐私简档($K$, $A_{\min}$)决定的,一个严格的隐私简档会得出一个很长的候选列表. 用户可以通过权衡基于位置的服务质量和其隐私要求设置隐私简档($K$, $A_{\min}$).

## 2 位置匿名器

如图1所示,位置匿名器将每个移动用户的精确点位置信息 $p$ 模糊到空间区域 $R$ 以满足每个用户隐私简档要求. 新模型的位置匿名器将满足以下4个目标.

(1) 准确度. 匿名区域 $R$ 应属于区域 $AR$,并包含满足和接近的 $KR$ 用户(即 $KR \geqslant K$,$AR \geqslant A_{\min}$);

(2) 质量. 攻击者只知道用户在匿名区域 $R$ 内,无法获知其具体位置;

(3) 效率. 匿名算法在计算上有效且可扩展. 能够应对大量的持续运动的移动用户和时空查询的实时要求;

(4) 灵活性. 每个注册用户均有指定其隐私要求和随时改变要求的能力.

### 2.1 数据结构

图2描述了基本位置匿名器的数据结构. 主要思想是采用基于网格的完整金字塔数据结构[7],将层次空间分解为 $H$ 级,其中高度 $h$ 的级别有 $4h$ 个网格单元. 每个金字塔单元表示为($cid$,$N$),其中 $cid$ 是单元标识符,$N$ 是单元格边界内移动用户的数量. 动态的维护金字塔结构用以追踪每个单元内当前移动用户的数量. 另外,追踪一个哈希表. 每个注册用户都有一个格式为($uid$,$profile$,$cid$)的条目,其中 $uid$ 是移动用户标识符,$profile$ 是用户的隐私简档,$cid$ 是移动用户所在的单元标识符. $cid$ 总是在金字塔的最低级别,如图2的阴影层所示.

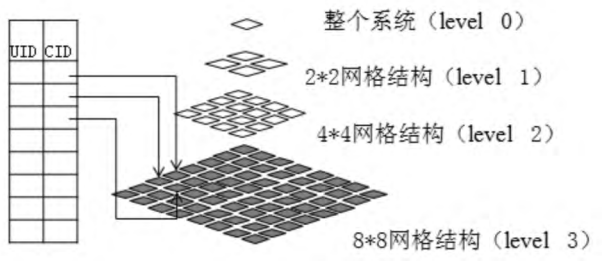



图2 基本位置匿名器

Fig. 2 The basic location anonymizer

### 2.2 维护

基于位置的应用程序,其高动态环境必然使它采用的数据结构需要持续频繁地更新. 位置更新以($uid$,$x$,$y$)的形式发送给位置匿名器,其中 $uid$ 是用户标识符,$x$ 和 $y$ 是用户新的位置坐标. 一旦位置匿名器接收到更新,散列函数 $h(x,y)$ 可以在最底层的网格层获取用户新的标识符 cidnew. 检查哈希表中的用户条目以获得其原始单元标识符 cidold. 若旧单元格标识符与新单元格标识符相同(cidold = cidnew),则不需要做任何处理; 若有变化(cidold ≠ cidnew),则执行:

(1) 更新哈希表中新的单元格标识符;

(2) 更新新、旧金字塔网格单元格中的计数器 $N$;

(3) 如果需要,把单元格计数器 $N$ 的变化发送到金字塔顶层. 若有新的注册用户,在哈希表中创建一个新条目,并将金字塔结构中所有受影响的网格单元的计数器加1. 类似地,若有用户退出,将其条目从哈希表中删除,并将所有受影响的网格单元的计数器减1.

### 2.3 匿名算法

算法1描述了一种基于网格的金字塔结构的自

下而上的匿名算法. 具体算法如下:

输入: 用户隐私简档( $K$, $A_{\min}$) 和用户当前处于网格单元的标识符 $cid$

输出: 满足用户要求的单元格区域 $Area(cid)$

算法 1　Bottom-up cloaking algorithm

Function *BOTTOM-UP CLOAKING* ( $K$, $A_{\min}$, $cid$ )

if $cid.N \geqslant K$ and $cid.Area \geqslant A_{\min}$ then

return $Area(cid)$; //if the initial area is satisfied, then initial area is returned.

end if

$cid_V \leftarrow$ The vertical neighbor cell of $cid$.

$cid_H \leftarrow$ The horizontal neighbor cell of $cid$.

$N_V = cid.N + cid_V.N$, $N_H = cid.N + cid_H.N$

if ( $N_V \geqslant K$ or $N_H \geqslant K$) and $2cid.Area \geqslant A_{\min}$ then

if( $N_V \geqslant K$ and $N_H \geqslant K$ and $N_H \leqslant N_V$) or $N_V < K$ then

return $Area(cid) \cup Area(cid_H)$;

else

return $Area(cid) \cup Area(cid_V)$;

end if　　//return the more satisfied $K$

else

*BOTTOM-UP CLOAKING*( $K$, $A_{\min}$, *PARENT* ( $cid$) ); //if the neighbor cells are not satisfied, use the father node of $cid$ to do the recursion.

end if

## 3　隐私感知查询处理器

如图 1 所示,隐私感知查询处理器嵌入在基于位置的数据库服务器内,其主要目标是提供高效、准确的服务. 存储在基于隐私的位置数据库服务器上的数据类型主要有: 公共数据和私有数据. 公共数据包括固定物体,如医院、餐馆和加油站等. 私人数据主要包含个人资料,具有非零 $K$ 或非零 $A_{\min}$ 的隐私简档的移动或固定用户的信息. 这些数据被位置匿名器隐藏在匿名区域内. 基于存储数据,通过其隐私感知查询处理器识别新模型支持的 3 种查询类型.

1) 私人查询公共数据. 例如一个人( 私人查询) 询问关于离他最近的加油站( 公共资料) . 隐私感知查询处理器没有发出用户的确切位置,而加油站的确切位置是已知的.

2) 公共查询私人数据. 例如管理员( 公共查询) 询问移动用户的数量( 私人数据) . 隐私感知查询处理器知道查询的确切信息,但不知道移动用户的确切位置.

3) 私人查询私人数据. 例如一个人( 私人查询) 询问离他最近的朋友( 私人资料) . 用户和他好友的确切位置在隐私感知查询处理器上均不可用.

最近邻查询的主要思想是计算出要发送给客户端的结果候选列表. 然后,客户在候选列表中本地评估他的查询,以获得查询结果. 通过实验证明该方法是高效、可扩展的,通过计算证明候选列表是可包容性的,即包含确切答案的最小表.

算法的主要思想是初始化选择一组可用于整个目标对象集上搜索的过滤目标对象. 不管匿名区域 $A$ 中用户的确切位置,使用过滤目标,识别空间搜索可能覆盖最近邻查询潜在答案的区域 $AEXT$. 最后,将 $AEXT$ 内的所有目标对象都作为候选名单返回给用户. 算法 2 给出了私人对公共数据进行最近邻查询的具体算法.

输入: 伪匿名区域 $A$

输出: 结果候选列表

算法 2　Private*NN* Queries over Public Data

Function *PRIVATE NN PUBLIC DATA* ( *Cloaked Area A* )

$AEXT$ is an extened area and initially set to $A$

for each vertex $v_i$ in region $A$ do

$t_i \leftarrow$ is the nearest target object to $v_i$

end for

for each edge $e_{ij} = v_iv_j$ of region $A$ do

if $t_i = t_j$ then

$m_{ij} \leftarrow$ NULL

else

$L_{ij}$ is a line connecting $t_i$ and $t_j$

$P_{ij}$ is a line that divides and is orthogonal to $L_{ij}$

$m_{ij}$ is a intersection point of $P_{ij}$ and $e_{ij}$

end if

$d_m \leftarrow Distance(t_i, m_{ij}) = Distance(t_j, m_{ij})$

$d_i \leftarrow Distance(v_i, t_i)$

$d_j \leftarrow Distance(v_j, t_j)$

$max_d \leftarrow MAX(d_m, d_i, d_j)$

Expand $AEXT$ by distance $max_d$ in $v_iv_j$ direction

end for

candidate_list$\leftarrow$ All target objects inside $AEXT$

return candidate_list

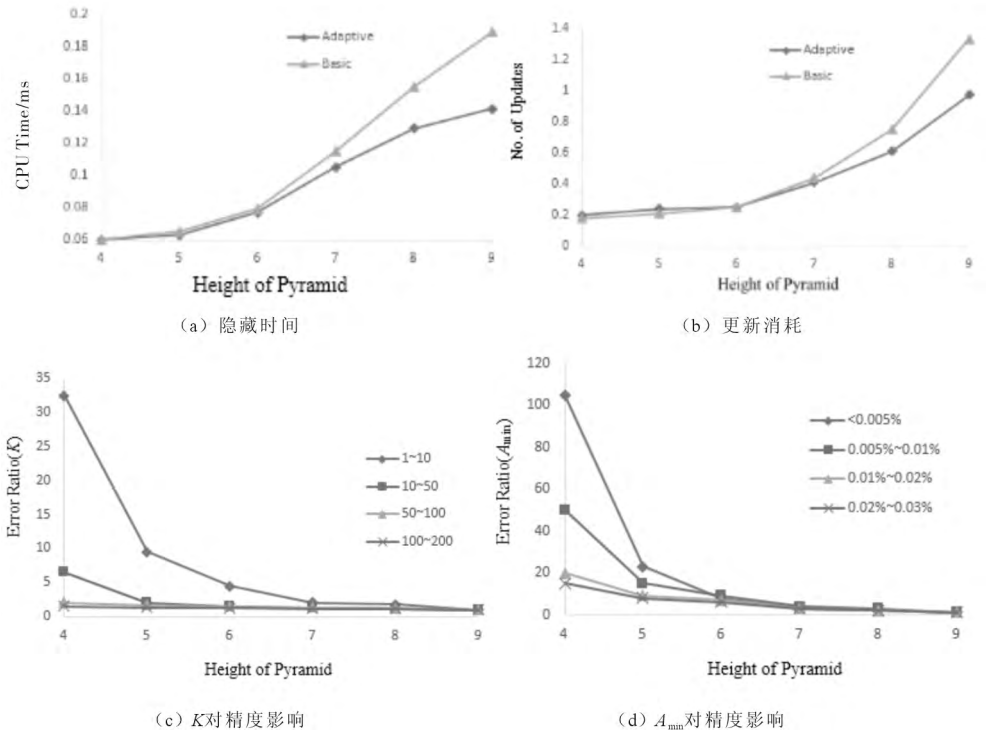
## 4 实验结果

通过实验评估新模型的两个主要组件(位置匿名器和隐私感知查询处理器)的性能来评价其框架性能.

### 4.1 位置匿名器

对基本位置匿名器和自适应匿名器在匿名时间、维护成本、准确性和可扩展性方面进行比较. 实验使用50000个注册用户的9级金字塔结构. 每个用户生成一个随机隐私简档,其中 $K$ 和 $A_{\min}$ 分别在[1-50]个用户和[0.005%,0.01%]的范围内均匀分配.

图3给出了金字塔高度对基本和自适应位置匿名器性能的影响. 图3(a)给出了金字塔高度对每个用户请求平均隐藏时间的影响(算法1). 图3(b)给出了每个位置更新所需的平均更新次数对金字塔高度的影响. 图3(c)和3(d)分别给出了关于 $K$ 和 $A_{\min}$ 的金字塔高度对匿名区域精度的影响. 基本和自适应方法都产生与算法1相同的匿名区域精度. 图3(c)中,测量精度为 $K'/K$,其中 $K'$ 是包括在匿名区域内用户的数量,而 $K$ 是用户的确切需求. 较低的金字塔水平给予要求宽松的用户非常不准确的答案. 然而,即使对于要求宽松的用户,更高的金字塔级别可以给出非常接近于最佳情况的准确匿名区域. 类似地如图3(d)中所示,测量精度为 $A'/A_{\min}$,其中 $A'$ 是计算的匿名空间区域,$A_{\min}$ 是所需的空间区域. 此外,对具有各种 $A_{\min}$ 要求的几组用户进行实验,同时将 $K$ 设置为1.




(a) 隐藏时间 (b) 更新消耗

(c) $K$对精度影响 (d) $A_{\min}$对精度影响

**图3 金字塔高度**

Fig. 3 The height of the pyramid structure

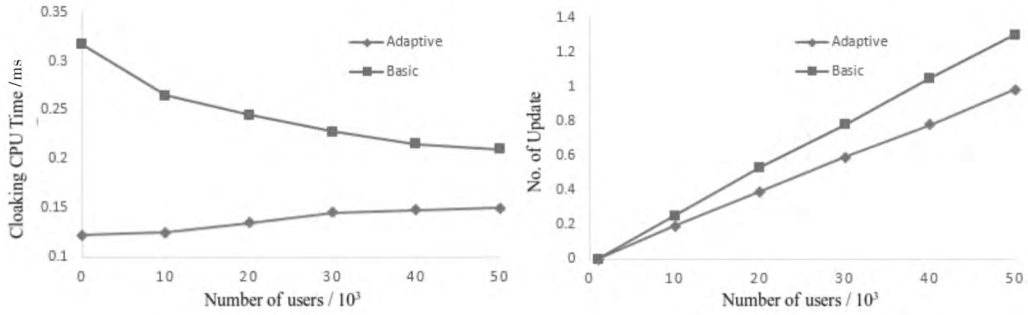
图4给出了将注册用户数量从1000改为50000时基本和自适应位置匿名器的可扩展性. 对于隐藏时间(图4(a)),随着用户数量的增加,基本位置匿名器的性能大大提高. 主要思想是,通过增加用户数量,移动用户的隐私要求将很可能在较低的金字塔级别中得到满足,即对算法1的递归调用较少. 自适应位置匿名器的情况不同,大量的用户增加了维护网格单元的数量,以适应具有各种需求的用户. 然而,适应性方法的隐藏时间总是小于基本方法的时间. 对于更新成本(图4(b)),随着用户数量的增加,由于维护单元数量较少,自适应方法的性能总是优于基本方法.

### 4.2 隐私感知查询处理器

以下研究隐私感知查询处理器对返回候选列表的大小和查询处理时间的效率和可扩展性.

图5给出了将公共目标数量从1000增加到10000时,隐私感知查询处理器的可扩展性. 对于1000目标时,使用更多的过滤器候选列表会大大减少(图5(a));对于10000目标时,使用4个过滤器导致候选列表大约只有一个过滤器返回列表的一

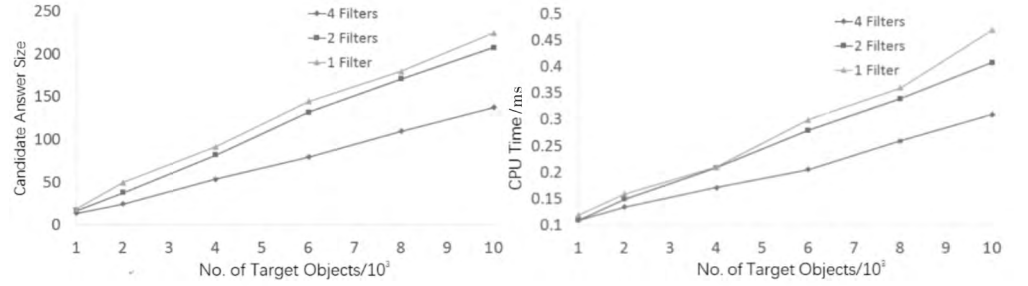
半. 关于查询处理时间( 图 5( b) ) ,4 个过滤器的计算开销是通过在搜索空间中的巨大修剪来获得候选名单,所以使用 4 个滤波器总能实现更好的性能.




(a) 隐藏时间　　(b) 更新消耗

**图 4　用户数量**

Fig. 4　Number of users




(a) 候选名单　　(b) 处理时间

**图 5　公共目标数量**

Fig. 5　Number of public target objects

## 5　结语

本文提出了一个新框架,移动用户无需泄漏其私人位置即可获取基于位置的服务,由位置匿名器和隐私感知查询处理器两部分组成. 该位置匿名器充当可信任的第三方,将每个用户的确切位置信息模糊隐藏在与用户隐私配置文件匹配的空间区域内. 位置匿名器具有高准确性、高质量、高效率和高灵活性等特性. 隐私感知查询处理器是嵌入到传统的基于位置的数据库服务器中,使其成为隐私感知具有隐藏的空间区域而不是精确的点位置信息. 此模型支持 3 种查询类型: 私人查询公共数据、公共查询私有数据和私人查询私有数据. 模型为处理这些查询提供一个框架,其返回候选列表,而不是确切的答案. 通过实验证明候选列表包含最小范围确切的答案. 通过实验评估研究了其组件,并显示在大量的移动用户和各种隐私要求下的效率、准确性和可扩展性.

## 参　考　文　献

[1] 雷建云,张镭钟. 基于 LBS 的连续查询位置隐私保护模型的动态规划算法[J]. 中南民族大学学报( 自然科学版) ,2015,34( 3) :83-87.

[2] Samarati P. Protecting respondents identities in micro data release[J]. IEEE Transactions on Knowledge and Data Engineering, 2001, 13( 6) : 1010-1027.

[3] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking[C]// ACM. International Conference on Mobile Systems Applications and Services. San Francisco: ACM, 2003: 163-168.

[4] 韩建民,林　瑜,于　娟,等. 基于位置 K-匿名的 LBS 隐私保护方法的研究[J]. 小型微型计算机系统,2014, 35( 9) :2088-2093.

[5] 王　璐,孟小峰. 位置大数据隐私保护研究综述[J]. 软件学报,2014,25( 4) :693-712.

[6] 叶阿勇,李亚成,马建峰,等. 基于服务相似性的 K-匿名位置隐私保护方法[J]. 通信学报,2014,35( 11) :162-169.

[7] Brinkhoff T. Framework for generating network-based moving objects[J]. GeoInformatica, 2002, 6( 2) :153-180.