基于位置服务的隐私保护

孟小峰 潘 晓
中国人民大学

关键词：基于位置服务 位置隐私 查询隐私

隐私保护问题

应用分类

根据服务面向对象不同，基于位置的服务可以分为面向用户和面向设备两种[1]。两种服务的主要区别在于，面向用户的基于位置的服务，用户对服务拥有主控权；面向设备的基于位置的服务，用户或物品属于被动定位，对服务无主控权。根据服务的推送方式的不同，基于位置的服务应用可以分为Push服务和Pull服务。前者是被动接受，后者是主动请求。下面将用四个例子（如表1）说明上述分类。当你进入某城市时收到欢迎信息属于面向用户（你）的Push服务（欢迎信息被主动推送到你的移动设备上）；而你在该城市主动提出寻找最近邻餐馆属于面向用户（你）Pull服务；假如你是某物流公司老板，当你公司负责运输的货物，偏离预计轨道时将向你发出警报信息，这属于面向设备（货物）的Push服务（消息被推送到物流公司老板的移动设备上）；如果你主动请求察看货物运送卡车目前所在位置属于面向设备（货物）的Pull服务。

表1 LBS应用分类

	Push服务	Pull服务
面向用户服务	当你进入某城市时收到欢迎信息	请求查找最近邻餐馆
面向设备服务	在货物追踪应用中，当货物运送偏离预计轨道时给与警报信息	请求查找卡车现在所在位置

基于位置的服务与隐私

很多调查研究显示，消费者非常关注个人隐私保护。欧洲委员会通过的《隐私与电子通信法》中对于电子通信处理个人数据时的隐私保护问题给出了明确的法律规定[2]。在2002年制定的Directive文本中，对位置数据的使用进行了规范，其中条款9明确指出位置数据只有在匿名或用户同意的前提下才能被有效并必要的服务使用。这突显了位置隐私保护的重要性与必要性。此外，在运营商方面，全球最大的移动通信运营商沃达丰（Vodafone）制定了一套隐私管理业务条例，要求所有为沃达 丰客户提供服务的第三方必须遵守。这体现了运营商对隐私保护的重视。

隐私泄露

基于位置服务中的隐私内容涉及两个方面：位置隐私和查询隐私。位置隐私中的位置指用户过去或现在的位置；查询隐私指涉及敏感信息的查询内容，如查询距离我最近的艾滋病医院。任何一种隐私泄露，都有可能导致用户行为模式、兴趣爱好、健康状况和政治倾向等个人隐私信息的泄露。所以，位置隐私保护即防止用户与某一精确位置匹配；类似地，查询隐私保护要防止用户与某一敏感查询匹配。

位置服务 VS 隐私保护

回想一下，我们似乎正面临一个两难的抉择。一方面，定位技术的发展让我们可以随时随地获得基于位置的服务；而另一方面，位置服务又将泄露我们的隐私……当然，你可以放弃隐私，获得精确的位置，享受完美的服务；或者，你可以关掉定

位设备，为了保护隐私而放弃任何位置服务。是否存在折中的方法，即在保护隐私的前提下享受服务呢？可以，位置隐私保护研究所做的工作就是要在隐私保护与享受服务之间寻找一个平衡点，让鱼与熊掌兼得成为可能。

隐私保护方法

下面将介绍在基于位置服务中的三种基本的隐私保护方法。

假位置

第一种方法是通过制造假位置[3]达到以假乱真的效果。如在图1中，用户寻找最近的餐馆。白色方块是餐馆位置，红色点是用户的真实位置。当该用户提出查询时，为其生成两个假位置，即哑元（如图1中的黑色点）。真假位置一同发送给服务提供商。从攻击者的角度，同时看到三个位置，无法区分哪个是真实的哪个是虚假的。

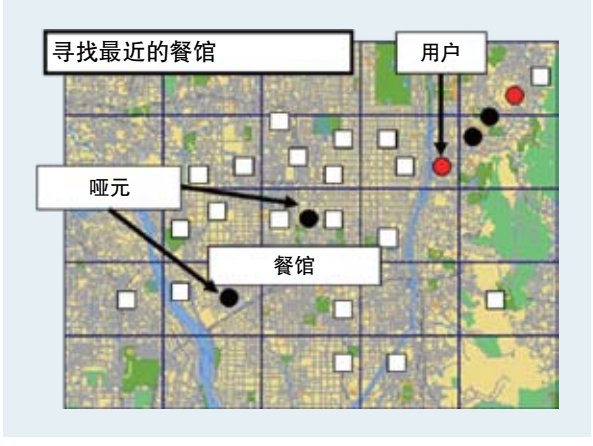

图1 假位置

时空匿名

第二种方法是时空匿名[4~6]，即将一个用户的位置通过在时间和空间轴上扩展，变成一个时空区域，达到匿名的效果。以空间匿名为例，延续图1寻找餐馆的例子，当用户提出查询时，用一个空间区域表示用户位置，如图2中的红色框。从服务提供商角度只能看到这个区域，无法确定用户是在整个区域内的哪个具体位置上。

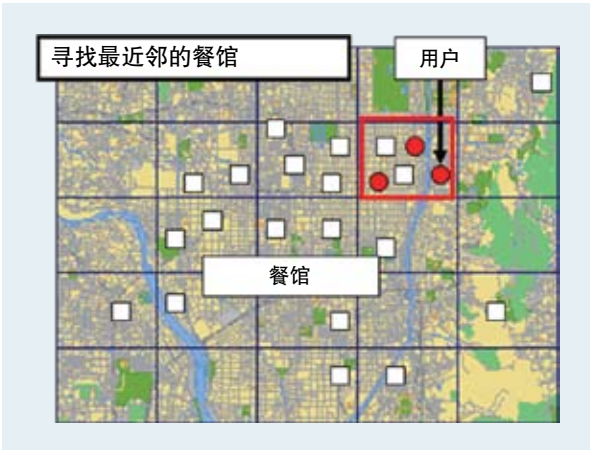

图2 时空匿名

空间加密

第三种方法是空间加密[7]，即通过对位置加密达到匿名的效果。继续前面的例子，首先将整个空间旋转一个角度（如图3），在旋转后的空间中建立希尔伯特（Hilbert）曲线。每一个被查询点P（即图3中的白色方块）对应的希尔伯特值如该点所在的方格数字所示。当某用户提出查询Q时，计算出加密空间中Q的希尔伯特值。在此例子中，该值等于2。寻找与2最近的希尔伯特值所对应的P，

图3 加密

即P1。将P1返回给用户。由于服务提供商缺少密钥，在此例子中即旋转的角度和希尔伯特曲线的参数，故无法反算出每一个希尔伯特值的原值，从而达到了加密的效果。

感知隐私的查询处理

在基于位置的服务中，隐私保护的最终目的仍是为了查询处理，所以需要设计感知隐私保护的查询处理技术。

根据采用匿名技术的不同，查询处理方式也不同：如果采用的是假数据，则可采用移动对象数据库中的传统查询处理技术，因为发送给位置数据库服务器的是精确的位置点。如果采用时空匿名，由于查询处理数据变成了一个区域，所以需要设计新的查询处理算法。这里的查询处理结果是一个包含真实结果的超集。如果采用空间加密技术，查询处理算法与使用的加密协议有关。

隐私度与效率对比

从匿名效率和隐私度两方面对上述三种隐私保护方法进行对比[8]（如图4），可以看出加密是安全度最高的方法，但是加密解密效率较低；生成假数据的方法最简单、高效但隐私保护度较低，可根据用户长期的运动轨迹判断出哪些是假数据；从已有的工作来看，时空匿名在隐私度与效率之间取得了较好的平衡，也是普遍使用的匿名方法。下面将以时空匿名方法为主进行介绍。

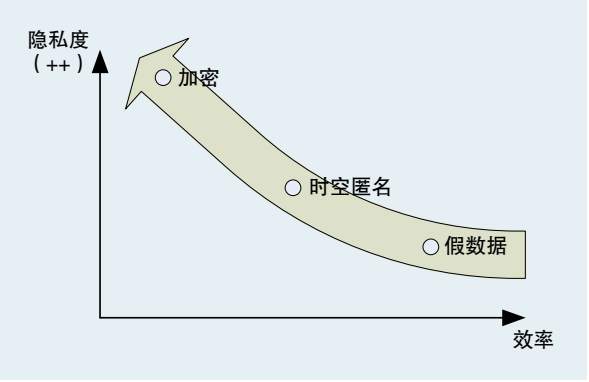

图4 隐私度与效率对比

存在的挑战

位置隐私研究中所面临的挑战包括四个方面：

1. 隐私保护与位置服务是一对矛盾；

2. 基于位置服务的请求，具有在线处理的特点，故位置匿名具有实时性要求；

3. 基于位置服务中的对象，位置频繁更新；

4. 不同用户的隐私要求大相径庭，所以隐私保护需要满足个性化的需求。

隐私保护系统结构

隐私保护系统基本实体包括移动用户和位置服务提供商，它具有如下有四种结构：独立结构[15]、中心服务器结构[4-6]、主从分布式结构[11]和移动点对点结构[9]。

独立式结构

独立结构是仅有客户端（或者移动用户）与位置服务器的客户端/服务器（Client/Server，C/S）结构。由移动用户自己完成匿名处理和查询处理的工作。该结构简单，易配置，但是增加了客户端负担，并且缺乏全局信息，隐私的隐秘性弱。

中心服务器结构

与独立结构相比，中心服务器结构在移动用户和服务提供商之间加入了第三方可信匿名服务器，由它完成匿名处理和查询处理工作。该结构具有全局信息，所以隐私保护效果较上一种好。但是由于所有信息都汇聚在匿名服务器，故可能成为系统处理瓶颈，且容易遭到攻击。

主从分布式结构

为了克服中心服务器的缺点，研究人员提出了类似主从分布式结构。移动用户通过一个固定的通信基础设施（如基站）进行通信。基站也是可信的第三方，与前者的区别在于它只负责可信用户的认证以及将所有认证用户的位置索引发给提出匿名需

求用户。位置匿名和查询处理由用户或者匿名组推举的头节点完成。该结构的缺点是网络通信代价高。

移动点对点结构

移动点对点结构与分布式结构工作流程类似，惟一不同的是它没有固定的负责用户认证的通信设施，而是利用多跳路由寻找满足隐私需求的匿名用户。所以它拥有与分布式结构相同的优缺点。

隐私保护研究内容

下面将介绍一些经典位置隐私和查询隐私保护方法以及感知隐私的查询处理技术。

隐私保护模型

先介绍一下迄今为止使用最广泛的位置k-匿名模型[10]，后面介绍的隐私保护方法均满足该模型。k-匿名是隐私保护中普遍采用的方法。位置k-匿名的基本思想是让一个用户的位置与其他（k-1）用户的位置无法区别。以位置3-匿名为例（如图5），将三个单点用户用同一个匿名区域表示，攻击者只知道在此区域中有3个用户，具体哪个用户在哪个位置，无法确定，达到了位置隐私保护目的。

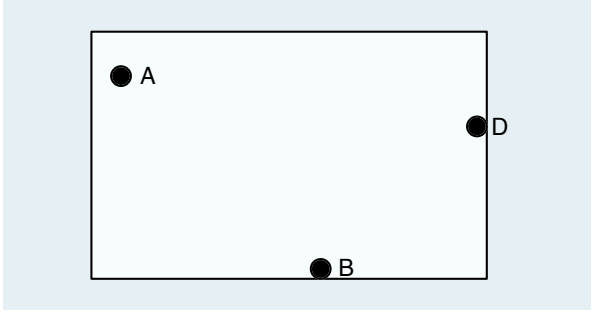

图5 位置3匿名

基于四分树的隐私保护方法

最早的匿名算法是基于四分树的隐私保护方法[10]。它解决了面对大量移动用户高效寻找满足位置k-匿名模型的匿名集的问题。其解决方法是：自顶向下地划分整个空间，如果提出查询的用户所在的区域的用户数大于 k，将整个空间等分为4份，重复这一步，直至用户所在的区域所包含的用户数小于 k，返回四分树的上一层区域，将其作为匿名区域返回。该方法的缺点是要求用户采用统一隐私度和返回匿名区域过大。

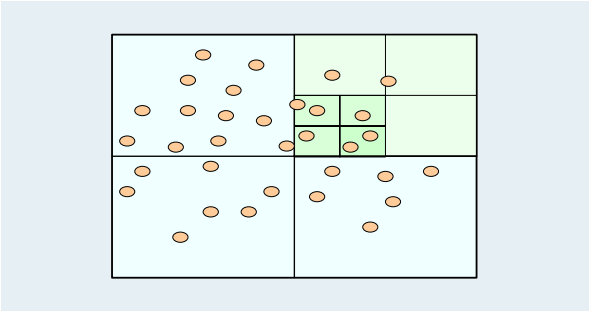

图6 基于四分树的匿名方法

个性化隐私需求匿名方法

在隐私保护中，不同的用户有不同的位置k-匿名需求，因此需要解决满足用户个性化隐私需求的位置k-匿名方法，这正是CliqueCloak[12]的贡献。其解决方法是利用图模型形式化定义此问题，并把寻找匿名集的问题转化为在图中寻找k-点团的问题。图7中，点是用户提出查询时的位置，k表示用户的最小隐私需求，圆圈代表用户可接受的最差服务质量。当新的对象m到达时，根据用户的隐私和质量要求，更新已有图，并找出m所在团。将覆盖该团所有点的最小边界矩形作为匿名区域返回。

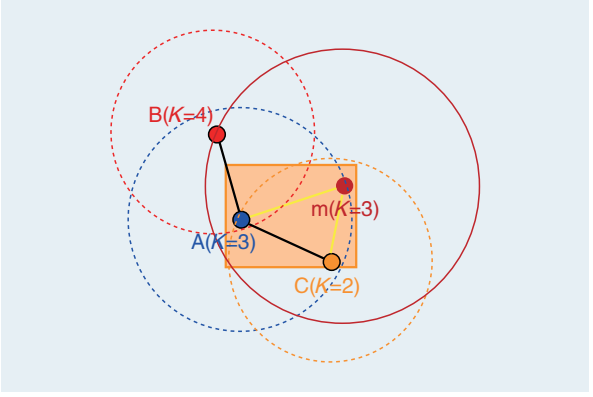

图7 个性化隐私需求匿名方法

连续查询隐私保护

前面的隐私保护工作都是针对Snapshot查询类型。如果将现有的匿名算法直接应用于连续查询隐私保护将产生查询隐私泄露。如图8所示，系统中存在6个用户{A,B,C,D,E,F}。攻击者知道存在连续查询，但并不知道连续查询是什么，以及是由谁提出的。在3个不同时刻t_i、t_{i+1}、t_{i+2}，用户A形成了3个不同的匿名集，即{A, B, D}、{A, B, F}、{A, C, E}。将三个匿名集取交，即可获知是用户A提出的查询Q1。

此问题主要是由同一用户（A）在其有效生命期内形成的匿名集不同而造成的。所以解决方法[16]是让连续查询的用户在最初时刻形成的匿名集在其查询有效期内均有效。同一个例子中，即用户A在t_i时刻形成的匿名集是{A,B,D}，则在t_{i+1},t_{i+2}时刻，匿名集依然是{A,B,D}，如图8（b）和（c）中虚线矩形所示。

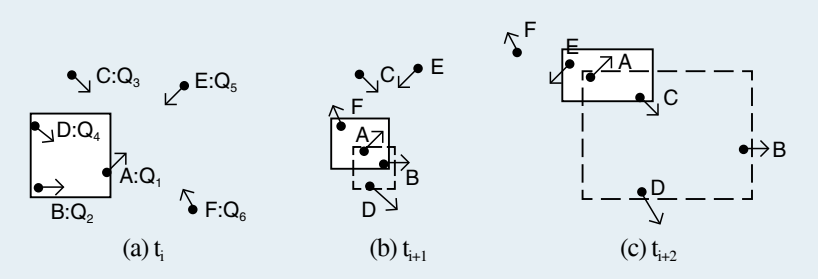

图8 连续查询隐私保护技术

感知查询差异性的隐私保护

位置k-匿名模型只能防止用户与查询建立关联，但不能切断用户与查询内容之间的关联。图9显示的是在位置匿名后发布的匿名位置和查询，符合位置3匿名。但是，攻击者可以确定，位置落于[（1,2）~（5,9）]的第一个匿名集中的用户，一定患了某种疾病。

解决此问题的基本方法[20]是在寻找匿名集的时候考虑查询语义，保证在一个匿名集中敏感查询所占比例不超过p%。如在图10中所示，即使攻击者拥有用户的真实位置，获知该用户落于哪个匿名集中，但他仍然无法获知该用户提出了何种查询。

位置	查询
[（1,2）~（5,9）]	医院
[（1,2）~（5,9）]	诊所
[（1,2）~（5,9）]	医院
[（2,5）~（4,7）]	加油站
[（2,5）~（4,7）]	加油站
[（2,5）~（4,7）]	学校

图9 查询隐私泄露

位置	查询
[（1,2）~（4,7）]	**俱乐部A
[（1,2）~（4,7）]	加油站
[（1,2）~（4,7）]	加油站
[（5,2）~（7,9）]	餐馆
[（5,2）~（7,9）]	诊所
[（5,2）~（7,9）]	学校

图10 感知查询差异性的隐私保护技术

感知隐私保护的查询处理

如何基于匿名后的位置（一个区域）为用户求得查询结果是隐私保护中必须考虑的另一重要问题。

在基于区域位置数据的查询处理技术中，位置数据可以分为两种：公开位置数据和隐私位置数据[16]。公开数据是指如加油站、旅馆和警车等公共信息，其位置是一个精确点；隐私数据属于个人数据，其位置是一个模糊的范围。根据查询点和被查询点是否是隐私数据，可以将查询分为四种（如表2）：基于公开数据的公开查询、基于隐私数据的公开查询、基于公开数据的隐私查询和基于隐私数据的隐私查询。

基于公开数据的公开查询可以用传统方法处理，基于隐私数据的公开查询和基于公开数据的

表2 感知隐私的查询类型

		被查询点	
		公开数据	隐私数据
查询点	公开数据	基于公开数据的公开查询 如：在某电影院200m内所有餐馆	基于隐私数据的公开查询 如：某加油站500米内所有出租车
	隐私数据	基于公开数据的隐私查询 如：距离我最近的加油站	基于隐私数据的隐私查询 如：离我最近的朋友

隐私查询是基于隐私数据的隐私查询的特例。所以这两种查询处理方法经过扩展可以适应第四种情况。

基于隐私数据的公开查询

首先以范围查询为例说明基于隐私数据的公开查询。如查询“某加油站500米内所有出租车”，出租车是空间匿名后得到的区域位置信息，圆是加油站附近500米的范围。最简单的方法[19]是将所有与查询范围相交的匿名框作为候选结果集，匿名框与查询范围重叠区域面积表示查询结果是真正结果的概率，在图11中查询结果即{（B, 50%），（C, 90%），（D, 100%），（E, 60%））}。

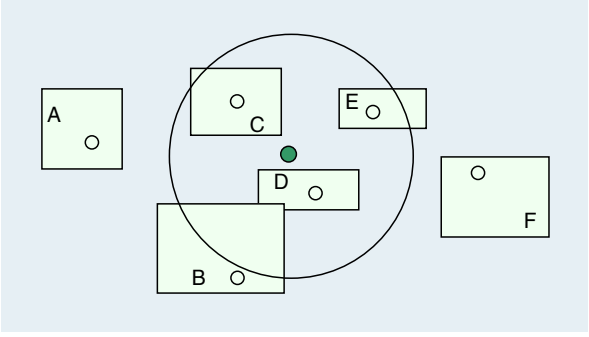

图11 基于隐私数据的公开查询

基于公开数据的隐私查询

查询距离用户现在所在位置最近的加油站。加油站如图12中的p1~p8点，用户的位置是一个匿名区域。为使候选结果集中包含真实结果，需要计算该匿名区域内每一个点的最近邻。这个结果集包含两部分[13]：所有被匿名区域覆盖的点和匿名框边上的每一个点所对应的最近邻。后者可以通过寻找被查询点间的垂直平分线与该边的交点获得。

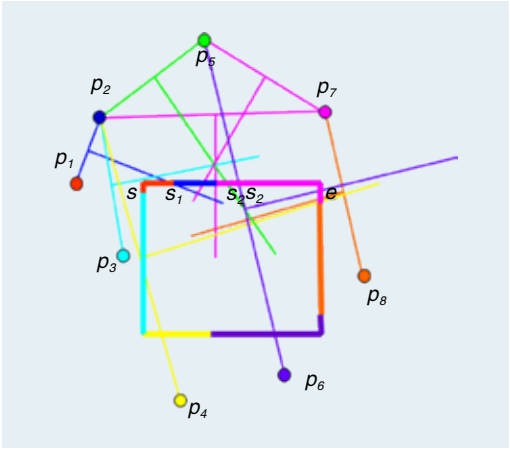

图12 基于公开数据的隐私查询

隐私保护技术面临的挑战

除了上述问题外，在基于位置的服务中隐私保护问题仍面临着很多挑战，如多技术混合的隐私保护技术、轨迹的隐私保护技术和室内位置隐私保护技术等。

多技术混合的隐私保护

如前所述，加密方法安全但不高效，时空匿名高效但相对加密方法而言不够安全。虽然目前大部分研究工作均集中在时空匿名方法上，但是我们试图在加密与时空匿名之间做些工作，研究结合加密算法的高隐秘性和空间匿名算法高效性的混合匿名模型与算法，同时保证利用此种匿名方法所获得数据的可用性，并研究基于混合匿名技术的查询处理算法。

移动轨迹的隐私保护

由于攻击者可能积累用户的历史信息分析用户的隐私，因此还要考虑对用户的连续位置保护的问题，或者说对用户的轨迹提供保护。现有大部分

的轨迹匿名技术[14]多采用发布假数据或丢掉一些取样点的方法。按照前面的分析，这样的方法不够安全，可能通过挖掘历史信息辨别真伪。因此需要研究基于时空匿名的轨迹匿名模型和算法，在保证挖掘结果正确性的前提下保证用户轨迹信息不泄漏。另外，现有的轨迹匿名多是离线（Offline）处理方式。在基于位置的服务中存在汽车导航的应用，用户需查询从A地到B地的行车路线。研究在线轨迹匿名模型和算法是另一个值得关注的问题。

室内位置隐私

研究工作大都专注于室外位置隐私保护，其实在室内也存在隐私泄露的问题。在室内安装无限传感器收集用户位置，可用于安全控制和资源管理，如当室内人数小于某个值时关掉空调设备。但是收集室内人员位置信息的同时可能会泄露个人隐私[12]。如在公司中，管理者可以监控雇员行为，并推测健康状况等。为了保护室内人员的个人隐私，需要针对室内环境特点，研究基于室内位置隐私的攻击模型、匿名模型、匿名算法和查询处理算法。

要解决以上问题，可以将现有技术如数据发布中的隐私保护技术、移动数据的查询处理技术和不确定数据的建模、查询处理技术相结合，这也许会给我们带来一些意想不到的惊喜。■

孟小峰

CCF理事、2009年CCF王选奖一等奖获得者。中国人民大学教授。主要研究方向为网络与移动数据管理等。

xfmeng@ruc.edu.cn

潘　晓

中国人民大学博士生。主要研究方向为移动数据管理，隐私保护。

smallpx@ruc.edu.cn

参考文献

[1] ABI Research, http://www.abiresearch.com/press/1483-Global+LBS+Revenues+to+Reach+ $2.6+Billion+ in+2009 4

[2] J. Schiller, Jochen, A. Voisard. Location-based Services, Elsevier Science Ltd, April 200

[3] H. Kido, Y. Yanagisawa, and T. Satoh. Protection of location privacy using dummies for location-based services, In Proc.the 25th International Conference on Distributed Computing Systems(ICPS’05), 2005

[4] M. Gruteser. and D. Grunwald. Anonymous Usage of Location-based Services Through Spatial and Temporal Cloaking, In Proc. of the International Conference on Mobile Systems, Applications, and Services (MobiSys’03), 2003，pp.163~168

[5] B. Gedik and L. Liu. Location Privacy in Mobile Systems: A Personalized Anonymization Model, In Proc. of the International Conference on Distributed Computing Systems (ICDCS’05), 2005

[6] M. F. Mokbel, C. Y. Chow, and W. G. Aref, The New Casper: Query Processing for Location Services without Compromising Privacy, In Proc. of the 32nd International Conference on Very Large Data Bases (VLDB’06), 2006

[7] A. Khoshgozaran and C. Shahabi. Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy. In Proc. of SSTD, 2007

[8] G. Ghinita. Understanding the Privacy-Efficiency Trade-off in Location-Based Queries, ACM SIGSPATIAL GIS Workshop on Security and Privacy in GIS and LBS (SPRINGL), November 2008

[9] G. Ghinita, P. Kalnis and S. Skiadopoulos. MobiHide: A Mobile Peer-to-Peer System for Anonymous Location-Based Queries, In Proceedings of International Symposium on Spatial and Temporal Databases (SSTD), July 2007

[10] C. Chow and M. Mokbel. Privacy in Location-based Services: A System Architecture Perspective. The SIGSPATIAL Special Newsletters, SIGSPATIAL Special, July 2009，Vol. 1, No. 2, pages 23~27

[11] G. Ghinita, P. Kalnis, and S. Skiadopoulos. PRIVE: Anonymous Location based Queries in Distributed Mobile Systems. In Proceedings of International Conference on World Wide Web, WWW, 2007，pages 1~10

更多参考文献请访问：www.ccf.org.cn的“中国计算机学会通讯”栏目

[12] C. Chow, M. F. Mokbel, and T. He. Tinycasper: a privacy-preserving aggregate location monitoring system in wireless sensor networks. In proceedings of SIGMOD08(demo), Vancouver, Canada ,2008，Pages 1307~1310

[13] H. Hu and D. Lee. Range nearest-neighbor query. IEEE Transactions on Knowledge and Data Engineering (TKDE), 2006，18(1):78~91

[14] O. Abul, F. Bonchi, and M. Nanni. Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases, In Proc. of International Conference on Data Engineering (ICDE’08), 2008，pp.376~385

[15] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar. Preserving User Location Privacy in Mobile Data Management Infrastructures, In Proc. of Privacy Enhancing Technology Workshop (PET’06), 2006

[16] C. Chow and M. F. Mokbel, Enabling Privacy Continuous Queries for Revealed User Locations, In Proc. of the International Symposium on Advances in Spatial and Temporal Databases (SSTD’07), 2007

[17] X. Pan, X. Meng, J. Xu. Distortion-based Anonymity for Continuous Query in Location-Based Mobile Services. In the proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2009), Seattle, Washington，2009，November 4~6

[18] X. Pan, J. Xu, X. Meng. Protecting Location Privacy against Location-Dependent Attack in Mobile Services. In Proceedings of the ACM 17th Conference on Information and Knowledge Management(CIKM2008), page 1475~1476, Napa Valley, California, 2008，October 26~30

[19] O. Wolfson, P.A. Sistla, S. Chamberlain, and Y. Yesha. Updating and Querying Databases that Track Mobile Units, Distributed and Parallel Databases, vol. 7, no. 3,1999，pp. 257~387

[20] Z. Xiao, J. Xu, and X. Meng. p-sensitivity: a semantic privacy protection model for location-based services, Proc. of the 2nd International Workshop on Privacy-Aware Location-based Mobile Services(PALMS’08), Beijing, China, 2008