

基于网格划分空间的位置匿名算法

邹永贵, 张玉涵

(重庆邮电大学空间信息系统研究中心, 重庆 400065)

摘要: 现有的位置匿名算法的匿名时间较长, 匿名后的空间区域较大, 严重影响查询的服务质量。为了解决这些问题, 提出了一种基于网格划分空间的位置匿名算法, 该算法基于位置 k -匿名模型, 采用网格结构划分空间后对用户位置进行位置匿名。实验结果表明, 该算法在满足用户位置隐私需求的前提下, 位置匿名时间更短, 用户的平均匿名空间减小, 从而大幅度提升用户查询的服务质量。

关键词: 基于位置的服务; 位置隐私; 网格划分; k -匿名模型

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2012)08-3059-03

doi: 10.3969/j.issn.1001-3695.2012.08.067

Location-cloaking algorithm based on grid-divided space

ZOU Yong-gui, ZHANG Yu-han

(Research Center of Spatial Information System, Chongqing University of Posts & Telecommunications, Chongqing 400065, China)

Abstract: However, the cloaking time of the existing algorithms is too long and the cloaked spatial regions are too big, which seriously affect the query service quality. To solve these problems, this paper proposed an improved location cloaking algorithm based on spatial division by grid to balance the trade-off between privacy preserving and quality of service. The improved algorithm was based on the k -anonymity model, divided the whole space by grid structure. Experimental results validate that it decreases the cloaking time, and average cloaked spatial regions get smaller with increasing privacy level.

Key words: location-based service (LBS); location privacy; grid division; k -anonymity model

0 引言

近年来, 伴随着无线网络、移动计算技术的日趋成熟和快速发展, 定位技术越来越成熟, 随时随地获得个人精确位置成为可能, 基于位置的服务 (location-based service, LBS) 得到广泛的应用, 如基于位置的紧急救援服务、信息娱乐服务、广告服务等。当用户在享受具有诱惑力的各种基于位置的服务时, 必须向基于位置的服务器提供自身的位置信息, 而攻击者可以通过定位位置传输设备、窃听位置信息传输通道等方式访问到原始的位置数据, 并计算推理获取到与位置信息相关的个人隐私信息, 使得用户的隐私安全受到严重的威胁。

一般来讲, 位置服务中的隐私保护可以分为位置隐私^[1]和查询隐私^[2]两种。例如, 用户利用自己带有 GPS 的手机查询“离我距离最近的工商银行”, 但又不想让任何人知道自己当前所在位置, 这属于位置隐私保护范畴; 又比如, 用户甚至不希望任何人知道自己提出了哪方面的查询请求, 这属于查询隐私保护范畴。任何一种隐私泄露都有可能对用户行为模式、兴趣爱好、健康状况等个人隐私信息的泄露, 而位置隐私泄露是令用户最担心的^[3]。

1 相关研究

近年来对基于位置服务中用户隐私保护的研究已经取得了一定的成果。迄今为止, 在位置匿名处理中使用得最多的模

型是位置 k -匿名模型。 k -匿名模型是由美国 Carnegie Mellon 大学 Sweeney 提出^[4], 最早使用在关系数据库的数据发布隐私保护中, 它指一条数据表示的信息至少和其他 $k-1$ 条数据不能区分。Gruteser 等人^[5]最先将 k -匿名的概念应用到位置隐私上, 提出了位置 k -匿名模型: 当一个移动用户的位置无法与周围其他 $k-1$ 个用户的位置互相区别时, 称此位置满足位置 k -匿名模型。以位置 3-匿名为例 (图 1), 将三个用户用同一匿名区域表示, 攻击者只知道在此区域中有三个用户, 具体哪个用户在哪个位置无法确定, 从而达到了保护用户位置隐私的目的。

基于位置 k -匿名模型, Gruteser 等人^[5]提出了时空匿名方法, 该方法假定用户拥有统一的 k -匿名需求, 对于每个用户的位置更新, 空间用四分树递归划分, 直至在某子空间内的用户数小于 k , 返回其上一级的子空间作为用户位置的匿名区域。该方法要求所有的用户采用统一的 k -匿名需求, 不能满足用户的个性化隐私需求, 并且返回的匿名区域过大。而且在极端情况下, 当一个位置 L 上有 k 个用户时, 尽管满足用户的 k -匿名需求, 但是位置还是会暴露。

Gedik 等人^[6]在此基础上提出了 CliqueCloak 算法, 该算法解决了所有用户拥有统一的 k -匿名需求问题。CliqueCloak 算法假定每个用户有不同的 k -匿名需求, 结合一组用户, 通过构建 Clique 图来决定哪些用户可以共享匿名空间区域, 但是这种方法并没有解决一个位置上有 k 个用户这种极端情况下的位置隐私泄露问题。此外, 由于计算 Clique 图的开销较大, 这种

收稿日期: 2012-02-03; 修回日期: 2012-03-05

作者简介: 邹永贵 (1969-), 男, 四川荣县人, 教授, 硕士, 主要研究方向为空间数据库 (zouyg99@gmail.com); 张玉涵 (1987-), 女, 河南温县人, 硕士, 主要研究方向为空间信息系统。

方法仅限于少量用户和很小的 k -匿名需求,即 k 是 $5 \sim 10$,在实际应用中很难满足用户需求。

近年来的 Bottom-up Cloaking 算法^[1]有效地解决了这些问题。它允许用户定义自己的隐私需求,满足用户个性化的隐私需要;增加匿名空间的最小值 $A_{\min}$ 参数,即返回的匿名空间必须要超过此值(可以是面积或半径等),该参数弥补了只有 k -匿名参数而引起的隐私泄露。这种方法采用不完整的金字塔数据结构,将空间分层分解为 H 层,这样一个高度为 h 的层有 4^h 个网格单元。金字塔的根高度为 0,只有一个网格单元,包含了整个空间。所谓不完整金字塔结构是指只维护那些可能会用到的网格单元。例如,如果所有的移动用户都有严格的隐私需求,金字塔的最底层不能满足任何的用户隐私协议,这样位置匿名服务器将不再维护这一层。如果用户当前所处的网格单元满足了用户的隐私需求,则返回当前网格作为该用户的匿名区域,如果不满足该用户的隐私需求,则返回该网格的父网格,直到产生一个符合该用户隐私需求的匿名空间区域。不过,这种方法的位置更新和匿名时间较长,金字塔的高度会随着用户隐私协议的改变频繁变动,维护成本较高,而且这种方法产生的匿名空间并非最优,查询服务质量还有待提高。

本文采用网格结构划分空间的方法,在 Bottom-up Cloaking 算法的基础上进行位置匿名算法改进,命名为 Grid-divide 位置匿名算法。该算法不但可以满足用户个性化的隐私要求,而且避免了位置更新慢、匿名时间长的这些缺点,同时可以获取更小的匿名空间,从而进一步提高查询服务质量。

2 位置匿名服务系统的体系结构设计

与大部分现有工作一样,本文采用中心服务器结构^[1,5-7](图2),即除了移动用户、服务提供商之外,加入可信任的第三方——位置匿名服务器。这种结构的计算开销较小,更适合实时查询。移动用户将查询请求发送给位置匿名服务器,位置匿名服务器收集用户的确切位置信息并将其转换为匿名区域,然后再转发给服务提供商。服务提供商对匿名后的位置进行查询处理,并将查询结果发送给位置匿名服务器,位置匿名服务器对查询结果求精后返回给移动用户,查询结果求精需要特定的算法,故不在本文讨论范围之内。

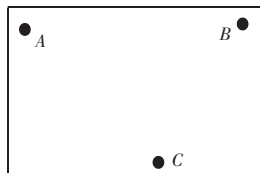


图1 位置3-匿名

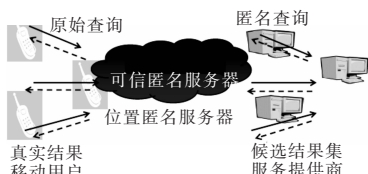


图2 中心服务器结构

移动用户在位置匿名服务器注册一个特定的隐私协议,该隐私协议被定义为一个二元组 $(k, A_{\min})$, k 表示用户希望自己是 k -匿名的,即该用户的匿名区域中至少包含 k 个用户, $A_{\min}$ 表示用户想要隐藏自己的位置信息在一个面积至少是 $A_{\min}$ 的区域。 $A_{\min}$ 的作用是为了防止在用户密集区时,即使很小的空间区域也可满足用户 k -匿名的需求时引起的隐私泄露。在极端的情况下,在一个位置 L 上有 k 个用户,虽然满足 k 值的需求,但是位置还是会暴露。一般情况下, k 值越大, $A_{\min}$ 值越大,该用户的匿名区域越大,即匿名度越高。当然,移动用户可以在任何时候改变自己的隐私协议。

位置匿名服务器收到移动用户不断更新的位置,并与该用

户的隐私协议 $(k, A_{\min})$ 匹配,在使用本文所提的位置匿名算法完成位置匿名后,将匿名空间区域发送到位置匿名服务器。在给位置信息匿名的同时,位置匿名服务器删除任何用户的 ID 来保证位置信息的匿名性。

3 Grid-divide 位置匿名算法设计

3.1 数据结构定义

在本文的算法中,整个位置空间被分为 $\alpha \times \beta$ 的网格(α 和 β 由系统自定义)。每个网格单元都表示为 (cid, N) , cid 是每个网格单元的 ID, N 是网格单元内的移动用户数量。每个移动用户移进或移出当前所在网格时向位置服务器发送位置,当位置服务器收到位置更新时,映射用户位置到当前网格。此外,建立一个哈希表,每个注册的移动用户以 $(uid, profile, cid)$ 的形式注册, uid 是移动用户的 ID, $profile$ 是用户的隐私协议,即 $(k, A_{\min})$,在本文中 $A_{\min}$ 的单位为一个网格单元, cid 是移动用户所在的网格单元的 ID。

当用户的位置发生变化,一个位置更新以 (uid, x, y) 的形式发送给位置匿名器, uid 是用户 ID, x 和 y 是用户新位置的空间坐标。当位置匿名器接收到更新后,用一个哈希函数 $h(x, y)$ 得到用户所在网格单元的 cid_{new} ,然后检测用户在哈希表登记的原始网格单元 cid_{old} 。如果 $cid_{old} = cid_{new}$,那么就不用进行更多的处理。如果 $cid_{old} \neq cid_{new}$,进行两个操作:a)在哈希表中更新用户所在的新的网格单元 ID;b)新旧网格单元的用户数量 N 都要更新。

3.2 Grid-divide 位置匿名算法描述

该算法的输入包括用户的隐私协议 $(k, A_{\min})$ 和用户当前所处的网格单元的 cid 。 k 和 $A_{\min}$ 应当比系统注册的用户数量和总的空间区域小。如果用户当前所处的网格单元满足了用户的隐私需求,即 $cid.N \geq k$, $cid.Area \geq A_{\min}$,则返回当前网格单元作为用户的匿名区域;否则,检查当前网格相邻的四个网格单元,如果当前网格和任意的邻居网格合并产生了一个符合匿名需求的空间区域,则返回这个组合的空间区域。如果和相邻网格单元的合并没有符合匿名需求的,若用户当前所处的网格单元不符合 k -匿名需求,合并当前网格和拥有最多用户的相邻网格,返回合并后的网格单元递归执行该算法直到有一个有效的网格单元产生;若用户当前所处的网格单元不符合最小空间区域面积 $A_{\min}$ 需求,合并当前网格和拥有最少用户的邻居网格,返回合并后的网格单元递归执行算法直到有一个有效的网格单元产生。

算法的伪代码如下:

```

1 Function CLOAKING ( $k, A_{\min}, cid$ )
2 if  $cid.N \geq k$  and  $cid.Area \geq A_{\min}$  then
3   return  $Area(cid)$ 
4 end if
5  $cid_N \leftarrow cid$  上边垂直相邻的网格单元编号
6  $cid_S \leftarrow cid$  下边垂直相邻的网格单元编号
7  $cid_E \leftarrow cid$  左边水平相邻的网格单元编号
8  $cid_W \leftarrow cid$  右边水平相邻的网格单元编号
9  $cid_N.N = \max(cid_N.N, cid_S.N, cid_E.N, cid_W.N) / *$  用户数量最多的网格单元编号  $*$ 
10  $cid_Y.N = \min(cid_N.N, cid_S.N, cid_E.N, cid_W.N) / *$  用户数量最少的网格单元编号  $*$ 
11 if  $(cid.N < k)$ 
12   if  $(cid.N + cid_N.N \geq k)$  and  $(cid.Area + cid_N.Area \geq A_{\min})$ 

```

```

13  return Area(cid + cidX)
14  else
15  return CLOAKING(k, Amin, (cid + cidY))
16  end if
17  if (cid.N ≥ k and cid.Area < Amin)
18  return CLOAKING(k, Amin, (cid + cidY))
19  end if
20  end function

```

4 实验结果与分析

该算法采用 Java 实现,在处理器 Intel Core 2 Duo T6400 2.0 GHz、内存 2 GB 的平台上运行。实验采用著名的 Thomas Brinkhoff 基于路网的移动对象生成器^[8],以城市 Oldenburg 的交通路网作为输入,生成输入城市路线图上移动的移动对象,目标对象均匀分布在空间。本实验取 5 000 个注册移动用户,为每个用户随机产生一个隐私协议, k 取值范围为 $[1, 50]$, $A_{\min}$ 取值范围 $[1, 50]$,本实验取 $A_{\min}$ 的单位为一个单元格。

为验证本文方法的有效性,实验将 Bottom-up 匿名算法和本文所提出的 Grid-divide 位置匿名算法使用相同的数据集进行比较。实验对算法的匿名时间和匿名区域两个指标进行评测。匿名时间是指触发查询提出请求到匿名成功的时间。取 $A_{\min}=4$ 不变,如图 3 所示,匿名时间随着 k 值的增加而增长,这是因为随着 k 值的增加,每一个用户位置需要更多的时间处理、等待才能匿名成功。匿名区域是指用户位置匿名成功后得到的空间区域的大小。取 $A_{\min}=2$ 不变,如图 4 所示,匿名区域随着 k 值的增加而增长,这是因为随着 k 值的增加,需要更大的空间才能满足用户的隐私需求。由图 4 中可以看到随着 k 值的增加,本文提出的 Grid-divide 位置匿名算法得到的匿名区域明显减小,因为在用户当前所在网格不符合条件的情况下,Bottom-up 匿名算法递归返回四倍于当前网格的父网格,而改进后的 Grid-divide 位置匿名算法递归合并和当前网格同样大小的邻居网格。由这两个指标可以看出,本文提出的算法可以有效减少匿名时间,减小匿名空间区域,从而为用户查询的服务质量提供了保障。

5 结束语

本文针对位置服务中的用户位置隐私保护问题,采用网格结构划分空间方法,在 Bottom-up 匿名算法的基础上进行改进。

实验结果表明,改进后的算法匿名时间更短,匿名区域较小,从而有利于提高用户的位置查询服务质量。

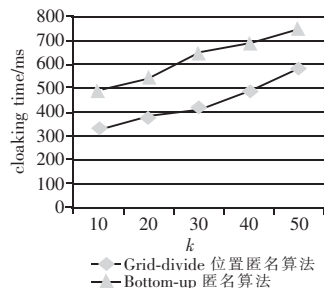


图 3 本文算法与 Bottom-up 匿名算法匿名时间比较

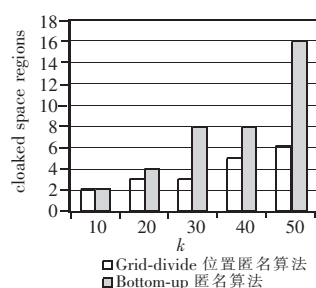


图 4 本文算法与 Bottom-up 匿名算法匿名区域比较

参考文献:

- [1] MOKBEL M F, CHOW C Y, AREF W G. The new casper: query processing for location services without compromising privacy[C]//Proc of the 32nd International Conference on Very Large Data Bases. 2006: 763-774.
- [2] CHOW C, MOKBEL M F. Enabling privacy continuous queries for revealed user location[C]//Proc of International Symposium on Advances in Spatial and Temporal Databases. Berlin: Springer-Verlag, 2007.
- [3] 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.
- [4] SWEENEY L. K-anonymity: a model for protecting privacy[J]. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10(5): 557-570.
- [5] GRUTESER M, GRUNWAL D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//Proc of International Conference on Mobile Systems, Applications, and Services. New York: ACM Press, 2003: 163-168.
- [6] GEDIK B, LIU L. Location privacy in mobile systems: a personalized anonymization model[C]//Proc of International Conference on Distributed Computing System. 2005: 620-629.
- [7] XIAO Z, MENG X, XU J. Quality-aware privacy protection for location-based services[C]//Proc of International Conference on Database System for Advanced Applications. 2007.
- [8] BRINKHOFF T. A framework for generating network-based moving objects[J]. Geoinformatica, 2002, 6(2): 153-180.