doi:10.11959/j.issn.1000-436x.2016157

基于位置语义的路网位置隐私保护

陈慧[1,2,3]，秦小麟[1,3]
（1. 南京航空航天大学计算机科学与技术学院，江苏 南京 211106；2. 南京信息工程大学电子与信息工程学院，江苏 南京 210044；
3. 江苏省物联网与控制技术重点实验室，江苏 南京 211106）

摘 要：移动用户在享受基于位置的服务（LBS）的同时受到位置隐私泄露的威胁，因而提供有效的位置隐私保护策略至关重要。传统的位置隐私保护方法主要采用空间匿名的方式，若攻击者获得了更多与匿名空间相关的背景知识，尤其是与位置相关的语义信息，就会严重降低匿名效果。为了防止由位置语义分析造成的敏感位置信息泄露，并根据移动用户活动范围大多限定为道路网络的特点，提出一种基于位置语义的路网位置隐私保护方法，充分考虑了用户的个性化隐私需求，并通过实验验证了方法的可行性及有效性。

关键词：位置语义；隐私保护；路网；敏感度；普及度
中图分类号：TP309 **文献标识码**：A

Location-semantic-based location privacy protection for road network

CHEN Hui[1,2,3], QIN Xiao-lin[1,3]
(1. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
2. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China;
3. Jiangsu Key Laboratory of Internet of Things and Control Technology, Nanjing 211106, China)

Abstract: Mobile users suffer location privacy leakage threat as enjoy location-based services (LBS). Therefore, it was important to provide effective policy for location privacy protection. Conventional protection approaches were mainly based on spatial cloaking, which leads to the anonymous effectiveness suffer great reduction if the attacker obtains more background knowledge with respect to the cloaking region, especially semantic information of the location. To prevent sensitive location information leakage for the location semantics being analyzed, and consider the characteristic that most users move on road networks, a location-semantic-based location privacy protection method for road networks was proposed. The proposed method considers users' personalized privacy requirements well. The feasibility and effectiveness of the proposed method are verified through experiments for many scenarios.

Key words: location semantic, privacy protection, road network, sensitivity, popularity

1 引言

近年来，无线通信技术、定位技术及移动对象数据库管理技术取得了迅速发展。与此同时，移动计算设备的计算能力和存储能力也在不断提高，各种终端如便携机、个人数字助理、移动电话、传感器、车载终端等逐渐普及。得益于物联网的兴起，基于位置的服务（LBS, location based service）从早期的应用于紧急情况下快速定位求助者位置逐步扩展到更为广泛的应用领域[1,2]。在 LBS 中，用户通过移动定位装置获取自己的位置信息，并利用位置信息取得相关服务。如个人用户可以查询离自己

收稿日期：2015-12-19；修回日期：2016-04-11
基金项目：国家自然科学基金资助项目（No.61373015, No.61300052, No.41301047）；江苏高校优势学科建设工程基金资助项目；江苏省重大科技成果转化基金资助项目（No.BA2013049）
Foundation Items: The National Natural Science Foundation of China (No.61373015, No.61300052, No.41301047), The Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, The Important National Science and Technology Specific Project (No.BA2013049)

最近的酒店；交通管理部门可以掌握某个时段某段道路上的路况信息；商家可以向位于店铺附近的客户发送电子广告或优惠券等。

在LBS中，用户将自己的位置信息及服务请求发送给位置服务器，服务器根据接收到的位置信息对用户的服务请求进行处理。因此，用户位置信息的准确性直接决定了LBS的服务质量。尽管服务提供商声称自己是安全可靠的，但这种服务方式的自身特点决定了用户在享受服务的同时，必然会泄露位置信息。根据这些位置信息，攻击者可以进一步推理分析出用户的身份、行为、习惯等其他信息，而这些信息往往是用户不希望泄露的隐私信息。如果服务器受到恶意攻击，甚至可能服务器本身就是攻击者，攻击者从服务器得到用户的隐私信息，将造成更大的损失。Duckham等[3]指出，位置隐私是一种特殊的信息隐私，强调由个人决定何时、以何种方式、以何种程度将自己的位置信息与别人分享。而位置隐私威胁是指攻击者在未经授权的情况下，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原始的位置数据，并根据这些数据进行推理分析，从而获取与位置信息相关的个人隐私信息[4]。移动对象位置隐私保护技术能够确保用户在享受LBS的同时，保护自己的隐私信息，具有重要意义。

至今，针对LBS中的位置隐私问题，研究者们相继提出了多种解决方案[5~13]。其中，大部分隐私保护策略采用位置模糊化方法并引入*k*-匿名思想，将发出LBS请求的用户的精确位置泛化为一个模糊区域作为匿名空间提交给位置服务器，使该匿名空间中，移动用户不能与其他至少$k-1$个用户相区分。这类措施降低了用户被攻击者推理分析出准确位置的可能性，对位置隐私起到了一定的保护作用。然而，这些方法大多没有考虑到：攻击者如果掌握了足够的有关匿名空间的背景知识，就可以根据这些背景知识推理分析出更多信息，提高攻击者的攻击能力，从而严重降低匿名空间的隐私度。值得注意的是，匿名空间除了包含移动用户，还包含很多位置。而这些位置是带有语义信息的，比如位置的类型（医院、学校等）就是一种语义信息。并且，不同类型的位置具有不同的敏感度，比如与公园相比，医院是一个更为敏感的位置。此外，不同类型的位置其热门程度也不同，比如与一个门禁森严的军事基地相比，商场的热门程度更高，人流量更大，即商场的普及度比军事基地的普及度高。如果这些丰富的位置语义信息被攻击者获得，攻击者即可根据这些信息排除匿名空间中的一些位置，从而提高攻击能力。下面通过2个例子更直观地阐述该问题。

例1 假设Bob刚结束治疗从诊所出来，发出一个查询请求，查找离自己最近的咖啡店。为了保护Bob的当前位置隐私，为其构造了一个匿名空间。假设该匿名空间中包含10个位置，其中有5个都与诊所相关，那么攻击者将Bob定位到诊所的可能性明显增加。而诊所恰恰是Bob不想让别人知道的位置，因此诊所对Bob来说是敏感的位置。

例2 同样以例1中的Bob为例，假设构造的匿名空间中包含2个位置，一个是Bob所在的诊所，另一个是一座废弃的工厂。那么，攻击者在获得了匿名空间中这2个位置的语义信息后，很容易排除掉废弃的工厂，从而将Bob定位在诊所。

例1中构造的匿名空间之所以隐私度低，是因为其中包含了多个敏感度较高的位置，从而提高了攻击者推理分析的准确性。即使用户不在敏感位置，如果把用户所在位置和敏感位置放在一起构成匿名空间，一旦被攻击，则被怀疑在这些敏感位置的可能性也会增加。而这同样是用户不愿意接受的，因为用户并不希望自己被误认为在敏感位置。例2中所构造的匿名空间之所以隐私度低，是因为其中包含了普及度低的位置。

另一方面，同一类型的位置对不同的人所包含的语义信息是不同的。在上述例子中，假设Bob是一名医生，这时，诊所对于Bob并不是一个敏感位置。

由此可见，匿名空间的隐私度受位置语义的影响，而位置语义又是因人而异的。因此，本文提出的位置隐私保护方法，在构造匿名空间的过程中，将位置的语义信息作为一个重要的考察因素。

此外，现有的隐私保护策略普遍存在以下2点局限性。

1) 大部分是针对欧式空间提出的，并不适用于路网空间。而现实生活中，移动用户的活动范围更多被限制在道路上。

2) 大部分在满足用户的个性化隐私需求方面存在不足。一些方法中用户的隐私需求是固定的。另一些方法即使允许用户设定隐私需求，通常也只能对匿名空间的k值进行设定。而在实际问题中，除k值外，还有很多其他因素，比如用户所处的环

境、用户本人的特征等都与匿名空间的隐私度息息相关。

针对上述问题，本文的贡献主要有以下几方面。

1) 提出了一种隐私度量方法，该方法在度量匿名空间隐私度时将位置的语义信息作为重要的评价因素。

2) 满足用户的个性化隐私需求。

3) 提出了一种适合路网空间的位置隐私保护方法。

2　相关工作

目前已有的移动对象隐私保护方法主要分为以下几类。

1) 基于策略的方法（policy-based scheme），这类方法主要是研究用户位置隐私策略，具体探讨特定的地址信息应该何时公布给哪些对象。

2) 空间匿名方法（spacial cloaking scheme），这类方法是较早提出的一类位置隐私保护方法，其本质是用一个空间区域来表示用户的真实位置。其中，使用最多的是位置 *k*-匿名模型（location *k*-anonymity model）。该模型由美国卡内基梅隆大学的 Sweeney[14]提出，最早在关系数据库的数据发布隐私保护中得到使用。*k*-匿名模型的主要思想是使一条数据表示的个人信息与其他至少 *k*–1 条数据不能区分。Gruteser[15]将 *k*-匿名思想引入位置隐私保护，提出了位置 *k*-匿名模型。位置 *k*-匿名的主要目的是使一个移动对象的位置无法与其他 *k*–1 个移动对象的位置相区分。Mokbel[16]提出一种 Casper 方法，采用一种金字塔数据结构搜索用户生成匿名集。

3) 位置扰动方法（obfuscation scheme），这类方法的主要思想是通过在真实位置周围添加假位置或某些固定位置，或对真实位置按照某种策略进行变换，从而降低真实位置的精确度。Kido[17]通过生成假位置来降低真实位置的精确度，达到保护位置隐私的目的。Duckham[18]采用添加固定位置（如岔路口）的方法实现对真实位置的扰动。Ardagna[19]采用 3 种扰动技术（扩大范围、转移重心、缩小范围）中的一种或多种对真实位置实施扰动。

4) 基于加密的方法（cryptography-based schemes），这类方法在用户端将位置通过设定的加密策略进行完全的转换，使服务器可以处理用户的查询请求，但无法获知用户的真实位置。Ghinita[20]提出一种基于隐私信息检索（PIR, private information retrieval）的方法，无需用户信任任何第三方匿名器即能返回准确结果。但该方法仅支持隐私最近邻查询，且计算代价较高。

目前的位置隐私保护策略大多是针对欧式空间提出的，较少给出面向路网的隐私保护方案[21]。Chow[22]提出一种面向路网的位置隐私保护方法，将用户位置模糊化为若干相邻的路段，并在构造匿名集时考虑查询代价和查询质量。Palanisamy[23]提出一种利用动态假名机制混合区域在道路网络中实现混淆的匿名技术。然而，上述方法在构造匿名集时并未考虑匿名集中位置的语义信息。Damiani[24]指出传统的 *k*-匿名方法无法保护敏感的位置，由此提出了一种匿名方法解决社交网络位置分享应用中的敏感位置泄露问题。Yigitoglu[25]提出了一种适合道路网络的社交网络位置分享应用中保护敏感信息的方法。上述 2 种方法仅以位置的普及度来判断位置的敏感程度，并未考虑不同位置对不同用户的敏感程度。

针对 LBS 中的位置隐私保护问题，本文给出一种面向道路网络的位置隐私保护方法，在匿名集构造过程中引入位置的语义信息。具体地，在考虑位置普及度的同时，考虑不同位置对不同用户的敏感度，进而考察匿名区域对不同用户的隐私度。此外，允许用户设置个性化的隐私需求。

3　背景知识与问题定义

3.1　系统结构

位置隐私保护策略中采用的位置匿名体系结构主要分为 3 类：独立结构、中心服务器结构和分布式点对点结构[26]。本文采用的是中心服务器结构，如图 1 所示。该结构的主要特点在于，在客户端和位置服务器之间增加了一个可信第三方，即匿名服务器。当用户发出查询请求时，将自己的精确位置、查询内容以及隐私需求一起发送给匿名服务器。匿名服务器根据用户的隐私需求对用户的精确位置进行匿名处理，并将处理后生成的匿名位置与查询内容一起提交给位置服务器。位置服务器计算出候选查询结果，并发送给匿名服务器。匿名服务器对查询结果进行分析处理，最后将经过求精处理的查询结果返回给用户。本系统中，为了构造有效的匿名集，匿名服务器需要保存当前的地图信息以及道路信息，并实时更新道路中移动用户的信息。

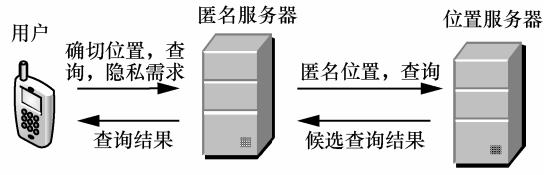


图 1 中心服务器结构

3.2 位置语义信息

本文提出的匿名算法将查询请求用户的真实位置隐藏于多条邻接的路段中，即由多条邻接路段构成匿名集 *RS*，且该集合满足用户的个性化隐私需求。整个道路网络表示为一个连通图 $G=(V,E)$，其中，V 是顶点的集合，表示路段的端点以及邻接路段的交点；E 是边的集合，表示路段。每一条路段 $Seg_i(sid,s,t)\in E$ 是道路网络中的一条边，其中，*sid* 是路段的编号，*s* 和 *t* 分别表示路段的起点和终点。

本文主要从位置的语义信息出发，探讨位置语义对位置隐私的影响，下面给出所涉及的语义信息定义。

定义 1 位置。用 $pos(lid, sid, x, y, tp)$表示道路上的位置，其中：*lid* 是位置的编号，*sid* 是该位置所在路段的编号，(x, y)表示该位置的坐标，*tp* 表示该位置所属的类型。所有位置划分为 *n* 种类型，则 $TP=\{tp_1, tp_2,\cdots, tp_n\}$为 *n* 种位置类型的集合。

定义 2 位置普及度。系统为每个位置类型 $tpi\in TP$ 设置一个普及度 pop_{tp_i}，用来表示该位置的热门程度，则 $POP=\{pop_{tp_1}, pop_{tp_2},\cdots, pop_{tp_n}\}$是所有位置类型所对应的普及度的集合。

定义 3 位置敏感度。用户为每个位置类型 $tp_i\in TP$ 设置一个敏感度 sen_{tp_i}，用来表示该位置相对该用户的敏感程度，则 $SEN_u=\{sen_{tp_1}, sen_{tp_2}, \cdots, sen_{tp_n}\}$是所有位置类型相对用户 *u* 所对应的敏感度的集合。

3.3 用户隐私需求

本方法允许用户定制个性化的隐私需求，主要有以下 3 方面。

1) 匿名集中的移动用户数量(*RS.UN*)。要求能够使用户在匿名集中至少与其他 *k*–1 个用户无法区分。该需求源于经典的 *k*-匿名需求，是位置隐私保护中最常用也是最基本的需求。

2) 匿名集中的路段数量(*RS.SN*)。要求匿名集中路段的数量不能少于用户设定的数值。如果匿名集中的路段数量过少，比如仅包含一条路段，那么即使匿名集中有多个移动用户，攻击者的攻击难度也显著降低了。

3) 匿名集的隐私度。要求匿名集中尽可能多地包含对该用户而言不敏感的位置。

此外，匿名集的普及度也是一个重要的隐私需求，要求匿名集中尽可能多地包含普及度高的位置，即人们经常光顾的位置。普及度较低的位置很容易被攻击者排除，从而降低攻击难度。为方便分析，本系统中的普及度由系统自定义，因此不属于个性化隐私需求范畴。

根据上述分析，对用户的隐私需求做如下定义。

定义 4 隐私需求。对一个发出匿名请求的用户 *u*，其隐私需求用 $PR_u(UN, SN, SEN_u)$表示。其中，*UN* 是用户自定义的匿名集中移动用户数量下限；*SN* 是用户自定义的匿名集中路段数量下限；SEN_u 是用户自定义的一组位置敏感度，根据个性化的隐私需求，相同类型的位置对不同用户的敏感度是不同的。

3.4 问题定义

在本文系统中，对攻击者做出如下假设：1)攻击者拥有地图和路段信息；2) 攻击者拥有移动用户的位置信息，但不知道移动用户的身份；3) 攻击者可以获知路段上移动用户的数量；4) 攻击者没有用户的背景信息。此外，本文假设匿名服务器对用户来说是可信的。

因此，本文需要解决的问题归结为根据用户的个性化隐私需求，以及与位置语义信息相关的普及度、敏感度等隐私需求，将用户的真实位置隐藏在由多条相邻路段构成的集合中，为用户构造一个匿名集 *RS*。

4 基于位置语义的匿名方法

4.1 隐私度量

根据上述分析，显然，匿名集中的位置所带有的语义信息直接影响匿名集的隐私度。因此，在度量一个匿名集的隐私度时，除了考虑匿名集所包含的移动对象数量以及路段的条数，还要考虑该匿名集的语义信息。本文分别用区域普及度和区域敏感度来表示位置的语义信息对匿名集的影响。

定义 5 区域普及度。一个区域 *reg* 的普及度 POP_{reg}，由该区域中所包含位置的普及度确定，其值为

$$POP_{reg}=\sum_{i=1}^{|TP|}\frac{|\,pos.tp=tp_i\,|}{NumPos(reg)}pop_{tp_i} \tag{1}$$

其中，$|TP|$为区域 *reg* 中所含位置类型的总数；*NumPos*(*reg*)为区域 *reg* 中所含位置的数量。

定义 6　区域敏感度。一个区域 *reg* 的敏感度 SEN_{reg}，由该区域中所包含位置的敏感度确定，其值为

$$SEN_{reg}=\sum_{i=1}^{|TP|}\frac{|\ pos.tp=tp_i\ |}{NumPos(reg)}sen_{tp_i} \tag{2}$$

式（1）和式（2）共同给出基于位置语义的匿名区域隐私度量标准。

定义 7　区域隐私度。一个区域 *reg* 的隐私度 PRM_{reg} 由区域普及度和区域敏感度共同确定，其值为

$$PRM_{reg}=\frac{POP_{reg}}{SEN_{reg}} \tag{3}$$

4.2　基于位置语义的匿名集构造

根据用户定义的隐私需求选择合适的相邻路段构造匿名集，主要解决的问题是如何确定哪些路段是可选的。本文采用边选择边判断的策略，其主要思想是：每次在候选路段中选择一条路段，使当前匿名集关于区域隐私度条件是最优的；每选择出一条路段，判断当前匿名集是否符合隐私需求。

匿名集由包含匿名请求用户所在路段在内的若干相邻路段构成，因此，从用户所在的路段出发，将该路段选为匿名集 *RS* 中的第一条路段。每选出一条路段，判断当前匿名集是否满足用户定义的隐私需求。若满足，则将当前匿名集作为最终的匿名集返回；若不满足，则找出当前匿名集中所有路段的相邻路段，在这些路段中选择一条最优路段加入当前匿名集。这里，所谓最优路段是指由该路段构成的匿名集关于区域隐私度条件在候选路段中是最优的。当一个匿名区域的普及度越高，同时敏感度越低时，这个匿名区域的隐私度就越高。显然，这里选出的路段应使当前匿名集的区域隐私度 PRM_{RS} 最大。根据上述原则，每次向当前匿名集增加一条路段，直至匿名集满足隐私需求。

算法 1 给出了最优路段选择算法伪代码。该算法的输入参数为当前匿名区 *CRS*、相邻路段集合 *SEGS* 以及位置敏感度集合 *SENS*。算法 1 首先对相关参数进行初始化（算法 1 第 1）行）；然后将每条相邻路段分别加入当前匿名区，计算合并区域的区域普及度、区域敏感度和区域隐私度（算法 1 第 2）~6）行），并记录区域隐私度最大的路段作为当前最优路段（算法 1 第 7）行）；最后返回最优路段（算法 1 第 9）行）。

算法 1　最优路段选择算法（OPTSEG）

输入：当前匿名区 *CRS*，相邻路段集合 *SEGS*，位置敏感度集合 *SENS*

输出：最优路段 *BESTSEG*

1) *BESTSEG* =*Φ*; MAX=0;

2) for each *Seg* in *SEGS*

3)　　将 *Seg* 赋值给 *e*;

4)　　$PR=POP_{(CRS\cup e)}$;

5)　　$SR=SEN_{(CRS\cup e)}$;

6)　　$MR=\frac{PR}{SR}$;

7)　　用 MAX 记录 *MR* 的当前最大值，并将对应的 *e* 赋值给 *BESTSEG*;

8) end for

9) return *BESTSEG*

算法 2 给出了基于位置语义的匿名集构造算法伪代码。该算法的输入参数为发起匿名查询请求的用户 *U*、用户请求的查询 *Q* 以及由用户定义的隐私需求 PR_u。在隐私需求 PR_u 中，用户根据自己的实际情况，定义了匿名集中至少需包含的移动用户数量 $PR_u.UN$、至少需包含的路段条数 $PR_u.SN$，更重要的是给出每种位置类型相对自己的敏感度 $PR_u.SEN_u$。算法 2 首先将发出请求的用户所在的路段放入匿名集 *RS*（算法 2 第 3）行），如果此时 *RS* 不满足用户隐私需求中的 $PR_u.UN$ 和 $PR_u.SN$，则循环执行如下步骤直至 *RS* 满足 $PR_u.UN$ 和 $PR_u.SN$ 条件为止（算法 2 第 4）行）：1）将当前 *RS* 中所有路段的相邻路段放入集合 *R* 中（算法 2 第 5）行）；2）对 *R* 中的每一条路段 e_i，计算区域 $RS\cup e_i$ 的隐私度，选择区域隐私度最大的一条路段加入匿名集 *RS*（算法 2 第 6）、7）行）。

算法 2　基于位置语义的匿名集构造算法（LSBASC）

输入：用户 *U*，查询 *Q*，隐私需求 PR_u

输出：匿名集 *RS*

1) *RS*=*Φ*;

2) 将用户所在的路段赋值给 *BestEdge*;

3) 将 *BestEdge* 放入匿名集 *RS*;

4) while *NumUser*(*RS*)< $PR_u.UN$ or *NumSeg*(*RS*) < $PR_u.SN$

5)　　*R*=*FindEdges*(*RS*);

6) *BestEdge*= *OPTSEG*(*RS*, *R*, $PR_u.SEN_u$);

7) 将 *BestEdge* 放入匿名集 *RS*;

8) *R*=*Φ*;

9) end while

10) return *RS*

下面以用户 User 为例阐述上述具体过程。为分析方便，这里仅给出 6 条路段，如图 2 所示。路段旁边用一对值(*Seg*$_i$, *num*)表示路段的编号以及该路段上当前的移动用户数量。图中 User 所在的位置由箭头标识。设置 *TP*={tp_1, tp_2, tp_3, tp_4}，其中，tp_1=医院，tp_2=酒吧，tp_3=商场，tp_4=学校。设定 pop_{tp_1}=0.3，pop_{tp_2}=0，pop_{tp_3}=0.4，pop_{tp_4}=0.3。用户定义的隐私需求为 PR_u(*UN*, *SN*, SEN_u)，其中，$PR_u.UN$=50，$PR_u.SN$=3，$PR_u.SEN_u$ ={0.5, 0.3, 0.2, 0}。根据算法步骤，首先将 User 所在的路段 Seg_1 加入匿名集 *RS* 中，由于 *NumUser*(*RS*)=10，*NumSeg*(*RS*)=1，不满足隐私需求，继续执行算法步骤：将与 Seg_1 邻接的路段 Seg_2、Seg_3 和 Seg_4 加入集合 *R*，根据区域隐私度量标准分别计算得到 $PRM_{(RS\cup Seg_2)}$=1.263，$PRM_{(RS\cup Seg_3)}$=1.87，$PRM_{(RS\cup Seg_4)}$=1.133，根据计算结果选择 Seg_3 加入匿名集 *RS*。此时，由于 *NumUser*(*RS*)=40，*NumSeg*(*RS*)=2，仍然不满足隐私需求，继续执行算法选择邻接路段。将与 Seg_1 和 Seg_3 邻接的路段 Seg_2 和 Seg_4 加入集合 *R*，计算得到 $PRM_{(RS\cup Seg_2)}$=1.737，$PRM_{(RS\cup Seg_4)}$=1.733，因而将 Seg_2 加入匿名集 *RS*。此时，*NumUser*(*RS*)=60，*NumSeg*(*RS*)=3，已满足隐私需求，算法结束，返回 *RS*={Seg_1, Seg_2, Seg_3}作为用户 User 的匿名集。

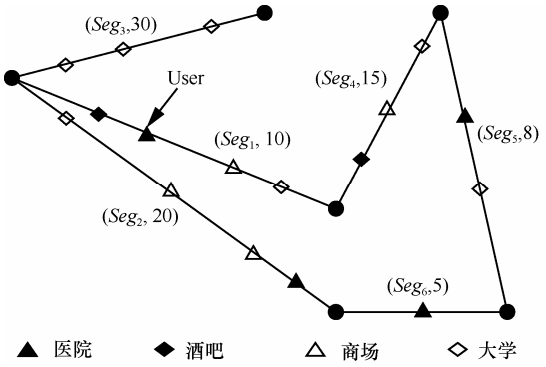


图 2 算法 1 实例

本文提出的匿名方法主要考虑位置语义对匿名空间隐私度的影响，将位置语义信息作为主要因素构造匿名空间。因此，匿名集的隐私度主要由区域隐私度决定，而区域隐私度仅与匿名集中位置语义有关，对服务质量没有直接影响。在匿名集构造过程中所考虑的 *k* 值和匿名集大小，仅为满足用户隐私需求而设置的限值，并不是影响服务质量的主要因素。值得注意的是，本方法中匿名集的构造是一个路段数量不断增加的过程，而路段数量直接决定候选集大小，最终影响服务质量。因此，在具体操作过程中，需对路段数量设置上限。

5 实验及结果分析

本文算法实验硬件环境为 3.20 GHz 的 Intel 四核 CPU，4 GB 内存。操作系统平台是 Microsoft Windows 7 Professional。

5.1 实验数据集和参数设置

实验数据采用德国 Oldenburg 市的公路网络数据[27,28]，共包括 6 105 条道路，7 035 个顶点。利用 Brinkhoff[29]的基于网络的移动对象生成器生成了 10 000 个均匀分布的移动用户，为了评估算法的匿名性能，用户数考察范围为 5 000~25 000。同时，通过修改数据生成器，在拟生成的位置数据属性中增加一维语义信息，生成了 4 种类型（包括医院、酒吧、商场及学校）的 10 000 个在公路网络中均匀分布的感兴趣点（位置）。另外，实验设置了 1 000 个发出匿名请求的移动用户。

用户发出的隐私需求包括 3 个参数，即（$PR_u.UN$, $PR_u.SN$, $PR_u.SEN_u$）。考虑实际隐私保护中，隐私需求的路段数不可能无限制增加，实验中需要设置最大值，表示为 $PR_u.SN_{\max}$。所有实验参数设置见表 1 所示。

表 1 参数设置

参数	默认值	评估范围
用户数	10 000	5 000~25 000
$PR_u.UN$	25	15~35
$PR_u.SN$	6	3~15
$PR_u.SN_{\max}$	20	
$PR_u.SEN_u$	系统随机生成	
感兴趣点	10 000	
发出匿名请求用户数	1 000	

5.2 算法性能评价标准

实验从匿名成功率、平均匿名执行时间、相对匿名度及相对空间粒度 4 个方面对提出算法进行评价。

1) 匿名成功率，即算法成功匿名的消息数在所有移动用户提出的匿名请求的消息数中所占的比

例[26]。该指标反应位置匿名算法对匿名请求的响应能力，匿名成功率越高，匿名算法越好。

2) 平均匿名执行时间，即在统计平均意义上，位置匿名器对用户真实位置进行匿名处理所花费的时间。该指标能够用来衡量位置匿名算法的运行效率，平均匿名时间越短，算法匿名执行效率越高。

3) 相对匿名度，指匿名执行后匿名集中包含的用户数量与用户位置隐私要求中的用户数量（$PR_u.UN$）的比值[26]，即用公式表示为

$$\text{相对匿名度}=\frac{RS.UN}{PR_u.UN} \tag{4}$$

在匿名成功的情况下，相对匿名度的值大于等于 1。一般地，随着相对匿名度值的增加，匿名效果提高。

4) 相对空间粒度，指匿名器针对用户匿名请求所限定的可容忍的最大匿名路段数 $PR_u.SN_{\max}$ 与匿名执行后匿名集中包含的路段数的比值[26]，用公式表示为

$$\text{相对空间粒度}=\frac{PR_u.SN}{RS.SN} \tag{5}$$

相对空间粒度越大，意味着在满足隐私需求的前提下，匿名空间越小，更接近最优解。因此，相对空间粒度越大越好。

5.3　实验结果分析

1) 匿名成功率。图 3 给出了匿名成功率相对于道路网总的移动用户数、隐私需求的移动用户数量以及隐私需求的路段数的变化情况。图 3(a)结果显示，道路网移动用户数越多，越有利于匿名执行，匿名成功率越高。这不难理解，因为用户数量越多，分布在各个道路上的用户也越多，隐私需求越容易满足，匿名成功率得到提高。反之，当用户数量少到一定程度，导致各个道路上的用户很少，在执行匿名请求过程中，当匿名集中路段数达到可容忍的最大限度时，很难保证匿名集中用户数同时满足要求，导致匿名失败。

在图 3(b)中，随着隐私需求的用户数 $PR_u.UN$ 的增加，匿名成功率下降。当道路网用户数在 10 000，$PR_u.UN$、$PR_u.SN$ 和 $PR_u.SN_{\max}$ 分别设定为 30、6 和 20 情况下，匿名成功率不到 50%。可见，隐私需求的用户数不宜设置过大。而从图 3(c)中可以看出，无论是道路网用户数相对较多还是相对较少，隐私需求的路段数 $PR_u.SN$ 的变化并未带来匿名成功率的较大变化。因此，为了减小匿名执行计算负担，从匿名成功率角度考虑，$PR_u.SN$ 不需要设定得过大。

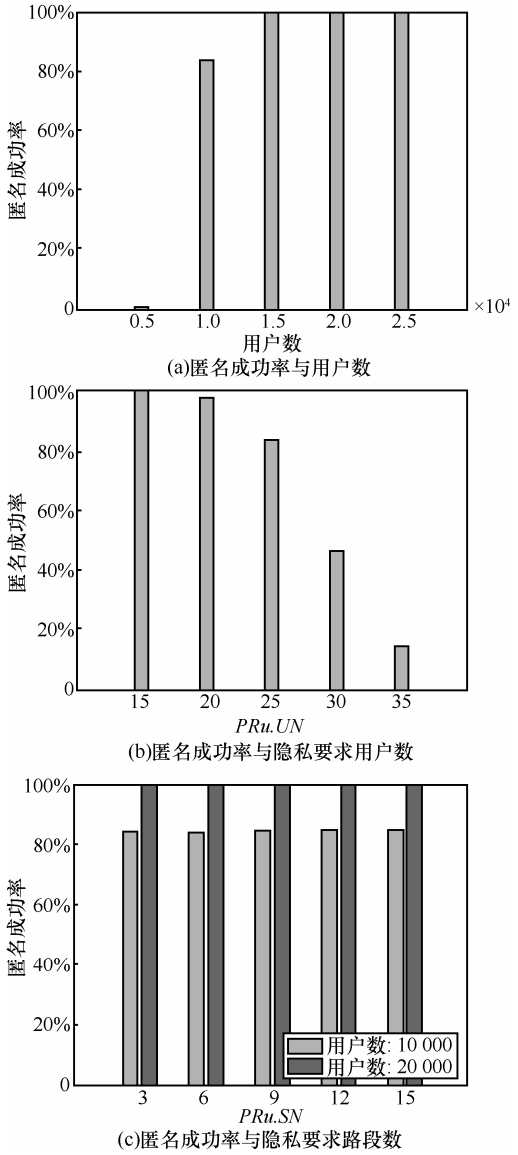


图 3　匿名成功率

2) 平均匿名执行时间。根据图 4，在道路网上的移动用户数一定的情况下，随着隐私需求用户数 $PR_u.UN$ 的增加或者隐私需求路段数 $PR_u.SN$ 的增加，平均匿名执行时间将增加。进一步，从图 4(b)可知，这种平均匿名执行时间增加的程度随着道路网上的用户数的增多而显得更为明显。其原因在于：当道路网上的用户数明显较多时，算法在判断邻近路段的选择所涉及的区域匿名度的计算随着隐私需求路段数的增加而需耗费更多的时间。另一方面，图 4 显示，在相同隐私需求条件下，即隐私需求用户数相同、隐私需求路段数相同，道路网上移动用户数量越

大，平均匿名执行时间越短。原因在于：当道路网上的移动用户数量越大，在同等路段选择时，隐私需求的用户数更容易满足，节省了判断计算时间。

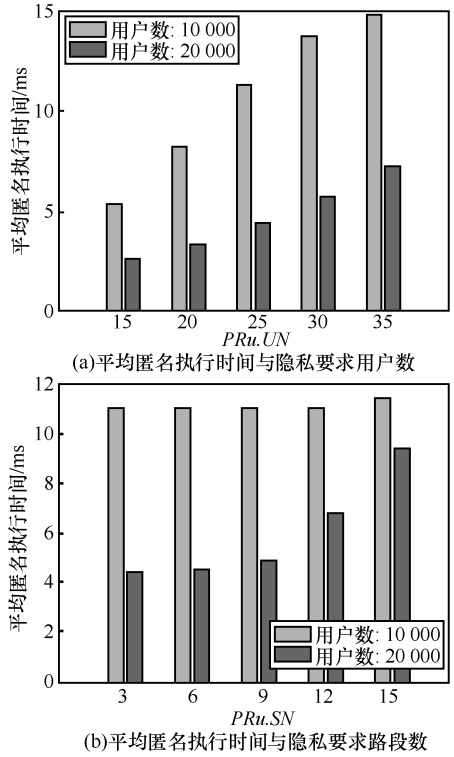


图 4　平均匿名执行时间

3) 相对匿名度。图 5 给出了相对匿名度与隐私需求用户数 $PR_u.UN$ 的关系。图中结果显示，随着 $PR_u.UN$ 的增大，相对匿名度略微减小，但在匿名成功的情况下，均大于等于 1。而在匿名失败的情况下，相对匿名度小于 1，这正是图中当道路网上的移动用户数为 10 000 时，$PR_u.UN$ 为 30 和 35 所对应的情形。因为对于这 2 种情形，匿名失败的情况可能较多，而被统计在内。同时可看出，随着道路网上的用户数增加，算法的匿名效果提高。

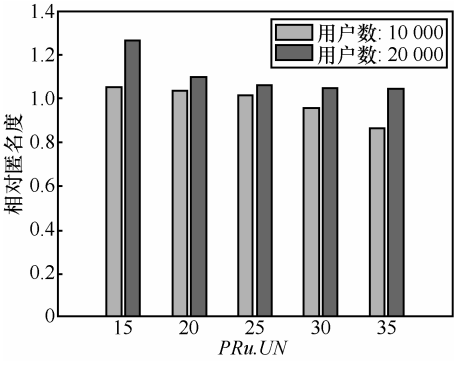


图 5　相对匿名度与隐私需求用户数 $PR_u.UN$ 的关系

4) 相对空间粒度。图 6 给出了相对空间粒度与隐私需求用户数 $PR_u.UN$ 的关系。从图 6 可以看到，随着 $PR_u.UN$ 的增大，相对空间粒度同样呈现下降趋势，且变化较明显。原因在于：$PR_u.UN$ 的增大将直接导致匿名集中路段数的增加，从而降低相对空间粒度。同时，图 6 显示，在较少的隐私需求用户数条件下能够获得较高的相对空间粒度，意味着适当低的隐私需求用户数能够获得较优的匿名空间。另一方面，道路网上移动用户数量的增加也能够带来更大的相对空间粒度。

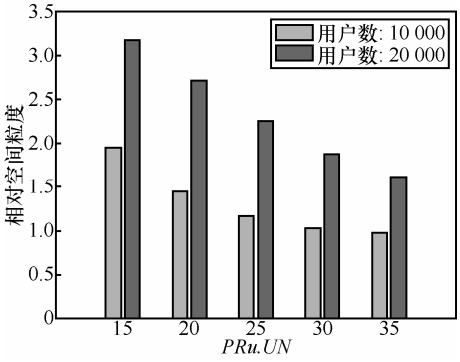
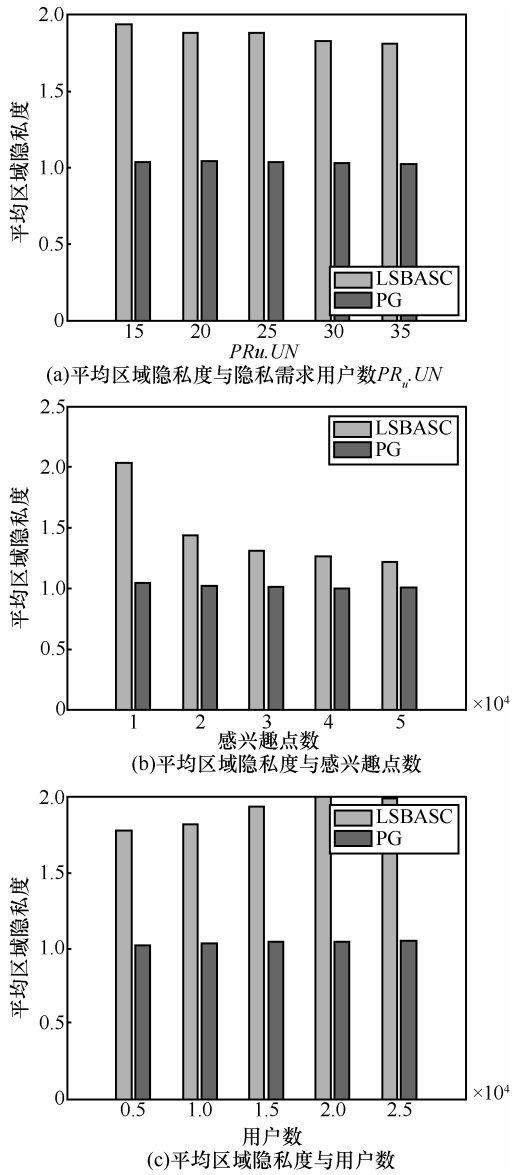


图 6　相对空间粒度与隐私需求用户数 $PR_u.UN$ 的关系

此外，选取了 Chow[22]针对路网的 LBS 隐私保护提出的匿名集构造方法（PG 算法），通过考察平均区域隐私度，进行了性能比对。图 7 给出了平均区域隐私度与隐私需求用户数 $PR_u.UN$、感兴趣点数及用户数的关系。由图 7 可知，在隐私需求用户数 $PR_u.UN$、感兴趣点数及用户数变化的情况下，本文 LSBASC 算法构造的匿名区域其隐私度均高于 PG 算法构造的匿名区域隐私度。原因在于：PG 算法没有考虑位置语义在构造匿名集过程中的影响，而本文 LSBASC 算法在构造匿名集过程中充分考虑了位置普及度和敏感度，使构造的匿名区域的隐私度明显提高。

6　结束语

本文给出了一种基于位置语义的路网位置隐私保护方法，该方法解决了现实道路网上的移动用户在发出 LBS 请求时如何进行个性化位置隐私保护的问题。与传统的位置隐私保护策略不同的是，本算法考虑了位置语义对匿名效果的影响，在定义了区域敏感度、区域普及度的基础上，给出了一种基于位置语义的匿名区隐私度度量，并以此为标准构造匿名集，使所构造的匿名集满足用户的个性化



(a)平均区域隐私度与隐私需求用户数$PR_u.UN$

(b)平均区域隐私度与感兴趣点数

(c)平均区域隐私度与用户数

图 7　平均区域隐私度

隐私需求。通过基于真实数据及模拟数据的大量实验，结合与其他匿名方法进行的对比实验，证明了本文方法的可行性及有效性。

参考文献：

[1] CHEN X, PANG J. Protecting query privacy in location-based services[J]. GeoInformatica, 2014, 18(1): 95-133.

[2] 周傲英, 杨彬, 金澈清, 等. 基于位置的服务：架构与进展[J]. 计算机学报, 2011, 34(7): 1155-1171.
ZHOU A Y, YANG B, JIN C Q, et al. Location-based services: architecture and progress[J]. Chinese Journal of Computers, 2011, 34(7): 1155-1171.

[3] DUCKHAM M, KULIK L. Location privacy and location-aware computing[J]. Dynamic & Mobile GIS: Investigating Change in Space and Time, 2006, 3: 35-51.

[4] MOKBEL M F. Privacy in location-based services: state-of-the-art and research directions[C]//The 8th International Conference on Mobile Data Management (MDM'07). Mannheim, Germany, c2007:228.

[5] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//The First International Conference on Mobile Systems, Applications, and Services. c2003: 31-42.

[6] MOKBEL M F, CHOW C Y, AREF W G. The new casper: Query processing for location services without compromising privacy[C]// VLDB. c2006:763-774.

[7] KALNIS P, GHINITA G, MOURATIDIS K, et al. Preventing location-based identity inference in anonymous spatial queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(12): 1719-1733.

[8] MASCETTI S, BETTINI C, WANG X S, et al. ProvidentHider: an algorithm to preserve historical *k*-anonymity in LBS[C]//The 10th IEEE International Conference on Mobile Data Management. Taipei, China, c2009:172-181.

[9] ZHANG C, HUANG Y. Cloaking locations for anonymous location based services: a hybrid approach[J]. GeoInformatica, 2009, 13(2): 159-182.

[10] TRAN M T, ECHIZEN I, DUONG A D. Binomial-mix-based location anonymizer system with global dummy generation to preserve user location privacy in location-based services[C]//2010 International Conference on Availability, Reliability and Security. c2010: 580-585.

[11] PAN X, XU J L, MENG X F. Protecting location privacy against location-dependent attacks in mobile services[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(8): 1506-1519.

[12] PAN X, MENG X F. Preserving location privacy without exact locations in mobile services[J]. Frontiers of Computer Science, 2013, 7(3): 317-340.

[13] KACHORE V A, LAKSHMI J, NANDY S K. Location obfuscation for location data privacy[C]//2015 IEEE World Congress on SERVICES. c2015: 213-220.

[14] SAMARATI P, SWEEENEY L. Protecting privacy when disclosing information: *k*-anonymity and its enforcement through generalization and suppression[J]. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10(5): 571-588.

[15] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]// Proceedings of the International Conference on Mobile Systems, Applications and Services (MobiSys'03). San Francisco, USA, c2003: 163-168.

[16] MOKBEL M F, CHOW C Y, AREF W G. The new casper: query processing for location services without compromising privacy[C]// Proceeding of the 32nd International Conference on Very Large Data Base. New York: ACM Press, c2006: 763-774.

[17] KIDO H, YANAGISAWA Y, SATOH T. An anonymous communica-

tion technique using dummies for location-based services[C]//IEEE International Conference on Pervasive Services (ICPS'05). Santorini, Breece, c2005:88-97.

[18] DUCKHAM M, KULIK L. A formal model of obfuscation and negotiation for location privacy[C]//The International Conference on Pervasive Computing. c2005:152-170.

[19] ARDAGNA C A, CREMONINI M, SAMAEATI P, et al. An obfuscation-based approach for protecting location privacy[J]. IEEE Transactions on Dependable and Secure Comput, 2011, 8(1): 13-27.

[20] GHINITA G, KALNIS P, KHOSHGOZARAN A, et al. Private queries in location based services: Anonymizers are not necessary[C]//The 2008 ACM SIGMOD International Conference on Management of Data. c2008:121-132.

[21] 李敏, 秦志光. 路网环境下位置隐私保护技术研究进展[J]. 计算机应用研究, 2014, 31(9): 2576-2580.
LI M, QIN Z G.. Survey of location privacy protection over road networks[J]. Application Research of Computers, 2014, 31(9): 2576-2580.

[22] CHOW C Y, MOKBEL M F, BAO J, et al. Query-aware location anonymization for road networks[J]. GeoInformatica, 2011, 15(3): 571-607.

[23] PALANISAMY B, LIU L. MobiMix: protecting location privacy with mix-zones over road networks[C]//The 27th IEEE International Conference on Data Engineering. c2011: 494-505.

[24] DAMIANI M L, SILVESTRI C, BERTINO E. Fine-grained cloaking of sensitive positions in location-sharing applications[C]//IEEE Pervasive Computing. 2011, 10(4):64-72.

[25] YIGITOGLU E, DAMIANI M L, ABUL O, et al. Privacy-preserving sharing of sensitive semantic locations under road-network constraints[C]//MDM. c2012:186-195.

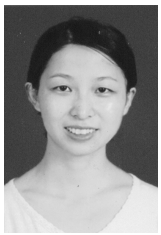
[26] 潘晓, 肖珍, 孟小峰. 位置隐私研究综述[J]. 计算机科学与探索, 2007, 1(3): 268-281.
PAN X, XIAO Z, MENG X F. Survey of location privacy-preserving[J]. Journal of Frontiers of Computer Science and Technology, 2007, 1(3): 268-281.

[27] CHEN S, JENSEN C S, LIN D. A benchmark for evaluating moving object indexes[J]. PVLDB, 2008, 1(2): 1574-1585.

[28] http://www.comp.nus.edu.sg/~spade/benchmark[EB/OL].

[29] BRINKHOFF T. A framework for generating network based moving objects[J]. GeoInfomatica, 2002, 6(2):153-180.

作者简介：



陈慧（1981-），女，安徽芜湖人，南京航空航天大学博士生，南京信息工程大学讲师，主要研究方向为移动对象数据库、隐私保护等。

秦小麟（1953-），男，江苏苏州人，南京航空航天大学教授，主要研究方向为分布式数据管理、物联网、数据安全与隐私保护、大数据管理与分析等。