CoPrivacy:一种用户协作无匿名区域的位置隐私保护方法

黄 毅 霍 峥 孟小峰
(中国人民大学信息学院 北京 100872)

摘 要 基于位置的服务的广泛应用给人们的生活带来了极大的便利.但是用户在享受这些便利服务的同时,个人的位置隐私也面临着严重的威胁.目前,典型的位置隐私保护技术是基于中心服务器的位置 k-匿名方法.该方法容易使中心服务器成为性能瓶颈和集中攻击点,也容易造成查询处理过程的复杂化,且牺牲了用户的服务质量.文中提出了一种用户协作无匿名区域的隐私保护方法 CoPrivacy,该方法通过用户之间协作形成匿名组,匿名组内的用户用该组的密度中心代替真实位置发出查询,并增量地从服务器获得近邻查询结果.组内成员通过近邻查询结果与自身位置之间的距离计算得出精确的查询结果.CoPrivacy 在不使用匿名区域的情况下达到了 k-匿名的效果,不牺牲用户的服务质量,并且提高了匿名系统的整体性能,简化了服务提供商的查询处理过程.文中在真实数据和模拟数据集上进行了充分的实验,验证了该方法的优越性.

关键词 基于位置的服务;位置隐私;隐私保护;用户协作;增量近邻查询
中图法分类号 TP311 **DOI 号**: 10.3724/SP.J.1016.2011.01976

CoPrivacy: A Collaborative Location Privacy-Preserving Method without Cloaking Region

HUANG Yi HUO Zheng MENG Xiao-Feng
(*School of Information, Renmin University of China, Beijing* 100872)

Abstract Serious location privacy problems arise with extensive application of location-based services. Nowadays, location k-anonymity is the one of the most popular location privacy-preserving methods, it requires a trusted third party as an anonymity server which is proved to be the performance bottleneck and aim point of attacks. In addition, it complicates the query processing by requiring an anonymity region instead of a point. This paper proposes a collaborative location privacy-preserving method without anonymity server and cloaking region. Anonymity groups are formed through user's collaboration, members in the group regard density center as their locations when requiring for LBS, and acquire kNN results incrementally from the service provider. At last, group members get the precise results through computing distances between their locations and the kNN results. The proposed method achieves k-anonymity without cloaking region, the efficiency of anonymity is improved, and query processing is simplified. Extensive experimental results show advantages of the proposed method.

Keywords location-based services; location privacy; privacy-preserving; collaboration; incremental query

收稿日期:2011-07-12;最终修改稿收到日期:2011-08-16.本课题得到国家自然科学基金(60833005,61070055,91024032)、中国人民大学科学研究基金(10XNI018)、核高基重大专项基金(2010ZX01042-002-003)、高等学校博士学科点专项科研基金(200800020002)资助.
黄 毅,男,1987 年生,硕士研究生,主要研究方向为位置隐私保护.E-mail: westyi@ruc.edu.cn. **霍 峥**,女,1982 年生,博士研究生,主要研究方向为轨迹隐私保护. **孟小峰**,男,1964 年生,教授,博士生导师,主要研究领域为 Web 数据管理、移动数据管理、XML 数据管理、云数据管理等.

1 引　言

随着无线通信和移动定位技术的不断发展，特别是移动网络和 GPS 定位技术的普及和广泛应用，促使一种新的应用模式——基于位置的服务（Location Based Services，LBS）的产生和发展. 简而言之，LBS 是由位置服务提供商提供的基于用户位置的增值服务[1]，目前主要有这几种应用：基于位置的旅游信息服务（如查询“离我最近的博物馆”等）、基于位置路线导航（如查询“最近的加油站/电影院”等）、基于位置的紧急救援服务（如查询“离我最近医院”等）和基于位置的广告服务（如“向超市 10 m 范围内的客人发送优惠券”等）[2] 等. 这些查询大多都依赖 k 近邻（k-Nearest-Neighbor，kNN）查询. 基于位置服务的广泛应用给人们的生活带来了极大的便利，然而，用户在使用这些服务时，可能面临着隐私泄露的威胁. 恶意的位置服务提供商或其它针对位置服务器的攻击者根据用户位置和查询内容信息鉴别出用户身份，进而获得用户的隐私信息.

为了解决位置服务中的隐私保护问题，文献[3]最早提出了位置 k-匿名模型. 它的基本思想是在发布用户位置的时候，用一个覆盖其它 $k-1$ 个用户的匿名区域代替用户的真实位置，从而使得位置服务提供商无法从 k 个用户中鉴别出某个用户.

目前大多数基于 k-匿名模型的研究[4-8] 都采用如图 1 所示的基于中心服务器（Trusted Third Party based，TTP-based）的结构. 用户将其位置信息、查询内容和隐私需求 k 发送给中心匿名服务器（TTP），匿名服务器根据用户的隐私需求 k，将用户的精确位置扩展为包含其它 $k-1$ 个用户的匿名区域，然后再将该区域和用户的查询发送给位置服务提供商. 得到结果集后，匿名服务器根据用户的位置，从结果集里计算出满足用户查询需求的精确结果，再返回给查询用户. 然而，使用中心服务器结构存在一些问题，主要体现为以下几点：（1）中心服务器容易成为系统性能瓶颈和集中攻击点；（2）中心服务器掌握所有用户的位置信息和查询信息，如果被黑客攻击，隐私泄露严重；（3）使用中心服务器代理查询会消耗额外的计算资源和通信代价.

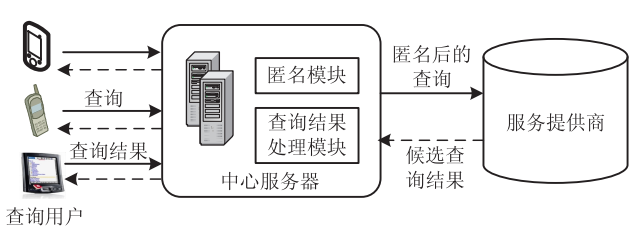


图 1　基于中心服务器的系统结构

鉴于中心服务器结构的诸多不足，越来越多的研究采用无中心服务器结构的隐私保护方法[9-13]. 大多数无中心服务器结构的方法都采取用户协作的方法来计算满足 k-匿名的区域，这样虽然避免出现性能瓶颈和集中攻击点，但基于匿名区域的查询仍然需要位置服务提供商和用户端进行大量的计算和较大的通信代价，此外，这些方法大多假设协作用户之间是可信的，如果恶意用户相互串通，其他用户的隐私将受到威胁. 文献[13]提出了 SpaceTwist 方法，用户随机选取自己真实位置附近的点作为锚点，然后使用该锚点代替自己的真实位置向位置服务提供商发起增量近邻查询，再根据用户真实位置和返回的结果进行计算，得到精确的查询结果. 虽然 SpaceTwist 避免了使用匿名区域查询造成的高计算代价和通信代价，但是 SpaceTwist 缺少用户之间协作，无法达到位置 k-匿名. 根据文献[13]的分析，攻击者通过分析用户的查询，可以将用户的位置限定在一个区域中，如果该区域只有一个用户发起查询，攻击者就有可能根据查询内容鉴别出用户，进而获得该用户的隐私.

结合用户协作结构和增量近邻查询处理的优点，本文提出了一种基于用户协作的隐私保护方法 CoPrivacy，该方法不需要中心服务器，不生成匿名区域，用户之间通过单跳和多跳协议形成匿名组. 组内用户使用该组形成区域的密度中心作为锚点，并使用该锚点代替自己的真实位置向服务提供商发起增量的近邻查询[14]，最后，每个用户根据自己的真实位置和增量近邻查询返回的结果计算得到精确的近邻查询结果. 此外，用户可以根据自己的隐私需求，指定个性化的匿名参数 k. 同时，为了达到更好的隐私保护效果，用户可以指定相对的隐私保护区域半径 s，进一步提高位置隐私保护度. 本文的贡献主要表现在以下几个方面：

（1）提出了一种新的基于用户协作的无匿名区域的隐私保护方法 CoPrivacy，不使用中心服务器，杜绝了性能瓶颈和攻击中心.

（2）以 SpaceTwist 增量近邻查询方法为基础设计了一种新的锚点的产生策略，该策略在不使用匿名区域的情况下达到位置 k-匿名，保证了匿名区域最小半径 s，提高了隐私保护度，简化了查询处理过程. 由于采用了增量近邻查询，保护位置隐私的同时

不用牺牲用户的服务质量.

(3)在用户协作的系统结构下,提出了一种通信协议,使系统中的节点可以通过较少的通信代价找到 k-匿名组.

(4)针对本文提出的方法在真实数据和模拟数据集上进行了充分的实验,验证了该方法的优越性.

本文第 2 节主要介绍位置隐私保护的相关工作;第 3 节介绍 CoPrivacy 的系统架构;第 4 节详细介绍 CoPrivacy 的实现算法;实验结果展示和分析在第 5 节;最后,第 6 节总结本文工作,展望未来研究工作.

2 背景及相关工作

从系统结构的角度来看,LBS 中的隐私保护方法主要分为两类:基于中心服务器的位置隐私保护方法和无中心服务器的位置隐私保护方法.其中无中心服务器的隐私保护方法又可进一步分为用户协作的方法和无协作的方法.下面将分别介绍一下基于中心服务器和无中心服务器的两类位置隐私保护方法.

2.1 基于中心服务器的位置隐私保护方法

基于中心服务器的隐私保护方法最早被提出来并获得了广泛的研究.文献[3]最早提出了时空匿名方法,将 k 匿名模型应用于位置隐私保护.但它假设所有用户拥有相同的隐私需求参数 k,用户无法根据自身需求指定个性化的隐私需求.此外,它独立处理每个用户的请求,伸缩性比较差.文献[4]提出了 CliqueCloak 方法,支持个性化的隐私需求参数 k,但限于计算复杂度,CliqueCloak 仅支持比较小的隐私需求参数 k,一般为 5～10.文献[7]对文献[4]的工作进行了改进,除了支持个性化的隐私需求参数 k 外,还允许用户指定可以容忍的最长匿名延迟时间和匿名空间的最大值.文献[5]提出了 casper 方法,主要关注基于匿名区域的查询处理方法.文献[6]使用有向图来表示用户之间的关系,要求用户同时被其它 k 个匿名区域覆盖,达到了更好的隐私保护效果.文献[8]提出了 PrivacyGrid 方法,将空间分割成网格,使用自顶向上的方法求取匿名区域.同时,它引入了位置 l-多样化的概念,增强了用户的隐私保护效果.

由于位置 k-匿名方法将一个点扩大为一个匿名区域,在达到位置隐私保护的同时,往往以牺牲用户的服务质量作为代价.隐私保护度越高,服务质量就越低,这也成为位置 k-匿名方法的主要缺点之一.

2.2 无中心服务器的位置隐私保护方法

无中心服务器的用户协作方法与我们提出的方法最为接近.用户通过相互协作,形成满足 k-匿名的用户组,再使用组内用户的最小边界矩形或者假位置代替用户真实位置向位置服务提供商发起查询.文献[9]提出了 P2P 空间匿名方法,用户之间通过自组织通信形成用户组,再计算组内用户的最小边界矩形,发起查询的用户从组内用户中随机选取一个邻居作为代理,由代理替用户向位置服务提供商发起查询,并将结果转发给发起查询的用户.文献[10]提出了 Prive 方法,它基于希尔伯特曲线,将所有用户信息构建成一颗分布式的 B^+ 树,根据 B^+ 树中节点信息来形成基于希尔伯特曲线的 k-匿名区域.文献[11]提出了一种基于假数据的方法,用户向其真实位置中添加高斯噪音,发送给 TTP,TTP 收到 k 个用户的位置信息后计算出一个假位置,k 个用户都使用该假位置代替其真实位置,达到 k-匿名的效果.文献[12]对文献[11]的工作进行了扩展,将基于中心服务器结构转换成基于用户协作的结构.

无用户协作的方法是指用户主动降低位置信息的质量,通常有使用假位置或加密的方法实现位置隐私保护.文献[13]提出了 SpaceTwist 方法,用户随机选取自己真实位置附近的点作为锚点,然后使用该锚点代替自己的真实位置向位置服务提供商发起增量近邻查询.如图 2 所示,SpaceTwist 维持供应空间和需求空间两个参数,其中供应空间是以锚点为圆心,包含位置服务提供商返回的所有近邻结果的圆形区域;需求空间是以用户真实位置为圆心,以用户与他已经发现的第 k 个近邻为半径的圆形区域.查询开始时,供应空间为空,需求空间为整个空间.位置服务提供商不断返回近邻查询结果,供应空间会不断扩展而需求空间会不断收缩,当供应空间完全覆盖需求空间时,用户已经找到查询的 k 个近邻,查询结束.

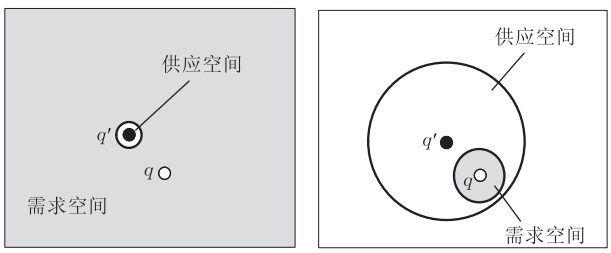


图 2 SpaceTwist 查询方法

虽然 SpaceTwist 避免了使用匿名区域查询造成的高计算代价和通信代价，但是 SpaceTwist 缺少用户协作，且无法达到位置 k-匿名，隐私保护度不高. 根据文献[13]的分析，攻击者通过分析用户的查询，可以将用户的位置限定在某个区域中，如果该区域只有一个用户发起查询，攻击者就有可能根据查询内容鉴别出用户，进而获得用户的位置隐私和查询内容隐私. 文献[15]中提出了一种使用中心服务器的改进算法，使得 SpaceTwist 达到 k-匿名的效果.

本文与上述工作的不同之处在于：本文提出的方法在不使用中心服务器的系统结构下，不使用匿名区域就可以达到位置 k-匿名的效果，避免了系统瓶颈和攻击中心，并且采用增量近邻查询方式向服务提供商发出查询，可以获得精确的查询结果. 该方法在提高了位置隐私保护度的同时并不牺牲用户的服务质量，并且简化了服务提供商的查询处理过程. 根据我们的调研，目前尚未发现类似的工作.

3　CoPrivacy 方法系统结构

随着移动设备的发展，客户端的计算能力和存储能力大幅提升，将计算模块放入客户端的系统结构变得切实可行. 本文提出的 CoPrivacy 系统结构由移动用户和服务提供商两部分组成. 移动用户通常为含有定位设备的手机或者其他终端，它包含了通信协议、位置匿名以及查询处理 3 个模块. 位置服务提供商为提供位置服务的服务器，它主要提供基于位置的近邻查询、范围查询等服务，如图 3 所示. 在我们提出的系统架构中，假定所有的移动用户都是可信的.

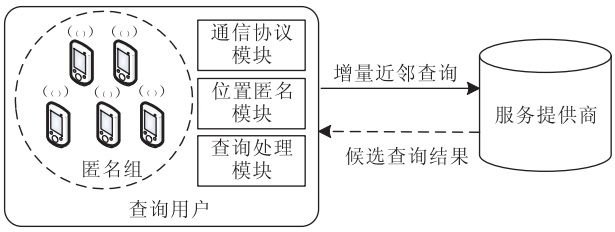


图 3　CoPrivacy 方法系统结构

在通信协议模块中，移动用户支持两种通信方式：P2P 通信和无线互联网通信. 其中 P2P 通信方式用来与其他移动用户进行自组织通信，无线互联网用来向位置服务提供商发起查询并取得查询结果. P2P 通信可以通过无线局域网(WLAN)或蓝牙等方式实现，无线互联网主要为基于移动基站的 2G 或 3G 网络. 该模块包含了一种用户协作的通信协议，移动用户通过单跳或多跳的方式向近邻的移动用户互相通信.

在位置匿名模块中，移动用户可以根据自己的隐私需求设置个性化隐私保护参数 k 和 s. 其中，k 表示相对匿名度，即用户与其他 $k-1$ 个移动用户无法区分. s 表示用户的相对匿名区域半径. 通信协议模块中得到近邻用户后，位置匿名模块根据用户设置的参数将至少 k 个用户组成一个匿名组，计算该匿名组所在区域的密度中心作为锚点.

查询处理模块中，匿名组中的用户使用该匿名组的锚点代替自己的真实位置向服务提供商发出增量近邻查询. 每个用户根据自己的真实位置和增量近邻查询返回的结果计算得到精确的近邻查询结果，并保证返回近邻查询结果覆盖的区域半径大于等于 s.

4　CoPrivacy 位置隐私保护方法

在上述系统结构下，本文提出了一种基于用户协作的无匿名区域的位置隐私保护方法 CoPrivacy. 这部分将对算法的具体处理过程做详细描述.

4.1　预备知识

定义 1(查询 Q).　用户发出的查询 Q 可以表示为以下形式：

$$Q=\{l,v,t,con,k,s\},$$

其中：

$l=(x,y)$ 表示查询发出的位置，x 表示位置的经度；y 表示位置的纬度；

v 表示查询发出时的运动速度；

t 表示查询发出的时刻；

con 表示用户输入的查询内容；

k 表示用户指定的匿名参数；

s 表示用户指定的匿名区域半径.

参数 l,v,t 可由 GPS 定位设备直接获得；参数 con,k,s 是需要用户指定的内容. 参数 k 和 s 是用户设置的隐私保护参数. k 越大，隐私保护效果越好，但需要更长的时间来发现近邻用户；s 越大，用户的隐私保护效果越好，但是增量近邻查询处理时间越长.

定义 2(k-匿名组 kAG).　可以形式化地表示为

$$kAG=\{gid,k,anchor\},$$

其中：

gid 表示该匿名组的标识符；

k 表示该匿名组中含有的成员个数；

$anchor$ 表示该匿名组的锚点，也就是每个成员发出查询时使用的位置，它可以通过计算匿名组的密度中心获得.

4.2 用户协作与位置匿名

在用户协作的隐私保护系统中,查询用户有三种状态:不在任何匿名组中;已在匿名组中但未获得锚点;已在匿名组中且已获得锚点.不在任何匿名组中的用户在发出查询时通过 P2P 单跳或多跳通信的方式发现近邻用户,如果近邻用户的数目大于 k,则形成满足 k-匿名的匿名组,此时,组内的用户已在匿名组中,但未获得锚点;然后,计算该组用户的 MBR 密度中心所在的位置作为锚点,最后以广播的形式将锚点发送给组内的所有用户,组内用户都在匿名组中,且获得了锚点.最后,用户用锚点代替自己的真实位置发出查询,查询结束后,匿名组解散,用户重新回到初始状态.具体来说,上述过程可以分为节点发现、计算锚点、广播锚点 3 个步骤.

步骤 1. 节点发现.不在任何匿名组中的查询用户 r_q 发出查询.r_q 首先生成新的组编号 gid,并将广播跳数 h 设置为 1,已发现邻居节点集 P 置为 $\{\varphi\}$,已发现节点个数 n 置为 $|P|$,即为 0,组内用户隐私需求参数 k' 为用户隐私需求参数 $r_q.k$(算法 1 第 2～5 行).然后广播节点发现的消息 FORM_GROUP,消息内容为参数 h 和 gid(算法 1 第 7 行),并等待邻居节点的响应.接收到响应消息的节点集合为 P' 后,r_q 将 n 置为 P' 中节点个数,并将 k' 更新为 P' 中所有节点最大的隐私需求参数 k(算法 1 第 8～10 行).然后再比较 n 和 $k'-1$,如果 $n>k'-1$,说明发现的节点数已经满足所有节点的匿名需求参数 k,不再广播节点发现的消息;否则先比较 P 和 P',如果二者相等,说明增加跳数无法发现更多的节点,无法通过广播节点发现的消息获得更多用户,节点发现结束,匿名失败;如果二者不相等,说明增加广播跳数可以发现更多用户,将 h 加 1,P 置为 P',继续广播节点发现消息 FORM_GROUP,等待用户响应(算法 1 第 12～16 行).节点发现完毕后,用户 r_q 将自身节点加入到已发现邻居节点集 P,已发现节点个数 n 加 1(算法 1 第 18 行).

算法 1. 节点发现 Discover-Peers.

1. //节点为 r_q
2. 生成 gid
3. 设置跳数 $h\leftarrow 1$
4. $r_q.gid\leftarrow gid$
5. 已发现的节点集合 $P\leftarrow\{\varphi\}$,发现的节点个数 $n\leftarrow|P|$,隐私需求参数 $k'\leftarrow r_q.k$
6. while $n<k'-1$
7. 广播发现节点消息 FORM_GROUP,消息内容为参数 h 和 gid
8. 接收到响应消息的节点集合为 P'
9. $n=|P'|$
10. $k'\leftarrow P'$ 中所有节点的最大 k 值
11. if $n<k'-1$ then
12. if $P=P'$ then
13. 循环结束
14. end if
15. $h\leftarrow h+1$; $P\leftarrow P'$
16. end if
17. end while
18. $P\leftarrow P\cup\{r_q\}$; $n\leftarrow n+1$
19. //计算锚点
20. $R\leftarrow P$ 中所有节点的 MBR
21. $c\leftarrow R$ 中 MBR 的密度中心点
22. $r_q.anchor\leftarrow c$
23. $r_q.gn\leftarrow n$
24. //广播锚点
25. 广播获得锚点的消息 ANCHOR_ACQUIRED,消息内容为 c, n, h 和 gid.

邻居节点 r_0 在收到节点 r_q 发送的节点发现消息 FORM_GROUP 后的处理流程如算法 2 所示.r_0 首先检查 $r_0.gid$ 是否为空,或者 $r_0.gid$ 是否等于接收到的 gid(算法 2 第 2 行),如果不满足条件,则说明用户已经加入别的组,不用响应当前接收到的 FORM_GROUP 消息,如果满足条件,r_0 首先将自己的组编号 $r_0.gid$ 置为接收到的 gid,将已发现的节点集合 T_p 置为 $\{\varphi\}$(算法 2 第 3～4 行).然后,r_0 检查接收到的广播跳数 h,如果 $h>1$,说明接收到的 FORM_GROUP 消息需要多跳广播,r_0 将 h 减 1 后继续广播 FORM_GROUP 消息(消息内容为 h 和 gid)并等待邻居节点的响应.接收到邻居节点的响应后,r_0 将响应节点 T'_p 集合设置为已发现的节点集合 T_p(算法 2 第 6～10 行).最后,r_0 将包含自身的编号 id,位置 $r_0.l$,隐私需求参数 $r_0.k$,运动速度 $r_0.v$ 和时间戳 t 的元组加入 T_p,并将 T_p 发送给请求节点 r_q(算法 2 第 11～12 行).

步骤 2. 计算锚点.用户 r_q 发现邻居节点集 P 后,首先计算 P 中所有节点的 MBR 为 R,然后计算 R 的密度中心 c 作为锚点.最后将自身的锚点 $r_q.anchor$ 置为 c,组内用户个数 $r_q.gn$ 置为 n(算法 1 第 20～23 行).

算法 2. 响应节点发现 Response-Discover-Peers.

输入:广播跳数 h,组编号 gid

1. //节点本身为 r_0,请求节点为 r_q
2. if $r_0.gid=$NULL or $r_0.gid=gid$ then
3. $r_0.gid\leftarrow gid$
4. 已发现结点集合$\leftarrow\{\varphi\}$

5. if $h>1$ then
6. $h\leftarrow h-1$
7. 广播发现节点的消息 FORM_GROUP，消息内容为参数 h 和 gid
8. 接收到响应消息的节点集合为 T'_p
9. $T_p\leftarrow T'_p$
10. end if
11. $T_p\leftarrow T_p\cup\{\langle id, r_0.l, r_0.k, r_0.v, t\rangle\}$
12. 将 T_p 发送给请求节点 r_q
13. end if

步骤 3. 广播锚点. 用户 r_q 计算得到锚点 c 后，将广播获得锚点消息 ANCHOR_ACQUIRED，消息内容为 c，n，h 和 gid，将获得的锚点广播给组内的所有用户. 邻居节点 r_0 接收到获得锚点消息 ANCHOR_ACQUIRED 的处理流程如算法 3 所示.

算法 3. 响应获得锚点 Response-Anchor.

输入：锚点 c，组内用户个数 n，广播跳数 h，组编号 gid

1. //节点本身为 r_0
2. if $r_0.gid=gid$ then
3. $r_0.anchor\leftarrow c$
4. $r_0.gn\leftarrow n$
5. if $h>1$ then
6. $h\leftarrow h-1$
7. 广播获得锚点的消息 ANCHOR_ACQUIRED，消息内容为 c，n，h 和 gid
8. end if
9. end if

r_0 首先检查自身的组编号 $r_0.gid$ 是否与接收到的 gid 相同(算法 3 第 2 行)，如果不相同则不做出响应；如果相同则将自身的锚点 $r_0.anchor$ 置为接收到的锚点 c，将组内用户个数 $r_0.gn$ 置为接收到的参数 n(算法 3 第 3～4 行). 最后检查接收到的广播跳数 h，如果 h 大于 1，说明接收到的ANCHOR_ACQUIRED 消息需要多跳广播，r_0 将 h 减 1，再广播一次获得锚点消息 ANCHOR_ACQUIRED，消息内容为 c，n，h 和 gid(算法 3 第 5～8 行).

通过上述 3 个步骤，发出查询的移动用户通过相互协作组成了匿名组，且组内的每个移动对象都得到了锚点. 此时，用户可以用锚点代替自己的真实位置发出查询.

4.3 增量近邻查询处理

用户获得锚点后，首先对比自己的隐私需求参数 k 和组内用户个数 gn. 如果 $gn\geq k$，用户所在的匿名组满足用户的 k-匿名需求，用户可以直接使用获得的锚点向位置服务提供商发起增量近邻查询，如果 $gn<k$，用户所在的匿名组不满足用户的 k 匿名需求，为了满足用户的 k 匿名需求，还需向位置服务提供商发起 k-gn 次假查询.

增量近邻查询的流程如算法 4 所示. 用户维持一个按照增量查询返回结果同用户距离 $dist(p, r_0.l)$ 顺序建立的大顶堆 W_k，用来记录目前已经发现的 k 个近邻，初始化插入 $r_0.n$ 组〈NULL，∞〉到 W_k，需求空间 γ 为中堆顶元组的距离，供应空间 τ 为 0. 用户使用锚点位置 $r_0.anchor$ 向位置服务提供商发起增量近邻查询，会不断接收的返回的 INN 查询结果. 对于每一个接收到的近邻结果 p，首先使用 $dist(p, r_0.anchor)$ 更新用户的供应空间 τ，然后判断此时用户与 p 点距离 $dist(p, r_0.l)$ 是否小于 γ，如果小于则用 p，$dist(p, r_0.l)$ 更新 W_k，用 W_k 堆顶元组的距离更新 γ. 当 $\tau>\gamma+dist(r_0, r_0.anchor)$ 并且 $\tau>r_0.s$，表示供应空间完全覆盖需求空间和用户指定的最小隐私区域，此时用户已经获得 $r_0.n$ 个最近邻 W_k，结束增量近邻查询. 由增量近邻查询处理过程可知，用户最终得到的是精确的查询结果，本文提出的方法在实现位置隐私保护的同时，并未牺牲服务质量.

算法 4. 增量近邻查询 INN-Query.

1. //节点本身为 r_0
2. $W_k\leftarrow$ 按照查询结果同用户距离顺序建立的大顶堆
3. 初始化插入 $r_0.n$ 组〈NULL，∞〉到 W_k
4. 供应空间大小 $\tau\leftarrow 0$
5. 需求空间大小 $\gamma\leftarrow W_k$ 堆顶元组的距离
6. 使用锚点位置 $r_0.anchor$ 向位置服务提供商发起增量近邻查询
7. while $\gamma+dist(r_0, r_0.anchor)>\tau$ and $\tau<r_0.s$
8. 从位置服务提供商接收到的响应包为 S
9. for S 中每一个点 p
10. $\tau\leftarrow dist(p, r_0.anchor)$
11. if $dist(p, r_0.l)<\gamma$ then
12. 用 p，$dist(p, r_0.l)$ 更新 W_k
13. $\gamma\leftarrow W_k$ 堆顶元组的距离
14. end if
15. end for
16. end while
17. 结束增量近邻查询
18. return W_k

4.4 算法分析

本节将主要对 CoPrivacy 算法的隐私保护度、服务质量以及算法时间复杂度进行分析.

4.4.1 隐私保护度与服务质量分析

CoPrivacy 算法从两个方面保证了算法的隐私度：k-匿名和用户指定最小区域 s. 即使在攻击者知晓

移动对象分布的前提下，位置 k-匿名能保证披露风险低于 $1/k$. 最小区域 s 可以保证 k 个用户的位置分布不至于过于密集. 当攻击者不知晓移动对象的分布情况时，位置 k-匿名的披露风险为 $1/ClocakArea$，其中 $CloakArea$ 表示相对匿名区域面积，即以锚点为圆心，以 s 为半径的圆形区域面积.

在服务质量方面，本文所采用的增量近邻查询是一种精确的查询方法[13]. 在保护移动对象位置隐私的同时并未牺牲任何服务质量.

4.4.2 算法复杂度分析

移动客户端的计算能力有限，所以匿名算法不能过于复杂. 在 CoPrivacy 中，仅最先发起查询的用户需要计算该组的锚点，其它客户端仅需进行与周围客户端通信、接受锚点等操作. 因此，客户端匿名算法的时间复杂度为 $O(k)$，k 为移动用户指定的匿名需求参数.

5 实 验

5.1 实验环境

算法采用 Java 实现，在 Q8400 2.6GHz 处理器、4GB 内存的 Windows 7 平台上运行，采用真实的数据集和模拟数据集两组数据进行实验. 真实数据使用北京市出租车一天中产生的 GPS 数据集，区域面积约为 25.07 km×16.78 km，以下简称为出租车数据集；模拟数据集由移动数据管理研究业界认可的 Thomas Brinkhoff 路网数据生成器[16]生成，它以城市 Oldenburg 的交通路网（区域面积为 23.57 km×26.92 km）作为输入，生成模拟的移动用户数据，以下简称为 Oldenburg 数据集. 如果没有具体说明，实验中使用数据的默认参数值如表 1 所示.

表 1 实验默认参数

参数名称	默认值
移动用户数量	4000
匿名需求参数 k	10
隐私保护区域半径 s	500 m
用户近邻查询需要的近邻个数 n	10
位置服务提供商中查询对象个数	10000

实验模拟一个半双工的网络通信信道，带宽为 1 Mbps，移动用户使用该信道进行 P2P 通信；同时假设移动用户与位置服务提供商之间使用 3G 网络通信，带宽为 2 Mbps. 移动用户之间 P2P 通信消息和位置服务提供商返回的消息都是 64 字节.

我们在真实数据集和模拟数据集上对 CoPrivacy 方法的平均响应时间、匿名成功率、用户协作平均通信消息量、增量近邻查询平均结果大小等方面进行实验，评估了算法的可伸缩性以及隐私保护度，并用 CoPrivacy 方法和基于中心服务器的 PrivacyGrid 方法进行了对比，充分证明了本文提出方法以及文献[9]提出的方法的优越性.

5.2 可伸缩性及隐私保护度

在可伸缩性实验中，使用出租车数据集和 Oldenburg 数据集评估系统平均响应时间、匿名成功率、用户协作平均通信消息量、增量近邻查询结果大小 4 个参数随移动用户个数从 1000 增加到 10000 的变化情况. 响应时间指用户发起从发现邻居形成匿名组开始到获得所需的 n 个近邻查询结果所耗费的时间；匿名成功率指成功匿名的移动用户（指匿名组内用户个数大于或者等于用户的隐私需求参数 k）同系统中全部移动用户个数的比率，需要注意的是匿名失败的情况下，通过发起假查询，用户仍然能获得不错的隐私保护效果；用户协作平均通信消息数量指移动用户通过发现邻居形成匿名组到获得锚点平均传输的消息数量；增量近邻查询结果大小指用户向位置服务提供商发起增量查询到获得精确所需的 n 个近邻总共传输的近邻个数.

如图 4 所示，从两类数据的结果来看，随着系统中移动用户数量的增加，用户查询的平均响应时间变短，并趋于平稳，匿名成功率增加，逼近 100%，用户协作通信消息量和近邻查询平均结果大小都减小并趋于平稳. 这是因为在区域大小未发生改变的情况下，随着移动用户数量增加，移动用户密度变大，CoPrivacy 方法形成匿名组并找到锚点的速度和成功率都升高了. 此外，密集的用户使得生成匿名组的

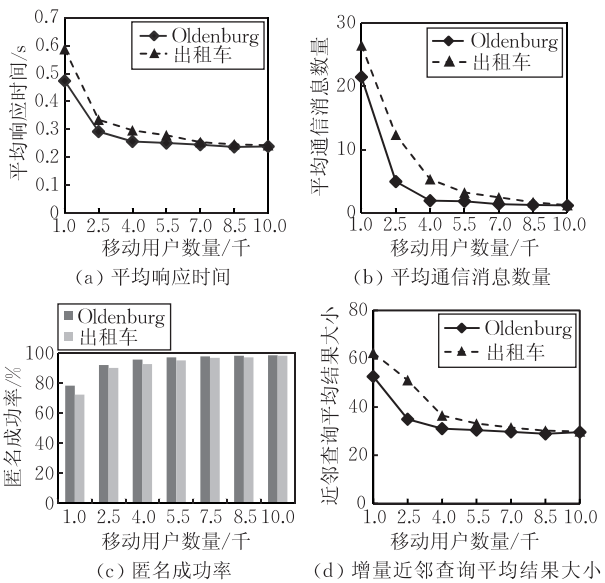


图 4 用户数量变化对系统性能的影响

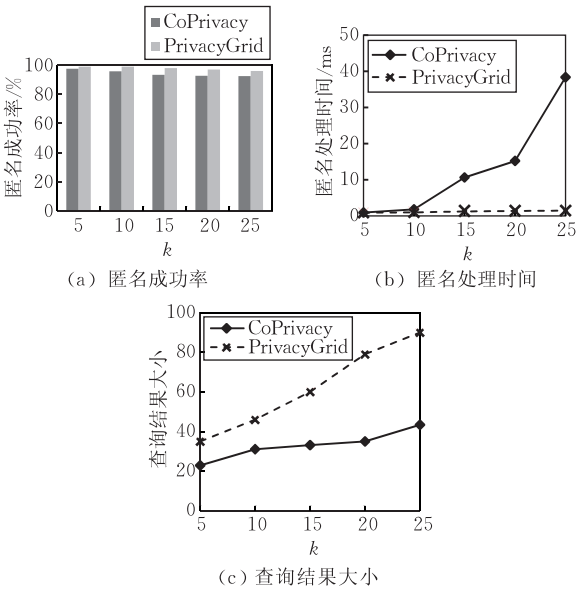
MBR 面积变小，生成的锚点位置更为接近用户的真实位置，使得增量近邻查询需要的查询结果减少. 当用户数大于 4000 后，曲线趋于平稳，这是因为用户数量大于 4000 后，移动用户已经足够密集，增加用户对用户形成匿名组获取锚点的过程没有太大影响，进而对用户增量近邻查询的过程也没有太大影响. 由上述实验结果和分析可以得出，CoPrivacy 具有良好的可伸缩性.

图 4 也表明，CoPrivacy 方法在北京市出租车的真实数据的整体运行效果会比在 Oldenburg 数据集上运行效果稍差，这是因为北京市的路网类似棋盘状的布局，出租车分布比较均匀，没有特别的密集区域，CoPrivacy 方法形成匿名组的时候需要更多的多跳通信，用户所在匿名组的 MBR 面积更大，导致了更高的通信代价和查询结果数量；而 Oldenburg 路网布局比较凌乱，会形成特定的密集和稀疏区域，更利于用户之间协作形成匿名组.

5.3　与传统匿名方法的对比

文献[8]中提出的 PrivacyGrid 方法是一种比较典型的采用中心服务器结构的位置隐私保护方法，因此，我们采用 PrivacyGrid 与本文提出的 CoPrivacy 方法在匿名成功率、匿名处理时间以及查询结果大小上进行对比. 本文将 PrivacyGrid 方法中用户的各项隐私需求参数同本方法设置一致，匿名参数 k 的取值从 5 变动到 25，其它参数为表 1 中的默认值. 使用 Oldenburg 数据集评估两种方法.

如图 5 所示，随着用户隐私需求参数 k 的增加，CoPrivacy 方法的匿名成功率有轻微的下降，匿名处理时间和查询结果大小都有比较明显的上升. 这是因为增加 k 意味着 CoPrivacy 方法需要发现更多的用户，而系统中移动用户数量没有增加，所以用户需要更长的时间来发现邻居，形成更大的匿名组，进而导致锚点与用户真实位置距离变大，增量查询结果大小变大. 图 5 同时也表明，与 PrivacyGrid 方法相比，CoPrivacy 方法可以达到几乎等同的成功率. 当 k 值增加的时候，CoPrivacy 方法需要更长的匿名处理时间，这主要是因为 CoPrivacy 方法使用 P2P 多跳通信来发现邻居用户，网络通信代价较大，而 PrivacyGrid 匿名处理在服务器端，没有网络延迟，然而 PrivacyGrid 通常无法获得精确的查询结果，而在本文提出的方法中，虽然查询时间有所增加，但用户得到的是精确的结果. 此外，在同样的 k 值下，CoPrivacy 方法查询结果大小比 PrivacyGrid 方法的查询结果小，当 k 值增加时差距更为明显，充分说明了 CoPrivacy 方法的优越性.



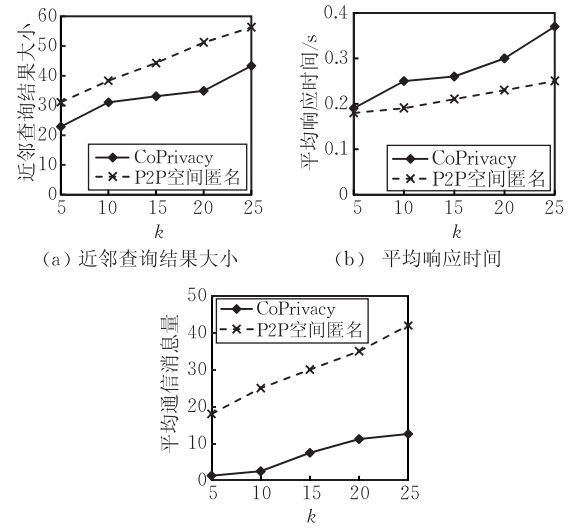
(a) 匿名成功率　(b) 匿名处理时间　(c) 查询结果大小

图 5　CoPrivacy 与 PrivacyGrid 对比

5.4　对比用户协作匿名方法

实验也将 CoPrivacy 与无中心服务器的方法进行了对比. 采用文献[9]提出的 P2P 空间匿名方法. P2P 空间匿名方法是一种在无中心服务器架构下的典型隐私保护方法[9]. 实验在 Oldenburg 数据集上进行，P2P 空间匿名方法中用户的各项隐私需求参数同 CoPrivacy 设置一致，匿名参数 k 从 5 变动到 25，其它参数为表 1 中的默认值. 分别从平均响应时间、用户协作平均通信消息量和增量近邻查询结果大小三个方面对两种方法进行评估.

如图 6 所示，随着用户隐私需求参数 k 的增加，



(a) 近邻查询结果大小　(b) 平均响应时间　(c) 平均通信消息量

图 6　CoPrivacy 与 P2P 空间匿名方法的对比

两种方法在平均响应时间、用户协作平均通信消息量和增量近邻查询结果三方面都有所上升.相比于P2P空间匿名方法,CoPrivacy方法在用户协作平均通信消息量和增量近邻查询结果方面有明显的优势,但在平均响应时间方面稍逊一筹.因为P2P空间匿名使用主动模式发现邻居用户,在形成匿名组后,不断发送消息维持匿名组信息,因此能更快地形成匿名区域,从而减小响应时间.但维持匿名组需要更多的用户协作通信,同时P2P空间匿名使用基于匿名区域的查询方法,导致更多的查询结果传输.

6 结论和展望

传统的位置隐私保护方法大多采用基于可信第三方的结构,在匿名服务器端使用满足 k 匿名的匿名区域代替用户真实位置进行查询处理,这往往需要服务器端和用户进行大量的计算,同时可信第三方容易成为系统的性能瓶颈和集中攻击目标.本文提出了一种用户协作的无匿名区域的隐私保护方法,该方法通过用户之间协作形成匿名组,匿名组内的用户用组的密度中心代替真实位置发出查询,并增量地从服务器获得近邻查询结果.组内成员通过近邻查询结果与自身位置之间的距离计算得出精确的查询结果.该方法在不使用匿名区域的情况下达到了 k 匿名的效果,不牺牲用户的服务质量,并且提高了匿名系统的整体性能,简化了服务提供商的查询处理过程.本文在真实出租车数据和模拟数据集上进行了充分的实验,证明了该方法的优越性.由于CoPrivacy方法中假定了参与协作的移动用户都是可信的,未来研究工作可以在系统中存在半可信或不可信的用户上展开.

参 考 文 献

[1] Mokbel M F. Privacy in location-based services: Start-of-the-art and research directions//Proceedings of the International Conference on Mobile Data Management (MDM'07). Mannheim, Germany, 2007: 228

[2] Solanas A, Domingo-Ferrer J, Martínez-Ballesté A. Location privacy in location-based services: Beyond TTP-based schemes//Proceedings of the International Workshop on PiLBA. Malaga, Spain, 2008, 397

[3] Gruteser M, Grunwal D. Anonymous usage of location-based services through spatial and temporal cloaking//Proceedings of the International Conference on Mobile Systems, Applications, and Services (MobiSys'03). New York, USA, 2003: 163-168

[4] Gedik B, Liu L. A customizable k-anonymity model for protecting location privacy//Proceedings of the IEEE International Conference on Distributed Computing Systems (ICDCS'05). Columbus, Ohio, USA, 2005: 620-629

[5] Mokbel M F, Chow C Y, Aref W G. The new casper: Query processing for location services without compromising privacy//Proceedings of the International Conference on Very Large Data Bases (VLDB'06). New York, USA, 2006: 763-774

[6] Xiao Z, Meng X, Xu J. Quality-aware privacy protection for location-based services//Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA'07). Bangkok, Thailand, 2007: 434-446

[7] Gedik B, Liu L. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. IEEE Transactions on Mobile Computing, 2008, 7(1): 1-18

[8] Bamba B, Liu L, Pesti P, Wang T. Supporting anonymous location queries in mobile environments with privacygrid//Proceedings of the International World Wide Web Conference (WWW'08). Beijing, China, 2008: 237-246

[9] Chow C, Mokbel M F, Liu X. A peer-to-peer spatial cloaking algorithm for anonymous location-based services//Proceedings of the Annual ACM International Symposium on Advances in Geographic Information Systems (GIS'06). Virginia, USA, 2006: 171-178

[10] Ghinita G, Kalnis P, Skiadopoulos S. PRIVE: Anonymous location based queries in distributed mobile systems//Proceedings of the International Conference on World Wide Web (WWW'07). Banff, Alberta, Canada, 2007: 1-10

[11] Solanas A, Martínez-Ballesté A. Privacy protection in location-based services through a public-key privacy homomorphism//Proceedings of the European PKI Workshop, Theory and Practice. Lecture Notes in Computer Science. Palma de Mallorca, Spain, 2007: 362-368

[12] Solanas A, Martínez-Ballesté A. A TTP-free protocol for location privacy in location-based services. Computer Communications, 2008, 31(6): 1181-1191

[13] Yiu M L, Jensen C S, Huang X, Lu H. SpaceTwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services//Proceedings of the IEEE International Conference on Data Engineering (ICDE'08). Cancun, Mexico, 2008: 366-375

[14] Hjaltason G R, Samet H. Distance browsing in spatial databases. ACM Transactions on Database Systems, 1999, 24(2): 265-318

[15] Gong Z, Sun G, Xie X. Protecting privacy in location-based services using k-anonymity without cloaked region//Proceedings of the International Conference on Mobile Data Management (MDM'10). Kanas City, USA, 2010: 366-371

[16] Brinkhoff T. A framework for generating network based moving objects. GeoInformatica, 2000, 6(2): 153-180

HUANG Yi, born in 1987, M. S. candidate. His major research interest is location privacy preservation.

HUO Zheng, born in 1982, Ph. D. candidate. Her research interests are trajectory privacy-preserving and mobile data management.

MENG Xiao-Feng, born in 1964, professor and Ph. D. supervisor. His research interests include Web data management, mobile data management, native XML database and cloud data management.

Background

This research is partially supported by the grants from the National Natural Science Foundation of China (Nos. 61070055, 91024032), the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (No. 10XNI018), National Science and Technology Major Project of Key Electronic Devices, High-end General-purpose Chips and Fundamental Software Products (No. 2010ZX01042-002-003), Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 200800020002).

Recent years, Location based services (LBS) develop fast based on the pervasive of location-aware devices, such as, GPS-enabled cell phones and PDAs, location sensors and RFID cards. Although development of LBS brings convenience to people's everyday life, it causes serious location privacy leakage, which may expose users' whereabouts or other privacy information. In order to protect LBS users' location privacy, researchers have proposed several techniques, such as dummy locations, location k-anonymity and spatio-temporal encryption. Among these techniques, location k-anonymity is the most popular one. Several techniques are proposed to achieve location k-anonymity. Most of the existing methods are based on the central server architecture, it requires a trusted third party as an anonymity server which is proved to be the performance bottleneck and aim point of attacks. In addition, it complicates the query processing by requiring an anonymity region instead of a point. In this paper, we propose a collaborative location privacy-preserving method without anonymity server and cloaking region. Anonymity groups are formed through user's collaboration, members in the group regard density center as their locations when requiring for LBS, and acquire kNN results incrementally from the service provider. At last, group members get the precise results through computing distances between their locations and the kNN results.

Location privacy-preserving is a rather young research area that has received lots of concerns recent years. We have studied location privacy-preserving since 2006, during our study, several key problems have been solved, such as location privacy-preserving against location dependence attack; location privacy-preserving for continuous query; location privacy-preserving for semi-honest users. Related research findings are published in DASFAA, CIKM, SIGSPATIAL GIS and TKDE.