轨迹隐私保护技术研究进展分析

胡兆玮[1,2] 杨 静[2]

(吉林师范大学计算机学院 四平 136000)[1] (哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001)[2]

摘 要 随着移动社会网络和信息技术的发展,基于位置的服务应用更加广泛,轨迹隐私作为一种特殊的位置隐私,受到了人们的广泛关注。轨迹数据蕴含了移动用户在时间、空间维度内的丰富信息,结合其他相关的背景知识对这些数据进行挖掘、分析,可以获得许多隐私信息,其可能对人身安全造成威胁。因此如何在保证用户获得高质量服务的同时,又较好地保护用户的轨迹隐私,是轨迹隐私保护技术研究的核心内容。首先介绍了轨迹隐私保护的概念、应用类型、技术分类、衡量标准和系统结构;其次研究了近年来国内外关于轨迹隐私保护研究的主要技术和方法;最后分析了该领域当前的研究热点,并对未来的研究方向进行了展望。

关键词 隐私保护,位置服务,轨迹隐私,匿名化

中图法分类号 TP393 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.4.003

Survey of Trajectory Privacy Preserving Techniques

HU Zhao-wei[1,2] YANG Jing[2]

(Computer College, Jilin Normal University, Siping 136000, China)[1]

(College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)[2]

Abstract As the development of mobile social networks and information technology, location-based service applications are more widely used. Trajectory privacy is a special location privacy, which receives more and more attention. The collected trajectory data contain extensive information about mobile user under the dimensions of time and space. Attackers may get much private information by mining and analyzing background information on mobile user, which could result in human safety threat. Therefore, it is an important content of trajectory privacy preserving technology that how to ensure users get high-quality services as well as protect the users' trajectory privacy. Firstly, this paper discussed the concept, type, classification, metric and architecture in terms of trajectory privacy preserving. Secondly, it described the recent major research on techniques and methods of trajectory privacy preserving. Finally, it generalized the current research focus in the field, and concluded the future research.

Keywords Privacy preserving, Location-based service, Trajectory privacy, Anonymity

1 引言

随着计算机技术和定位技术的发展,移动用户随时随地享受信息服务成为可能,通过向服务器提供自己的实时位置以及查询请求内容,位置服务商将相应的基于位置的查询信息进行处理并将处理结果反馈给用户。位置服务质量的高低与移动用户位置数据的准确度有很大的关系,位置信息越精确,用户所获得的服务体验也越好[1]。然而如果位置服务器是不可信的,则会对移动用户的位置及轨迹隐私安全造成很大的威胁,因此在提供精确的位置服务的同时,如何保证移动用户位置及轨迹隐私的安全,是一个亟待解决的问题[2]。

轨迹隐私保护是在位置服务基础上产生和发展起来的,研究发现,保护了移动用户的位置隐私,并不能完全保护移动用户的轨迹隐私。在基于位置的服务中,用户将包含自身位置信息的查询内容发送给位置服务器来处理,如果信息被攻击者截获或位置服务器被攻击者监听,则攻击者会获得用户的位置信息[3]。如果移动用户连续提出位置查询请求,攻击者不断获取用户的位置,将这些位置连接起来,就很容易获得用户的运动轨迹。通过对这些轨迹的分析,结合其他相关的背景知识,攻击者可以推断出用户的许多隐私信息,如工作单位、家庭住址、行为习惯、经常去什么场所等,从而造成个人隐私信息的泄露,甚至对人身安全造成威胁。因此,如何保护移动用户的轨迹隐私,已成为移动用户隐私保护研究的重要课题。

本文第2节介绍轨迹隐私保护的概念、应用类型、技术分类、衡量标准和轨迹隐私保护模型的2种系统结构;第3节到第5节介绍了目前轨迹隐私保护的相关技术和方法;第6节对轨迹隐私保护未来的研究方向进行了展望;最后总结全文。

到稿日期:2015-03-11 返修日期:2015-07-24 本文受国家自然科学基金项目(61370083,61073043,61073041,61402126),四平市科技发展计划资助项目(2014056)资助。

胡兆玮(1980—),男,博士生,讲师,主要研究方向为轨迹隐私保护、数据挖掘等,E-mail:huzw19@126.com;杨 静(1962—),女,教授,博士生导师,主要研究方向为数据库理论与应用、隐私保护、数据挖掘等。

2 轨迹隐私保护中的主要问题

2.1 基本概念

隐私是个人、组织、机构等不愿意被外界知道的信息，如个人的兴趣、爱好、疾病、收入、生活习惯，机构的客户信息、经营状况等。由于意识观念、工作性质的不同，人们对隐私的认定标准也不相同，但任何不愿意被窃取或披露的私密信息，都可以称为隐私。

轨迹是指移动用户的位置信息随着时间变化形成的序列，可形式化表示为 $T=\{id,(x_1,y_1,t_1),(x_2,y_2,t_2),\cdots,(x_n,y_n,t_n)\}$，其中 id 是移动用户的标识或者数据项的标识符，(x_i,y_i,t_i) 表示移动用户在 t_i 时刻所处的位置为 (x_i,y_i)。将移动用户的位置信息关联起来，便可得到移动用户的轨迹信息。

轨迹隐私是一种特殊的隐私，是根据连续的位置信息形成的轨迹里所透露出的敏感信息，或通过其他人的轨迹信息推断出的敏感信息，如“在哪里”、“去过哪里”、“正在做什么”、“做过什么”和“即将去哪里”等。轨迹隐私保护既要保证不泄露轨迹自身的隐私信息，还须防止通过部分信息或轨迹获取其他个人的隐私信息。

2.2 轨迹隐私保护的类型

轨迹隐私保护是以社会关系层和位置层为基础，利用用户的移动关系和情境关系对位置轨迹层中位置关联形成的轨迹信息，包含了详尽的时空信息，既有动态的信息，也有静态的信息，对其研究可支持许多实际应用。目前轨迹隐私保护主要解决连续位置查询中的轨迹隐私保护、数据发布中的轨迹隐私保护两方面的隐私问题[4]。

连续位置查询中形成的数据属于动态数据，对其的保护需要综合考虑轨迹、位置、环境、情境等相关信息；数据发布中的轨迹数据属于静态数据，在保护其隐私的同时，还需要考虑所公布数据的使用率。

2.2.1 连续位置查询中的轨迹隐私保护

保护了移动用户的位置隐私，并不能完全保护移动用户的实时运行轨迹，攻击者可以通过其他手段获得移动用户的运行轨迹。在 LBS 服务中，用户为了获得高质量的位置服务，必须提供精确的位置信息；如果想获得连续的位置服务，必须持续提供位置信息，即连续的位置查询[5]。现有的位置隐私保护算法仅保护单一位置，且容易泄露连续位置的轨迹隐私信息，如使用 K 匿名方法可以实现用户的位置轨迹隐私保护，但当用户连续发出位置服务请求时，根据 K 匿名算法的基本思想，会在用户的周围形成一系列匿名框，将这些匿名框关联起来，就可以获得用户活动的大致轨迹，再结合其他的背景知识或相关技术，便有很大概率获得用户的轨迹信息[6]。同时，现有的轨迹匿名算法没有考虑轨迹的情境信息，又存在着由于匿名区域过大而导致的查询服务不准确和等待时间过长等不足[7,8]，因此保护连续位置查询中的轨迹隐私对移动用户具有重要的现实意义。

2.2.2 轨迹数据发布的隐私保护

轨迹数据本身包含了详尽的时空信息，对轨迹数据的分析和研究可获得许多有用的信息。在轨迹数据发布中，最简单的隐私保护方法是隐藏、删除每条轨迹数据的准标识属性，或者将用户的真实信息用虚假信息代替，但这些方法有时候并不有效，攻击者利用已获得的匿名时空数据也能推测出移动用户的隐私信息，同时直接删除某些属性也会影响对数据的使用效果[9,10]。如某医院发布的数据中，每个轨迹数据集中并未包含病人的姓名等身份信息，而是随机赋予一个统一的、随机的 RID，但它包含了一个敏感属性：疾病。如 $RID=1$ 的病人在时间戳 2、4、5、8 分别到过位置(1,5)、(6,7)、(8,10)、(11,8)。如果一个攻击者知道一个叫 Alice 的用户在时间戳 2 和 8 到过位置(1,5)和(8,10)，在发布的信息库中只有 $RID=1$ 的病人满足该临时属性，则攻击者可以断定 Alice 患了 $RID=1$ 所对应的疾病。因此，保护数据发布中的轨迹数据信息也是轨迹隐私保护研究的重要内容。

2.3 轨迹隐私保护技术分类与衡量标准

2.3.1 轨迹隐私保护技术分类

轨迹隐私保护是伴随着位置隐私保护技术的发展而产生和发展的[11,12]，主要解决轨迹数据发布和连续位置查询中的轨迹隐私保护问题。其基本思想是从保护轨迹上的敏感位置点、保护用户轨迹数据的孤立性和保护用户移动属性的安全性等方面，来实现隐私保护的目标[13]。目前的轨迹隐私保护技术大致可以分为以下几类。

(1)泛化法，是指对用户的位置、轨迹进行模糊化处理，将点位置扩展为面位置区域，从而降低攻击者识别用户的概率[14]。泛化法能够保证发布数据的真实性，但由于加入了大量的辅助数据，系统的计算开销以及响应时间增加了。

(2)抑制法，是指对原始轨迹数据有限制性地进行发布，对一般数据直接发布，对敏感数据限制发布或经过转换处理后再发布。该方法最大的缺点是如果抑制的点太多，会影响发布数据的可用性，降低服务质量[15]。

(3)假数据法，指通过用虚假的轨迹代替真实的轨迹或在真实数据中添加虚假的数据，以达到混淆真实数据的目的[16]。虚假的轨迹数据一般通过旋转生成法或随机生产法来产生，再加入真实数据中，这种方法通常情况会造成服务质量下降，查询的准确度会受到一定的影响。

2.3.2 轨迹隐私保护衡量标准

轨迹隐私保护的目的是在保护用户隐私的同时，保证发布的数据具有较高的可用性和移动用户能够获得良好的服务体验[17]。为了评估轨迹隐私保护的水平和隐私保护技术在实际应用中所达到的效果，主要从以下几个方面进行轨迹隐私保护方法的衡量。

(1)隐私保护度。一般来说，隐私保护度越高，隐私泄露的风险就越小，隐私泄露的概率也越小。但在实际应用中，隐私的保护程度不仅与隐私保护算法的优劣以及攻击者所掌握的背景知识有关，还与移动用户所处的空间区域、地理环境、用户数量等因素有关。

(2)服务质量。在轨迹隐私保护中，轨迹数据发布的信息扭曲度越小，服务质量越好；在移动用户的位置服务中，移动用户所获得的服务体验越好、响应时间越短，则服务质量越好。

(3)信息熵。假设攻击者结合自己掌握的知识能识别出

移动用户的概率为 $P_i(i=1,2,\cdots,k)$，则隐私保护强度能用信息熵来权衡，$H=-p_1\log_{10}p_1-p_2\log_{10}p_2-\cdots-p_k\log_{10}p_k$。信息熵越大，隐私保护强度越好，攻击者能识别出用户的概率就越低。

(4)匿名成功率。指成功匿名的轨迹数在所有发送的轨迹中所占的比例，能够体现轨迹隐私保护算法的性能。匿名成功率越高，算法越好。

(5)匿名时间。指匿名算法将请求用户的轨迹进行匿名处理所需的时间，能够体现出隐私保护算法的执行效率。匿名时间越短，算法的效率越高。

(6)查询时间。指匿名服务器进行匿名查询时所花费的时间，主要用来衡量匿名查询服务器端的查询代价，查询花费的时间越短，系统执行效率就越高，算法的性能就越好。

(7)候选结果大小。指匿名处理后返回的查询结果集的大小，用来衡量数据服务器和匿名服务器之间通信代价的大小，候选结果越小，通信代价也越小，算法性能越好。

2.4 轨迹隐私保护的系统结构

轨迹隐私保护是位置隐私保护的一个方面，其系统结构是构建在位置隐私保护系统结构的基础上的。轨迹隐私保护模型根据移动用户提出位置服务请求时，终端移动节点请求位置服务的方式不同(尤其是在连续位置查询中的轨迹隐私保护，用户是通过自身实现轨迹匿名，还是通过第三方的匿名服务器进行匿名处理)，可分为分布式结构和中心服务器结构。

2.4.1 分布式结构

分布式结构由客户端和数据库服务器两部分组成，客户端自身完成轨迹匿名和查询结果的求精工作。每个提出查询请求的移动用户根据匿名算法查询适合的其他用户构建的匿名区域，然后再选择适合的用户节点将匿名集合转发，最后将查询结果反馈给请求用户[19]。分布式网络具有一定的自治性，没有固定的第三方，比较适合移动用户轨迹隐私保护机制的实现；同时，数据库和匿名服务相互独立，增强了容错能力和抗攻击能力。但其结构复杂，网络负载不易均衡。

2.4.2 中心服务器结构

中心服务器结构由客户端、数据库服务器和匿名服务器组成，匿名服务器负责收集用户精确的位置轨迹信息，对轨迹信息进行匿名，并对查询结果进行求精等[18]。中心服务器结构在一定程度上减轻了客户端的负担，可以满足移动用户提出的服务请求。但由于要完成移动用户轨迹信息的匿名和查询结果的求精等大量的工作，系统的开销会增大，从而成为系统处理的一个瓶颈。

在轨迹隐私保护中，中心服务器由于结构简单，易于实现和维护，能对全局数据进行整体匿名保护，因此在实际应用中较为普遍。

3 泛化法

轨迹是移动用户的位置信息按时间形成的序列，当前关于轨迹隐私保护的研究主要面向一般应用的轨迹隐私保护技术，主要分为泛化法、抑制法和假数据法。

泛化法是将点、线位置扩展为面、区域位置的方法，轨迹k-匿名技术是轨迹隐私保护中最基本的泛化技术，由Sweeney最早提出[32]，主要用于保护关系数据库中的数据隐私，其本质是将数据库记录中的标识符属性进行泛化处理，使特定的记录和其它 $k-1$ 条记录无法区分。随后 Marco Gruteser 把关系数据库中的 k-匿名技术用于位置隐私保护中，提出了 k-匿名位置隐私保护模型[33]，即移动用户当前所处的位置和其它 $k-1$ 个用户的位置无法区分时，认为用户的位置满足k-匿名位置隐私保护模型。其中 k 表示隐私保护强度，k 越大，匿名效果越好。随着 LBS 技术的深入发展，人们发现保护了用户的位置隐私，并不能完全保护用户的轨迹隐私[20]，于是 k-匿名技术被引入到轨迹隐私保护中，形成了k-匿名轨迹隐私保护模型，其核心思想是将一条轨迹和其它 $k-1$ 条相似的轨迹泛化为一个匿名区域，以达到保护隐私的目的。

3.1 k^m-匿名轨迹数据隐私保护方法

用户的身份信息会通过记录其运动的数据泄露。有时将身份信息删除也可能导致轨迹隐私信息的泄露。在 k-匿名方法的基础上，Giorgos 提出了一种改进的 k^m-匿名方法来保护移动用户的轨迹隐私信息[21]。k^m-匿名方法是一种泛化的隐私保护模型，用于降低交易数据发布中用户身份信息泄露的可能，其中 k、m 为隐私保护参数，k 表示隐私保护强度，m 表示攻击者已经掌握用户之前访问过的 m 个位置点的相关信息。该方法利用概括法来最小化原始轨迹与匿名轨迹之间的距离，通过利用基于距离的归纳方法来实现轨迹数据的 k^m-匿名。该模型的最大优点是在轨迹数据发布之前不需要了解准标识符属性的详细信息，也不需要区分敏感信息和非敏感信息。

定义 1(支持集 $Sup(s,T)$) T 是一个轨迹集合，s 是 T 的一个子集，$sup(s,T)$表示 T 中包含集合 s 的轨迹数[21]。

当轨迹集合 T 中的每个轨迹 t 至少包含 k 个轨迹，且每一个支持子集 S 至多包含 m 个位置点时，该轨迹集合 T 实现了 k^m-匿名隐私保护。该匿名方法确保一个攻击者掌握一个移动用户的任何支持子集 S 的势不大于 m，也确保不能以高于 $1/k$ 的概率识别出该用户[34]。

例如在轨迹集合 $T=\{t_1,t_2,t_3,t_4,t_5,t_6\}$ 中，$t_1=(d,a,c,e)$，$t_2=(b,a,e,c)$，$t_3=(a,d,e)$，$t_4=(b,d,e,c)$，$t_5=(d,c)$，$t_6=(d,e)$。轨迹集合 T 满足 2^1-匿名和 1^3-匿名，而不满足 2^2-匿名，因为子集(d,a)仅包含于 t_1 中，未达到隐私保护度为 $k=2$ 的条件。

k^m-匿名方法中引入了 seqanon 算法，通过最小化原始轨迹和匿名轨迹之间的欧氏距离，利用归纳化的方法实现轨迹信息的匿名化。在上例轨迹集合 T 中，设 $k=2,m=2$，即实现轨迹集合 T 的 2^2-匿名化。首先计算出 T 的所有最小支持集 S，如表 1(a)所列。在最小支持集(d,a)中，距离 a 最近的轨迹为 b，因此将 a 和 b 归纳为集合(a,b)，并分别替代 S 与 T 中的 a 和 b，形成的匿名集如表 1(b)所列。此时 $s=(a,b)$，$Sup(s,T)=1<k=2$，不满足 2^2-匿名。算法再次循环计算距离(a,b)最近的轨迹为 c，因此将(a,b)和 c 归纳为(a,b,c)，然后分别替代(a,b)和 c，形成的匿名集如表 1(c)所列。此时$s=(a,b,c)$，$Sup(s,T)=k=2$，满足 2^2-匿名。

表 1　轨迹 T 实现 2^2-匿名化的过程

(a)

subT	sup
(d,a)	1
(c,e)	1
(b,a)	1
(a,d)	1
(b,d)	1

(b)

id	trajectory
t_1	(d,{a,b},c,e)
t_2	({a,b},{a,b},e,c)
t_3	({a,b},d,e)
t_4	({a,b},d,e,c)
t_5	(d,c)
t_6	(d,e)

(c)

id	trajectory
t_1	(d,{a,b,c},{a,b,c},e)
t_2	({a,b,c},{a,b,c},e,{a,b,c})
t_3	({a,b,c},d,e)
t_4	({a,b,c},d,e,{a,b,c})
t_5	(d,{a,b,c})
t_6	(d,e)

3.2　基于前缀树的隐私保护方法

针对移动社会网络的签到服务中存在的轨迹隐私泄露风险，霍峥等人提出了基于前缀树的轨迹隐私保护方法Private-CheckIn[22]。该方法首先根据签到序列建立签到缓存序列的前缀树，在执行剪枝、重构、遍历操作后，得到新的 k-匿名的签到序列，以达到对轨迹隐私的保护。

3.2.1　PrivateCheckIn 体系结构

该方法是建立在中心服务器结构的基础上的，其体系结构如图 1 所示，由客户端、隐私保护服务器和数据服务器 3 部分组成[22]。隐私保护服务器位于客户端和数据服务器之间，是可信的第三方服务器，主要负责注册移动用户、储存用户设定的隐私参数和进行签到处理，它根据移动用户提交的“预签到”请求的位置类型进行相应的签名处理，并根据用户的签到序列生成前缀树，执行剪枝、重构操作后生成 k-匿名前缀树，遍历 k-匿名前缀树后生成符合隐私保护要求的签到序列，最后将其发送给数据服务器，完成查询服务工作。

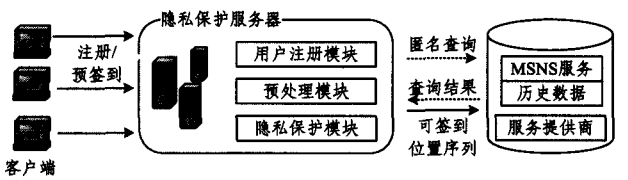

图 1　PrivateCheckIn 方法的系统结构

3.2.2　PrivateCheckIn 轨迹隐私保护方法

该方法的实质是将对数据轨迹的 k-匿名化过程变成生成签到序列的 k-匿名前缀树的过程。因此当生成签到序列的前缀树结构后，算法首先从树的根节点开始执行剪枝操作，剪去支持度小于隐私保护度 k 的节点 n_j，生成支持度不小于 k 的前缀树。然后结合剪枝过程中损失的签到序列再重构前缀树，以减小实际签到轨迹序列的损失，提高轨迹数据的可用性。最后遍历生成 k-匿名签到序列。表 2 和表 3 所列为原始的签到序列和经过剪枝处理生成的符合隐私保护要求的 k-匿名签到序列。

表 2　原始签到序列

u_1 : $L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_4 \rightarrow L_5$
u_2 : $L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_4$
u_3 : $L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_4$
u_4 : $L_1 \rightarrow L_3 \rightarrow L_4 \rightarrow L_5$
u_5 : $L_1 \rightarrow L_3 \rightarrow L_4 \rightarrow L_5$
u_6 : $L_1 \rightarrow L_3 \rightarrow L_4 \rightarrow L_5 \rightarrow L_6$
u_7 : $L_2 \rightarrow L_7 \rightarrow L_8$
u_8 : $L_2 \rightarrow L_7 \rightarrow L_8$
u_9 : $L_2 \rightarrow L_7 \rightarrow L_8 \rightarrow L_9$
u_{10} : $L_2 \rightarrow L_7 \rightarrow L_{10}$
u_{11} : $L_3 \rightarrow L_4 \rightarrow L_5$

表 3　k-匿名签到序列

u_1 : $L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_4$
u_2 : $L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_4$
u_3 : $L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_4$
u_4 : $L_1 \rightarrow L_3 \rightarrow L_4 \rightarrow L_5$
u_5 : $L_1 \rightarrow L_3 \rightarrow L_4 \rightarrow L_5$
u_6 : $L_1 \rightarrow L_3 \rightarrow L_4 \rightarrow L_5$
u_7 : $L_2 \rightarrow L_7 \rightarrow L_8$
u_8 : $L_2 \rightarrow L_7 \rightarrow L_8$
u_9 : $L_2 \rightarrow L_7 \rightarrow L_8$
u_{10} : $L_2 \rightarrow L7 \rightarrow L_8$
u_{11} : $L1 \rightarrow L_3 \rightarrow L_4 \rightarrow L_5$

3.3　基于语义匿名代理的轨迹隐私保护方法

当前针对轨迹隐私的保护仅停留在纯粹的地理位置的保护上，这种轨迹隐私保护方法具有很大的局限性，如在移动用户密度高的地方，很容易泄露用户的位置隐私和轨迹隐私。为了防范速度关联攻击，弥补单纯的轨迹隐私匿名保护的缺陷，耿技等人提出了基于语义匿名代理的轨迹隐私保护方法[23]，将语义位置融入轨迹隐私的保护中。该方法首先将城市地图进行建模，将其抽象为带有语义的无向带权图，再构造区域化的语义地图。每个构造的区域都要满足隐私门限要求 τ。当用户在当前位置区域提出服务请求时，获取当前用户在区域中的敏感位置的概率必须小于 τ，从而实现保护用户的语义轨迹隐私。

该方法首先使用希尔伯特空间填充曲线将二维地址转换成一维的数值序列；然后将所得序列组织到分布式的匿名结构 SAP-tree 中[35]，在形成的序列中随机选取提出查询的用户周围的连续的 $k-1$ 个用户形成 k-匿名集合，构建 k-匿名空间区域，其结构如图 2 所示；最后再随机选取匿名集合中的一个用户，将匿名空间区域集合发送给服务器。这样使攻击者辨不清信息的源和目的，增大了将用户与位置信息关联的难度以及将同一用户先后位置信息关联而获得用户轨迹的难度，有效地保护了轨迹隐私信息。

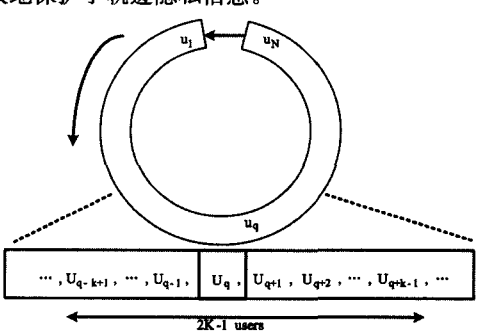

图 2　希尔伯特环和序列

该方法引入了语义匿名的概念，可有效地防范速度关联攻击[36]。通过构造区域化的语义地图，完成语义轨迹匿名；通过在 k-匿名框中随机选择用户 u 将匿名集合转发给服务器，降低了用户和位置之间的关联性，使攻击者无法区分位置和用户之间的关系，提高了隐私保护度。但是，由于转发用户也是匿名框集合中的成员，匿名框集合中本身就包含了该转发用户的信息，攻击者通过一定的背景知识也能以一定概率识别出发送请求的用户。

3.4 基于网格的 l-差异性原则

在轨迹 k-匿名模型中，为防止因 k 条轨迹相似而造成隐私泄露，要求匿名集合中的轨迹必须满足 l-差异性原则[37]。郭旭东等在此基础上提出了一种改进的 l-差异性方法[24]，将轨迹所在的二维空间划分成大小相等的单元格，将轨迹对应到相应的单元格中，通过轨迹所经过的单元格序列之间的关系来判断轨迹的相似性，从而使匿名的轨迹满足 l-差异性原则。

该方法先将轨迹数据集 D 进行 k-匿名化，转化成包含若干个等价类的轨迹集合 D^*，D 即 $D^*=\{S_1,S_2,\cdots,S_n\}$；然后对所有轨迹点的二维空间进行划分，形成对应的单元格；再根据划分的单元格，对每条轨迹 tr_i 生成对应的单元格列表，最后利用单元格列表生成轨迹的聚类，形成满足条件的匿名集[24]。

在轨迹数据集合 D^* 中，记录每个等价类中经过这个单元格的轨迹编号，然后遍历所有单元格，列举每个单元格记录的任意两条轨迹，将它们两两之间的关联度结合，最后形成一个关联度表 CT。例如轨迹 $T_1=\{1,2,3\}$ 表示其经过的单元格编号为{1,2,3}，轨迹 $T_2=\{1,3,5\}$，轨迹 $T_3=\{2,4,5\}$，那么可以得到单元格的记录表。通过分析记录表，可以得到在每个单元格中经过这个单元格的轨迹之间的关联度，最后形成关联度表，如表 4 所列。关联度表反映了数据集 D^* 中任意两条轨迹之间的关联度，利用这种关联关系，可以判断任意两条轨迹间是否相似，从而构造轨迹匿名集合。

表 4 轨迹 T_1,T_2,T_3 形成的关联度表

轨迹	T_1	T_2	T_3
T_1	3	2	1
T_2	2	3	1
T_3	1	1	3

3.5 基于时间混淆的轨迹隐私保护方法

目前许多隐私保护技术将轨迹信息简单地分为敏感信息和非敏感信息，并未考虑时间、情境等信息；而实际上轨迹的敏感性也依赖于情境和上下文信息[38]，当在一个连续的时间序列中进行基于位置的查询服务时，攻击者可能根据背景知识获得用户的轨迹信息。Huang 结合用户当前的上下文信息提出了一个时间混淆的轨迹隐私保护方法[25]，该方法通过引入时间混淆方法，综合 r-匿名、k-匿名、s-段范式 3 个隐私保护参数，来隐藏用户当前的真实轨迹隐私；同时提出了一种综合的隐私保护算法，将用户提出查询的时间序列完全打乱，再随机地发送给 LBS，从而减小攻击者通过连接匿名框来推断出移动用户的可能。

3.5.1 隐私参数

为了解决 LBS 可能泄露用户轨迹信息的问题，该方法引入了 r-匿名、k-匿名、s-段范式和时间混淆 4 个隐私保护参数。

r-匿名：这是一种预处理技术，当一个用户开始计划一条运行路线，可信的服务器会从用户的历史轨迹数据库中获得该用户的 $r-1$ 条相似的历史轨迹，使得 LBS 很难把用户的当前轨迹和其它 $r-1$ 条历史轨迹区别开来[39]。

k-匿名：每一查询位置都包含 k 个用户，使得攻击者很难将当前用户和其它 $k-1$ 个用户区别开来[40]。

s-段范式：在使用公路网络数据时，每个查询序列不仅要包含 k 个用户，还要至少包含 s 段公路。

时间混淆：匿名服务器并不依据查询请求的提出时间顺序形成的逻辑序列发送查询请求，而是完全打乱查询提出时间，然后随机地发送给 LBS 服务器，降低攻击者重构用户实际轨迹的可能性，从而防止用户轨迹隐私信息的泄露。

3.5.2 隐私保护算法结构

该隐私保护技术的主要结构如图 3 所示，主要分为 3 部分：地图网络构建、r-匿名、时间混淆。

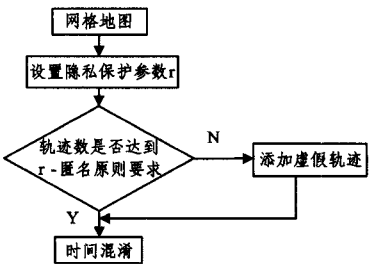

图 3 算法结构

(1)地图网络构建：区域地图被划分为固定长、宽的网格，用于储存每个对应轨迹所经过的网格数，隐私保护算法将位置查询信息的检索建立在网格的索引结构上。

(2)r-匿名：在该过程中需要输入用户的隐私设置参数 r、历史轨迹集合 R、用户当前的轨迹 T_u，如果隐私保护算法在历史轨迹集合中获得的与 T_u 相似的轨迹数少于 $r-1$，隐私保护算法会生成一定数量的虚假轨迹，实现 r-匿名。

(3)时间混淆：在该过程中需要设置用户隐私配置参数 r、k、s 和用户当前轨迹 T_u 以及用户等待返回查询结果的有效时间 $Timeout$。当前用户 T_u 向目的地出发时，匿名服务器会从网格地图中随机选择若干位置点发送给 LBS，以获得用户所需的位置服务。当用户在网格地图中未达到位置点 loc_i 时，匿名服务器可能已经将该位置点作为查询点发送给了 LBS。当用户真正到达该位置点时，首先从匿名服务器中检索用该位置点查询返回的结果，如果该结果没有超过有效时间 $Timeout$，则将该结果作为查询结果，匿名系统不再提出查询请求；否则，将重新发送查询请求。该方法可以有效提高服务器的服务效率并增强隐私保护水平。每次服务器都会向 LBS 发送多个查询，以使 LBS 无法从中识别当前用户 T_u。同时每个集合中必须包含其它 $k-1$ 个相似的用户，且必须包含 s 段路线；如果不满足，算法会继续进行搜索，直到满足匿名化要求为止[25]。

4 抑制法

抑制法是轨迹隐私保护中最常用的方法，通过对原始轨迹数据进行有选择性的发布，即限制发布某些敏感信息，来实

现隐私保护的目标。抑制法是基于攻击者背景知识的一种简单、高效的隐私保护方法，但当无法掌握攻击者的背景知识时，该方法将无法起到有效的隐私保护的作用。

4.1 基于扰动的轨迹数据隐私保护方法

针对轨迹数据发布中的隐私保护和数据可用性问题，翁国庆提出了基于扰动的轨迹隐私保护方法[26]，即用出现频率最低的同类节点来代替存在隐私泄露风险的节点，从而实现对具有隐私泄露风险的节点的抑制。

定义 2(隐私泄露概率) 在发布的数据集 T 中，攻击者 A 根据自己掌握的某投影轨迹 t^A 推测出 t^A 的拥有者可能访问过某隐私敏感节点 $P_j \notin P_A$ 的概率称为隐私泄露概率[26]。假设轨迹 t^A 的支持集为 $S(t^A,T)$，则隐私泄露概率 $P_{br}(P_j,t^A,T)$ 用公式表达为：

$$P_{br}(P_j,t^A,T)=\frac{|\{t|t\in S(t^A,T)\wedge P_j\in t\}|}{|S(t^A,T)|}$$

该方法通过以下两步来实现对移动用户轨迹隐私数据的保护：(1)首先对轨迹数据进行检测，依据攻击者可能拥有的背景知识以及用户设定的隐私保护参数，找出可能导致隐私泄露的点(P_j,t^j)。(2)利用数理统计方法，在同一个聚类分组中找到出现频率最低的节点来代替隐私泄露节点，从而形成新的轨迹，此时再循环检测新的轨迹中是否出现新的隐私泄露风险，如果没有，则进行替代操作，否则继续查找出现频率较低的节点进行替代操作，直至替代成功。如果在所有序列中都没有找到合适的替代节点，则对该隐私节点进行抑制。

该方法基于数理统计，使用出现频率较低的同类节点来代替存在隐私泄露风险的节点，在保持轨迹数据的内部结构、增加数据的可用性方面具有一定的优势[26]。但它对轨迹泄露风险的检测是建立在攻击者可能的背景知识的基础上，而在实际应用中要想穷尽攻击者所有的背景知识是很困难的。同时，在轨迹数量很多时，使用数据统计方法计算频率出现低的节点会增大系统的开销，增加系统的响应时间。

4.2 基于轨迹频率抑制的轨迹隐私保护方法

抑制法是轨迹隐私保护中的一种重要方法，通过限制发布轨迹数据中的敏感点来实现隐私保护，但如何选择合适的抑制点来平衡隐私保护和数据可用性是一个重要的研究课题。赵婧等人提出了基于轨迹频率抑制的轨迹隐私保护方法[27]，通过向有问题的轨迹数据集中添加假数据或局部抑制有问题的轨迹的方法来实现轨迹数据的隐私保护。

该方法先根据用户设定的隐私需求对轨迹数据集进行预处理，将不满足隐私需求的轨迹序列集合按照出现的频率排序，再采用添加假轨迹数据或局部、整体抑制的方法进行处理。

(1)添加假数据法。根据移动用户设定的隐私保护参数，计算达到隐私保护需求时在原始轨迹中需要添加的假轨迹数据数 *sum_add*。

(2)轨迹抑制法。按照隐私保护度和用户设定的隐私保护度进行轨迹数据的局部或整体抑制，计算达到隐私保护需求时所需要抑制的轨迹数 *sum_delete*。

(3)选择适当的方法。当 *sum_add*>*sum_delete* 时，抑制有问题的轨迹，采用轨迹数据抑制法；反之，则采用添加假数据法。

该方法对问题轨迹的判断也是基于攻击者所有可能的背景知识的，因此隐私保护最低容忍度的设置比较困难。

5 假数据法

虚假轨迹数据法的基本思想是为每一条轨迹生成一些虚假的轨迹，使用虚假的轨迹来代替真实的轨迹数据；或者通过对真实轨迹数据进行扰动形成一组虚假轨迹数据，并将其加入到原始轨迹数据中[28]。

针对移动终端在连续查询中的轨迹隐私保护问题，Feng 提出了一种简单的、分布式的轨迹隐私保护策略 Vurtual Avatar[29]，该方法通过在每个用户周围填充多个虚假的轨迹来隐藏用户的真实轨迹。当移动用户想获得一个基于位置的服务时，他不仅发送真实的轨迹信息，同时还发送多个虚假的请求信息。VAvatar 策略采用无序的虚假位置选择、查询进度安排。

以 S 为起点，D 为终点，D_1 和 D_2 为两条虚假的轨迹，其对应的区域地图如图 4 所示。假设用户的真实轨迹为 $S\rightarrow3\rightarrow6\rightarrow5\rightarrow9\rightarrow D$，整个查询时间周期可以分为 4 个阶段，在每个阶段用户将真实的查询信息和若干虚假的查询数据一起发送。每个阶段包含的位置点为：阶段一：1,2；阶段二：5,3,7；阶段三：4,9,6；阶段四：D_1,D,5,8,D_2。

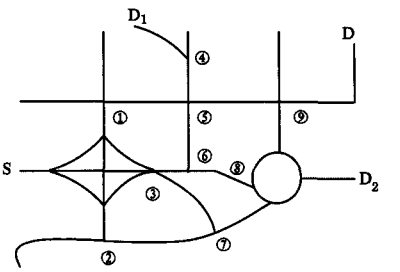

图 4 区域网络地图

所有可能形成的路径如图 5 所示，用户共发送了 13 个查询，用户发送的真实查询是无序的。在攻击者看来，每条路径都是可行的，因此攻击者无法将真实查询信息和虚假信息区别开来，从而实现了用户的轨迹隐私保护。

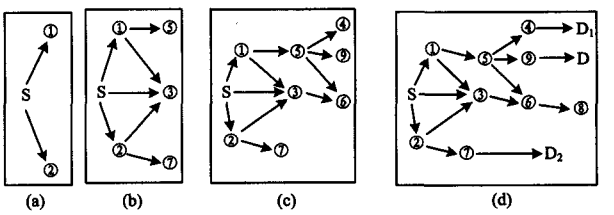

图 5 可能的路径

6 未来的研究方向展望

6.1 隐私位置点的优化处理和隐私保护匿名框的合理构造

在轨迹隐私保护中，如果不考虑用户所处的环境和上下文等信息，直接将用户所处的位置点分为敏感位置点和非敏感位置点，对敏感位置点进行匿名处理后发布，对非敏感位置点则直接发布。实际上这种方法不太现实，位置点的敏感性依赖于情境和上下文，如医院对某些病人是敏感的，而对医生则不是，不能一概而论地把医院作为敏感位置点。应该结合用户的个人信息和所处的环境、情境、上下文信息等因素来划分位置点的隐私性，从而提供个性化的隐私保护[21]。

同时，依据 k-匿名化原则形成匿名区域时，如果不考虑用户所处的实际环境，则生成的虚假匿名位置有时会不合理，就起不到隐私保护效果，如匿名区域位于山川、河流、森林、废弃的建筑物等。当准备形成匿名的区域内用户较少时，在构造 k-匿名区域的过程中，必须增加匿名区域的范围，这样势必增加系统的响应时间，影响系统的效率。

因此，需要对隐私位置点进行合理优化，科学、有效地构造匿名框。例如在设定隐私保护参数 k 的同时，再设定一个最大响应时间 T_{max}、最大匿名区域范围 A_{max}、最小匿名区域范围 A_{min}。当进行匿名保护时，综合考虑用户所处的位置点 W 和用户查询的内容 Q 等因素，从而形成一个科学、合理的匿名区域。当匿名区域形成时间超过最大响应时间 T_{max} 时，判定此次匿名失败；若形成的匿名区域超过设定的最大区域 A_{max} 时仍未达到 k-匿名要求，也判定此次匿名失败；当匿名区域小于设定的最小区域 A_{min} 时，则表示匿名效果不好，未达到要求的匿名效果，也表明此次匿名失败。

6.2 构建新的隐私保护模型，保护连续位置查询中的轨迹信息

在基于位置的连续查询中，根据轨迹形成的位置点顺序来发送查询服务请求，这种隐私保护方法并未考虑保护查询提出的时间顺序；按用户从起点到终点的顺序，将位置点逐一转换成匿名区域进行用户的隐私保护，而 LBS 收到的虽然是泛化后的用户匿名组或匿名区域，但根据匿名化的构建过程和查询信息的发送时间模式，攻击者很容易根据查询的位置点发现用户的轨迹。实际上，用户每提出一个查询服务请求，在每一个时间点都进行 k-匿名化，单纯从时间点来看，在每个时刻都满足 k-匿名要求。但如果根据用户发出查询请求的时间顺序，则可以推断出用户的运行轨迹。因此应该基于服务请求的提出时间，提出新的隐私保护模型，来保护轨迹隐私信息，实现服务效果和隐私保护的最优化合理利用[25]。例如在用户从起点到终点的运行过程中，不根据用户轨迹形成的位置点的顺序发送查询信息，而是打乱这种顺序，如用户运行 n 个位置点后，将该 n 个位置点作为一个区间，打乱顺序后随机选择其中的位置点发送服务请求。但该方法会产生服务的时间延迟，也可能降低查询的精确度，影响服务效果及数据的使用率。

6.3 大数据环境下的隐私保护技术研究

匿名化技术是当前最常用的隐私保护技术，对单一数据源的隐私保护效果比较好。但在大数据时代，面对规模大、类型多的大数据，当攻击者拥有其它一些公共的或者隐私的数据源时，通过多个数据源的融合，再利用链接攻击对匿名后的数据源进行攻击，就能够推断出个人绝大部分的敏感信息，造成个人隐私泄露[41]。此外，大数据资源的连续公开性，使得隐私管理技术面临新型的隐私攻击与隐私泄露风险，大数据引发的多源数据融合使当前的匿名技术的隐私保护效果显著降低，当前的隐私保护理论和技术无法涵盖大数据隐私的内涵，因此如何在大数据环境下突破现有隐私保护技术由于数据关联性导致的隐私泄露问题，以及如何预防资源数据融合带来的隐私威胁等问题，是未来隐私保护研究的一个重要方向。

结束语 移动社会网络中基于位置的服务的深入发展，给人们的生活带来了极大的便利，同时也给用户的个人隐私带来了极大的威胁，移动用户的位置、轨迹隐私保护成为关注和研究的重点。目前，国内外研究者在轨迹隐私保护方面已进行了深入的研究，并获得了一定的研究成果。本文对当前的研究技术和方法进行了深入的研究，分析了轨迹隐私保护中基于泛化法、抑制法和假数据法 3 种方法的研究现状以及各自的特点，并对未来的研究方向进行了展望。

参 考 文 献

[1] Chow C Y, Mokbel M F, Liu Xuan. Spatial Cloaking for Anonymous Location-based Services in Mobile Peer-to-Peer Environments[J]. GeoInformatica, 2011, 15(2): 351-380

[2] Lee K C K, Zheng Bai-hua, Chen C, et al. Efficient Index-Based Approaches for Skyline Queries in Location-Based Applications [J]. IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE), 2013, 25(11): 2507-2520

[3] Chow C-Y, Mokbel M, He T. A privacy preserving location monitoring system for wireless sensor networks[J]. IEEE Transactions on Mobile Computing, 2011, 10(1): 94-107

[4] Huo Zheng, Meng Xiao-feng. A Survey of Trajectory Privacy-Preserving Techniques[J]. Chinese Journal of Computers, 2011, 34(10): 1820-1830(in Chinese)
霍峥，孟小峰. 轨迹隐私保护技术研究[J]. 计算机学报，2011，34(10)：1820-1830

[5] Aslam B, Amjad F. PMTR: Privacy-enhancing Multilayer Trajectory-based Routing Protocol for Vehicular ad hoc Networks [C]// Proceedings of the 2013 IEEE Military Communications Conference. 2013: 882-887

[6] Mano K, Minami K. Privacy-preserving Publishing of Pseudonym-based Trajectory Location Data Set [C] // Proceedings of the 2013 International Conference on Availability, Reliability and Security. 2013: 615-624

[7] Yigitoglu E, Damiani M L. Privacy-preserving sharing of sensitive semantic locations under road-network constraints [C] // Proceedings of the 2012 IEEE 13th International Conference on Mobile Data Management. 2012: 186-195

[8] Gao Sheng, Ma Jian-feng. Balancing trajectory privacy and data utility using a personalized anonymization model[J]. Journal of Network and Computer Applications, 2014, 38(1): 125-134

[9] Chen Rui, Fung B C M. Privacy-preserving trajectory data publishing by local suppression[J]. Information Sciences, 2013, 231(1): 83-97

[10] Sui Pei-pei, Wo Tian-yu. Privacy-Preserving Trajectory Publication Against Parking Point Attacks[C]//2013 IEEE 10th International Conference on Ubiquitous Intelligence & Computing and 2013 IEEE 10th International Conference on Autonomic & Trusted Computing. 2013: 569-575

[11] Ghasemzadeh M, Fung B C M. Anonymizing trajectory data for passenger flow analysis[J]. Transportation Research Part C, 2014, 39(2): 63-79

[12] Kaplan E, Pedersen T B. Discovering private trajectories using background information [J]. Data & Knowledge Engineering, 2010, 69(7): 723-736

[13] Sanchez D, Castella-Roca J, Viejo A. Knowledge-Based Scheme to Create Privacy-Preserving but Semantically-Related Queries for Web Search Engines [J]. Information Sciences, 2013, 218(1): 17-30

[14] Tramp S, Frischmuth P, Arndt N. Weaving a Distributed, Semantic Social Network for Mobile Users[C]//8th Extended Semantic Web Conference. 2011:200-214

[15] Rowe M. Applying semantic social graphs to disambiguate identity references[C]//6th European Semantic Web Conference. 2009:461-475

[16] Ferrer J D, Rasua R T. Microaggregation and permutation-based anonymization of movement data[J]. Information Sciences, 2012,208(21):55-80

[17] Zhou Li-na, Li Ding, Finin T. How is the semantic web evolving? A dynamic social network perspective [J]. Computers in Human Behavior, 2011,27(4):1294-1302

[18] Riboni D, Pareschi L, Bettini C. Shadow attacks on users' anonymity in pervasive computing environments[J]. Pervasive and Mobile Computing, 2008,4(6):819-835

[19] Sun X, Sun L, Wang H. Extended k-anonymity models against sensitive attribute disclosure[J]. Computer Communications, 2011,34(4):526-535

[20] Bonchi F, Lakshmanan L V, Wang H W. Trajectory anonymity in publishing personal mobility data[J]. SIGKDD Explorations Newsletter, 2011,13(1):30-42

[21] Poulis G, Skiadopoulos S. Distance-based k^m-anonymization of trajectory data[C]//Proceedings of the 2013 IEEE 14th International Conference on Mobile Data Management. 2013:57-62

[22] Huo Zheng, Meng Xiao-feng, Huang Yi. PrivateCheckIn: Trajectory Privacy-Preserving for Check-In Services in MSNS[J]. Chinese Journal of Computers, 2013,36(4):716-725(in Chinese)
霍峥,孟小峰,黄毅. PrivateCheckIn:一种移动社交网络中的轨迹隐私保护方法[J]. 计算机学报,2013,36(4):716-725

[23] Geng Ji. Trajectory Privacy Protection Starategy Based on Semantic Anonymity Proxy[J]. Computer Applications and Software, 2014,31(7):307-310(in Chinese)
耿技. 基于语义匿名代理的轨迹隐私保护方法[J]. 计算机应用与软件,2014,31(7):307-310

[24] Guo Xu-dong, Wu Ying-jie, Yang Wen-jin. l-diversity algorithm for privacy preserving tra-jectory data publishing[J]. Computer Engineering and Applications, 2015,51(2):125-130(in Chinese)
郭旭东,吴英杰,杨文进. 隐私保护轨迹数据发布的 l-差异性算法[J]. 计算机工程与应用,2015,51(2):125-130

[25] Hwang R H, Hsueh Y-L. A Novel Time-Obfuscated Algorithm for Trajectory Privacy[J]. Proceedings of the 2012 International Symposium on Pervasive Systems, Algorithms and Networks. 2012:208-215

[26] Weng Guo-qing. A perturbation-based privacy preserving trajectory publication method[J]. Journal of Southeast University (Natural Science Edition), 2014,44(1):51-57(in Chinese)
翁国庆. 一种基于扰动的轨迹数据隐藏发布方法[J]. 东南大学学报(自然科学版),2014,44(1):51-57

[27] Zhao Jing, Zhang Yuan, Li Xing-hua. A Trajectory Privacy Protection Approach via Trajectory Frequency Suppression[J]. Chinese Journal of Computers, 2014, 37(10): 2096-2106(in Chinese)
赵婧,张渊,李兴华. 基于轨迹频率抑制的轨迹隐私保护方法[J]. 计算机学报,2014,37(10):2096-2106

[28] Gao Sheng, Ma Jian-feng. TrPF: A Trajectory Privacy-Preserving Framework for Participatory Sensing[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(6): 874-887

[29] Feng Yun-xia, Liu Peng. A Mobile Terminal Based Trajectory Preserving Strategy for Continuous Querying LBS Users[C]// Proceedings of the 2012 8th IEEE International Conference on Distributed Computing in Sensor Systems. 2012:92-98

[30] Wu Ying-jie. A Clustering Hybrid Based Algorithm for Privacy Preserving Trajectory Data Publishing[J]. Journal of Computer Research and Development, 2013,50(3):578-593(in Chinese)
吴英杰. 基于聚类杂交的隐私保护轨迹数据发布算法[J]. 计算机研究与发展,2013,50(3):578-593

[31] Cao J N, Karras P, Kalins P, et al. SABRE: a Sensitive Attribute Bucketization and REdistribution framework for t-closeness[J]. The VLDB Journal, 2011,20(1):50-81

[32] Sweeney L. K-anonymity: a model for protecting privacy[J]. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002,10(5):557-570

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

[34] Terrovitis M, Mamoulis N, Kalnis P. Local and global recoding methods for anonymizing set-valued data[J]. VLDB J, 2011, 20(1):83-106

[35] Chow C Y, Mokbel M F, Liu X. A Peer-to-Peer Spatial Cloaking Algorithm for Anonymous Location-based Services[C]// ACM International Symposium on Advances in Geographic Information Systems. New York, ACM Press, 2006

[36] Beresford A R, Stajano F. Mix zones: user privacy in location-aware Services[C]//Proceedings of the 2th IEEE Annual Conference on Pervasive Computing and Communications Workshops. 2004:127-131

[37] Machanavajjhala A, Gehrke J, Kifer D, et al. L-Diversity: Privacy Beyond k-Anonymity[C]// Proceedings of the 22nd International Conference on Dat Engineering, 2006. Atlanta, GA, USA, 2006:24

[38] Pingley N, Zhang Z, Fu X, et al. Protection of Query Privacy for Continuous Location Based Services[C]// Proc. IEEE INFOCOM. 2011:1710-1718

[39] Shokri R, Theodorakopoulos G, Troncoso C, et al. Protecting Location Privacy: Optimal Strategy Against Localization Attacks [C]// Proceeding ACM Conference Computer Communcations Security. 2012:617-627

[40] Yao L, Lin C, Kong X, et al. A Clustering-Based Location Privacy Protection Scheme for Pervasive Computing[C]// Proc. CoRR. 2010:719-726

[41] Meng Xiao-feng, Zhang Xiao-jian. Big Data Privacy Management [J]. Journal of Computer Research and Develoment, 2015, 52(2):3-18(in Chinese)
孟小峰,张啸剑. 大数据隐私管理[J]. 计算机研究与发展,2015,52(2):3-18