位置隐私保护技术综述*

贾金营，张凤荔
（电子科技大学 计算机科学与工程学院，成都 611731）

摘　要：随着如智能手机和平板电脑等移动设备的普及，基于位置的服务（LBS）变得越来越流行，人们通过网络进行查询的同时，将自己的位置信息暴露给了 LBS 提供商。如何保护用户的位置信息不被潜在地泄露给 LBS 提供商，对一个 LBS 系统来说是至关重要的。目前关于 LBS 的隐私保护的研究已经取得了一定的成果，为了更深入地解决位置隐私保护技术中还没有解决的诸多问题，展开对相关课题的深入研究，从非 *k*-匿名位置隐私技术、*k*-匿名位置隐私技术、P2P 架构下的 *k*-匿名技术和连续查询轨迹匿名技术四个方面对相关文献进行了综述，分别介绍了相关的算法。最后，总结了位置隐私保护技术当前存在的问题及未来的发展方向。

关键词：匿名集；匿名区；候选集；兴趣点（POI）
中图分类号：TP309.2　　**文献标志码**：A　　**文章编号**：1001-3695(2013)03-0641-06
doi:10.3969/j.issn.1001-3695.2013.03.001

Overview of location privacy protection technology

JIA Jin-ying, ZHANG Feng-li
(*School of Computer Science & Engineering*, *University of Electronic Science & Technology of China*, *Chengdu* 611731, *China*)

Abstract: As the proliferation of mobile devices such as smartphones and tablets, location-based services(LBS) are becoming increasingly popular. LBS, albeit useful and convenient, posed a serious threat to users' privacy as they are enticed to reveal their locations to LBS providers via their queries for. How to protect users' privacy against potentially compromised LBS providers is of vital importance to the well being of the LBS ecosystem. Some solutions have been proposed to deal with location privacy problems in location-based services, however, there is still a distance to be widely used. In order to resolve the contradictions which have been not solved, and deeply study with the related topics, this paper overviewed the related literatures from four aspects: the not *k*-anonymous location privacy technologies, the *k*-anonymous location privacy technologies, the *k*-anonymous location privacy in the P2P environment technologies, and the anonymous location privacy for continuously querying technologies. It introduced the related algorithms. Finally it concluded the problem remaining and development orientation of the location privacy protection technologies.

Key words: anonymous set; anonymous spatial region; candidate set; points of interest(POI)

0　引言

基于位置的服务（location-based services，LBS）是指使用某种定位技术（如全球卫星定位系统、手机定位和通过 WiFi 接入点的定位），为移动用户提供与当前的位置有关的个性化服务。基于位置的服务包括无线 110、120、119 等紧急服务，交通咨询，位置感知广告，旅游服务，基于位置的游戏和导航；结合用户当前的位置，以帮助用户寻找离用户最近的饭馆、宾馆、卫生间、医院、车站、加油站，以及娱乐场所等。在享受这些新技术带来的便利应用的同时，人们开始担心新的隐私威胁，即位置信息暴露给不受信任的第三方所带来的隐私泄露问题。位置信息是敏感的，因为人们日常生活中 80% 的信息与位置有关，并可能引出许多其他信息，例如经常光顾的地方，如医院会释放出一个人的健康状况。因此，如何在使用这些服务的同时，又不把自己的位置暴露出去，即位置隐私保护越来越受到人们的重视。近年来已经出现了由于位置信息泄露而导致的纠纷甚至打官司的事件。

1　非 *k*-匿名位置隐私技术

针对 LBS 系统存在的用户位置泄露的安全隐患，国内外不少研究人员针对 LBS 系统的安全隐患提出了许多相关的方案和算法，如冗余查询方案、SpaceTwist 方案、PIR（private information retrieaval）方案等。

1.1　冗余查询方案

文献[1]提出的冗余查询方案可以使得 SP（service provider）无法寻找到用户的真实位置。LSP（location service provider）在收到用户的 LBS 业务请求后，使用伪号码向 SP 提出用户 LBS 业务请求时，可以在一段时间内发送多个同一用户的请求。这些请求得到响应后，LSP 对于其中用户真实需要的那个定位请求按照正常的流程进行定位，对其余的冗余请求，以随机的方式向 SP 发送一些随机位置坐标。最后 SP 将查询的结果返回给 LSP 后，LSP 将其中用户真实的查询结果发回用户，丢弃其余的结果。

该方案不需要引入第三方平台，简单可行；但是无法保证

收稿日期：2012-08-12；**修回日期**：2012-09-23　　**基金项目**：国家“863”计划资助项目（2011AA010706）；国家自然科学基金资助项目（61133016）

作者简介：贾金营（1982-），男，河南鲁山人，博士研究生，主要研究方向为 LBS、GIS、移动计算等（jiajinying@126.com）；张凤荔（1963-），女，教授，博士，主要研究方向为计算机应用技术、信息安全.

用户身份不会被泄露,同时此方案还会占用 LSP 和 SP 的大量资源。

1.2 SpaceTwist 方案

文献[2]提出的 SpaceTwist 方案可以防止 POI 应用中用户位置的泄露,该方案需要引入可信的第三方平台。LSP 收到 SP 的 LBS 业务请求并定位用户后,并不是把用户位置坐标直接发送给 SP,而是将用户的所在位置坐标发送给一个第三方平台;然后第三方平台向 SP 发送的坐标并不是用户的真实坐标,而是一个虚假坐标。如图 1 所示,初始阶段(initial phase)中用“×”表示真实的用户坐标,用“·”表示名为 anchor 的虚假坐标;中间阶段(intermediate phase),SP 通过其服务器存储的地理信息找寻虚假坐标 anchor 附近的 POI,并不断地扩大搜索 POI 的范围(即 supply space),demand space 是第三方根据用户所需制定的实际需求 POI 的搜索范围;结束阶段(final phase),supply space 完全覆盖 demand space,第三方将搜索到的 POI 发回给 LSP,LSP 再将定位结果发回用户。

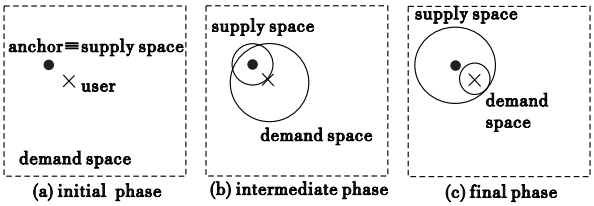

图1 SpaceTwist方案时序图

通过这一方案,使得 SP 无法得知用户真实的地理位置信息,但会增加 SP 的运算开销,延长服务时间;同时,引入第三方平台也增加了运营商的成本。另外,SpaceTwist 方案也不能保证用户的身份信息不会被泄露。

1.3 PIR 方案

文献[3]提出在 LBS 系统中引入 PIR 方案以增加其安全性。PIR 方案如图 2 所示,请求 $q(i)$ 是被加密的,任何人都无法从 $q(i)$ 中解析出 i,客户端 client 可以通过 $r(X,q(i))$ 轻易地计算出 X_i。这样便实现了客户端 Client 在不让服务器 server 知道其请求数据的情况下,秘密地获得所需的信息。

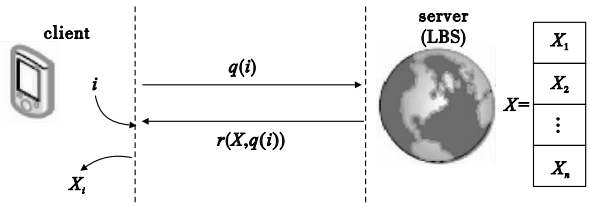

图2 PIR方案

通过引用这个方案,LSP 将每个 POI 的位置信息 p_i 作 Hilbert curve 变换,得到 $H(p_i)$,将所有 $H(p_i)$ 的信息上传到 SP 上。这个变换的参数对 SP 绝对保密,保证 SP 不能从 $H(p_i)$ 中计算出 p_i。在查询 POI 过程中,LSP 首先根据用户的位置坐标 u 计算出 $H(u)$;然后向 SP 请求最接近的 $H(p_i)$ 值;最后,对于该数值做 H^{-1} 变换,从而计算出 POI 的位置,并将其发回给用户。这一方案不需要增加第三方平台,但是所提供的位置信息的精确度较差,并且私密性保证主要依赖于 Hilbert Curve 变换参数的大量选取和更新。

2 k-匿名位置隐私技术

k-anonymity(k-匿名)保护方案可以从根本上防止用户身份信息泄露。文献[4]提出的 k-anonymity 保护方式不仅可以防止用户位置信息泄露,而且也可以防止用户身份信息泄露。k-匿名保护需要引入一个集中式的第三方平台即匿名器。当用户向服务器发送 LBS 服务请求时,先把位置信息发送给匿名器。匿名器将用户的位置坐标泛化成一个具有 k-匿名性质的区域即匿名区 ASR(anonymous spatial region),并且该区域在面积上不小于一定的值,而且区域内至少有 k 个用户,用户的身份在该区域内被识别出来的概率为 $1/k$。采用该方案时,SP 需要添加一个查询处理器,专门用来快速地处理空间区域的查询,找出所有的候选结果即候选集(candidate set,CS),并返回给用户,让用户从中选择一个最优的。虽然该方案考虑周全,能够在很大程度上解决位置隐私保护存在的问题,但是增加了 SP 的运算资源,加重了网络的负载。

2.1 常见的匿名算法

文献[5]提出了间隔匿名(interval cloak)算法。其算法的基本思想为:匿名服务器构建一个四个叉树的数据结构,将平面空间递归地用十字分成四个面积相等的正方形区间,直到所得到的最小的正方形区间的面积为系统要求的允许用户所采用的最小匿名区面积为止,每一个正方形区间对应于四叉树中的一个节点。系统中的用户每隔一定的时间将自己的位置坐标上报给匿名服务器,匿名服务器更新并统计每个节点对应区间内的用户数量。当用户进行匿名查询时,匿名器通过检索四叉树为用户 U 生成一个匿名区 ASR,间隔匿名算法从包含用户 U 的四叉树的叶子节点开始向四叉树根的方向搜索,直到找到包含不低于 k 个用户的节点(包括用户 U 在内),并把该节点所对应的区域作为用户 U 的一个匿名区 ASR。如图 3 所示,如果用户 u_1 发起 $k=2$ 的匿名查询,间隔匿名算法将首先搜索到象限区间 [(0,0),(1,1)],其中包含不少于 2 个用户。然后,它就向根的方向上升一级搜索象限区间[(0,0),(2,2)],该象限区间包含 3 个用户,大于要求的 2 个,算法停止搜索,并将该区间作为用户 u_1 的匿名区 ASR。由于该算法所得到的匿名区所包含的用户数量可能远远大于 k,因此会加大 LBS 服务器的查询处理负担和网络流量的负荷。

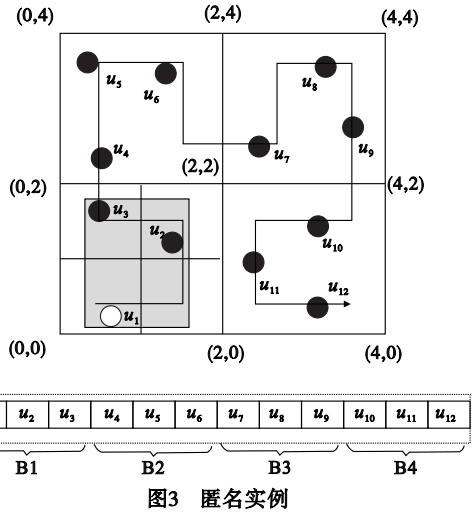

图3 匿名实例

文献[6]提出了 Casper 匿名(Casper cloak)算法,与间隔匿名算法比较类似,都采用四叉树数据结构。不同点有两个:a) Casper 匿名算法直接通过 hash 表来识别和访问四叉树的叶子节点;b) Casper 匿名算法在对节点进行搜索时,当所搜索的节点不满足 k 时,不像间隔匿名算法那样直接搜索其父节点,而

是依次检查它的两个相邻的兄弟节点和它所组合起来的区间中所包含的用户数量是否大于等于 k,如果满足则直接将组合区间作为用户的匿名区,否则再对其父节点进行搜索。并且允许每个用户自由地决定 k 值的大小和最小的匿名区面积 $A_{\min}$。如图 3 所示,如果用户 u_1 发起 $k=2$ 的匿名查询,Casper 匿名算法将首先搜索到象限区间 [(0,0),(1,1)],其中包含不少于 2 个用户;然后,先检查它的两个相邻兄弟区间[(0,1),(1,2)]和[(1,0),(2,1)],如果它们中的一个与自己组合起来的区间中包含的用户数量大于等于 k,则直接把组合起来的矩形区间作为匿名区 ASR,在这个例子中矩形[(0,0),(1,2)]将被作为匿名区 ASR。与间隔匿名算法相比,Casper 匿名算法所生成的匿名区 ASR 平均要比间隔匿名算法更小,从而提高了 LBS 的效率并减少了对网络的负荷。

文献[7,8]提出了 Hibert 匿名(Hilbert cloak)算法,也称做小集团算法(clique cloak)。文献[7]指出在某些情况下上面两种匿名算法会被攻破,并且证明了如果互惠主义被采纳,绝对匿名可以被保证。互惠主义的定义如下:

定义 1　$AS_k(u)$ 表示用户 u 匿名度为 k 的匿名集(anonymous set)。一个匿名算法当且仅当同时满足以下两个条件时,才满足互惠主义。a) $AS_k(u)$ 至少包含 k 个用户;b) 在 $AS_k(u)$ 中每一个用户 u',都有 $AS_k(u')=AS_k(u)$(也就是在 $AS_k(u)$ 中所有的用户有同一个匿名集 AS)。

Hibert 匿名算法就满足这个性质,将所有的用户依照 Hibert 空间填充曲线(Hilbert space-filling curve[9])进行整理,然后按照整理后的顺序相邻的每 k 个用户依次被放入一个水桶中。用户的匿名集就是包含有该用户的水桶内的所有用户。通过计算匿名集的最小绑定矩形从而得到匿名区 ASR。如图 3 所示,如果用户 u_1 发出一个 $k=3$ 的匿名查询,图的下面是创建 4 个 Hibert 水桶,该用户的匿名集 AS 包含用户 u_1、u_2、u_3,最小绑定矩形即匿名区 ASR,如图中阴影所示。

2.2　LBS 服务器端查询处理算法

查询一般可被分为两类,一类是范围查询,另一类是最近邻居查询。给出匿名区 ASR 及查询的类型和参数,LBS 服务器需要查找满足匿名区 ASR 中的任意一个可能的用户的 POI。一般 LBS 将 POI 放到辅助储存器中,并用 R 树来索引。对于一个范围查询,LBS 计算候选集 CS 为所有落入匿名区 ASR 中的 POI 和距离匿名区 ASR 边界距离小于 R 的 POI。如图 4(a)所示,LBS 将匿名区 ASR 扩张 R,从而转变成了一个普通的范围查询,候选集 CS 包括 P_1、P_2 和 P_3。

对于一个 K-最近邻居查询[10],候选集 CS 将包括距匿名区 ASR 内任意一点 k 个最近的 POI 的并。如图 4(b)所示,为得到一个 $k=1$ 的最近邻居查询的候选集 CS,LBS 需要检索所有落入匿名区 ASR 内的对象,也就是 P_1,并且还要检索沿着匿名区 ASR 边沿任意位置,也就是 P_2、P_3、P_5。

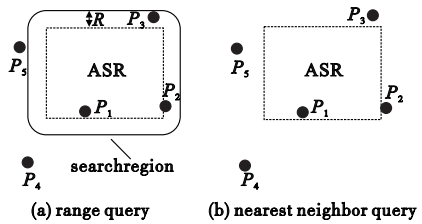

图4　LBS查询处理实例

2.3　匿名区分割算法

在用户相对比较稀少的地区,为了完成 k-匿名查询,通常匿名化后的匿名区 ASR 相当大,当使用过分大的匿名区进行查询时,会加重 LBS 服务器查询处理的负担,并且生成的候选集 CS 也会很大,回传候选集 CS 也加重了网络的负荷。为此文献[11]提出了分割匿名区的想法,将生成的过大的单一的矩形空间匿名区 ASR,划分成为几个不连续的小矩形空间匿名区 ASR,几个小的匿名区内的用户总和不小于 k,几个小的匿名区的面积之和不小于用户要求的最小匿名区面积 $A_{\min}$,并且至少有一个匿名区 ASR 包含要查询的用户 u。算法的基本思想是:采用 Hilbert 空间填充曲线将原来生成的单一的匿名区 ASR 中的用户分割到不同的 Hilbert 水桶里,然后分别生成不同 Hilbert 水桶中用户的最小绑定矩形。

2.4　提高匿名度 k

针对以固定的最小匿名度为服务质量(QoS)指标,基于位置的服务中的 k-匿名机制不能在计算资源允许的条件下为用户提供更高匿名度的问题,文献[12]提出定义新的 QoS 指标匿名结果集势指标 R, 用于度量和约束匿名服务和 LBS 为用户的每次查询所平均消耗的计算资源,在此约束下选择可接受的匿名度。其从理论上分析了匿名结果集势与匿名度的估算函数关系,据此构造相应的匿名度调节算法,从而在一次查询所允许的最大消耗不变的情况下增加 k 值来提高匿名度,为用户提供更高的安全保障。

3　P2P 架构下的 k-匿名技术

k-匿名保护需要引入一个集中式的第三方即位置匿名器,匿名效果较好,但所有信息都由第三方服务器处理,易导致信息阻塞、系统崩溃,并且第三方的可信度也很难保证;如果位置匿名器被攻击,则会泄露用户的位置。因此文献[13]引入了 P2P 的思想,取消了第三方的位置匿名器,由系统中的对等点来充当第三方位置匿名器,从而解决了集中式第三方的处理瓶颈问题。

3.1　基本算法

文献[13]给出了基于 P2P 的 k-匿名查询的基本算法。前提条件是每个用户都有两个独立的无线网络,如一个 GPRS 或者是 3G 网络,一个是 WiFi 无线网络,也可以是两个 GPRS(或 3G)即双卡双待机。一个网络用于与 LBS 通信,另一个网络用于 P2P 通信,并且系统中的用户都是安全可信的。一个完整的查询过程包括如下三步:

a)对等点查询。移动用户需要通过单跳或多跳网络查找不少于 $k-1$ 个对等点邻居。

b)生成匿名区。移动用户与他所查找到的 $k-1$ 个对等点邻居形成一个组,将他准确地隐藏到一个覆盖整个组内所有用户的一个区域中即匿名区 ASR。如果生成的 ASR 面积小于用户所要求的最小匿名区面积 $A_{\min}$,那么就需要扩大这个匿名区 ASR 到最小匿名面积 $A_{\min}$。

c)处理返回结果集。为了防止攻击者通过移动蜂窝定位技术进行攻击,移动用户需要在所形成的组内随机找一个对等点邻居作代理。通过专门用于 P2P 通信的网络将生成的匿名区 ASR 和查询的参数内容告诉代理,由代理通过另一个网络

与 LBS 服务器联系,发送查询参数和匿名区 ASR,以及接收候选集 CS,然后代理通过专门用于 P2P 的网络将候选集 CS 传回给查询用户。最后查询用户对返回的候选集 CS 进行过虑,得到查询的结果。

3.2 单跳网络内对等点邻居不足问题

为了达到 k-匿名的要求,用户必须查询到至少 $k-1$ 个对等点邻居。由于一个 WLAN 基站下的用户可能达不 $k-1$ 个,为了完成任务,文献[13]给出的对等点查询算法中直接通过增加网络路由的跳数来完成发现多个邻居的任务。

然而多跳通信会增加一定的网络传输延迟,并且增加了网络通信负担。为了更快速地完成查询任务,文献[14]提出了更好的解决算法,其算法的基本思想是:系统中的每个用户在进行一次匿名查询后,都暂时储存所用过的对等点邻居信息和相应的匿名集 AS。当一个用户要进行匿名查询时,他首先查找自己上次用过的匿名集 AS,如果上次查询时间离现在时间的间隔不大,满足要求,则直接用自己上次用过的匿名集 AS,由于移动用户一般都不停地以不确定的速度和方向移动,所以需要以上次邻居所在的地方为圆心,以上次查询的时间到现在时间的间隔与移动用户最大的合理移动速度的积为半径来确定邻居现在的位置范围,然后根据现在的邻居的位置范围生成最小绑定矩形即匿名区 ASR;如果自己上次查询的时间间隔大于一定的值,则不能使用自己的,应去查询自己直接单跳网络内邻居节点上暂存的匿名集,如果匿名集的用户数不少于 $k-1$ 个,并且时间间隔满足要求,则利用邻居的匿名集 AS 加上用户自己形成新的匿名集,然后采用上面的方法对匿名集内用户的位置进行修正,最后生成新的匿名区 ASR。如果邻居的匿名集 AS 也不能用,则只有通过增加网络的路由跳数来解决问题。

3.3 匿名区调整算法

在移动环境中,移动用户通过单跳和多跳网络查找对等点邻居,如果这些对等点分布得非常均匀,那么就很容易导致查询用户出现在匿名区 ASR 的中心。在这种情况下攻击者如果能获得匿名区 ASR,将很容易发现用户的位置坐标。为了防御这种攻击,移动查询用户就得查找多于 $k-1$ 个对等点邻居,从而随机交换他在匿名区 ASR 中的位置。

文献[14]给出了一种算法,其算法的思想为:首先,如图 5(a)所示,在匿名查询的需求用户 u 的匿名集 AS 中随机选择一个用户 P,计算出匿名区 ASR 的中心位置为 C_A;如果用户正好是离匿名区 ASR 中心最近的用户,则不对匿名区 ASR 进行调整;否则,在匿名集 AS 中找到离用户 P 最近的用户 P_N,以用户 P 为圆心,以用户 P 和用户 P_N 之间的距离的一半为半径画圆,交匿名区中心点 C_A 与用户 P 点之间的连线于点 M;然后生成一个大于 C_AM 之间的距离、小于 C_AP 之间的距离的随机数 R,将 C_A 向着用户 P 的方向移动距离 R,从而得到新的匿名区 ASR 的中心 C_A';最后,如图 5(b)所示,计算出新的匿名区中心点 C_A' 与原匿名区中心点 C_A 在 x 坐标和 y 坐标的差即 Δ_x 和 Δ_y,将原匿名区 ASR 的两个边界分别平移 $2\Delta_x$ 和 $2\Delta_y$ 即得到新的匿名区 ASR。

文献[15]给出了 RRS 匿名区调整算法,RRS 算法采用圆形的匿名区 ASR,其算法的基本思想如下:首先对生成用户 m 的匿名区 ASR 得到圆 C,圆 C 的圆心坐标为(c_x,c_y),圆 C 的半径为 r,用户 m 的坐标为(m_x,m_y);然后生成两个满足 $m_x-c_x-\frac{r}{\sqrt{2}}\leqslant\delta_x\leqslant m_x-c_x+\frac{r}{\sqrt{2}}$和 $m_y-c_y-\frac{r}{\sqrt{2}}\leqslant\delta_y\leqslant m_y-c_y+\frac{r}{\sqrt{2}}$的随机数 δ_x 和 δ_y,以坐标$(c_x+\delta_x,c_y+\delta_y)$为新的圆心,以原半径 r 为半径,得到一个新圆 C';最后重新计算落入圆 C'内的对等点邻居数,如果小于 $m-1$,则扩大圆 C'的半径后继续搜索新的对等点邻居,直到满足要求为止,所得到的圆 C'就是调整后的匿名区 ASR。由于随机数 δ_x 和 δ_y 满足两个特定的条件,保证了新生成的匿名区 ASR 中肯定包含用户 m。

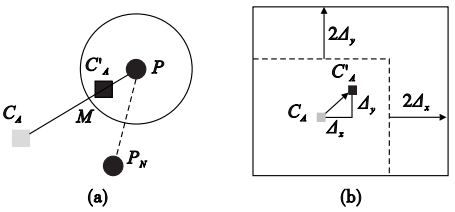

图5 匿名区中心调整和匿名区调整实例

文献[16~22]都对 P2P 架构下的 k-匿名位置隐私查询技术进行了研究,限于篇幅不在对其内容和算法进行详细阐述。

4 连续查询轨迹匿名技术

4.1 连续查询攻击案例

在连续查询匿名保护算法中,用户每次查询,匿名器都通过匿名算法将用户泛化到有 k 个用户的匿名集 AS 中,使攻击者无法从匿名集 AS 中区分出真实查询的用户。如果攻击者得知用户为连续查询,就可以采用连续攻击模型找出是真实查询的用户。假设用户分别在$\{t_i,t_{i+1},t_{i+2}\}$时刻发送了查询信息,如图 6 所示。

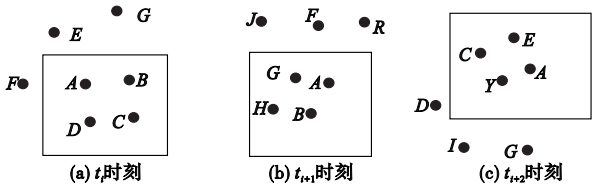

图6 连续攻击模型

若系统的匿名度 $k=4$,则攻击者在任意时刻攻击匿名服务器都会得到含有四个用户的匿名集。攻击者按照时间序列$\{t_i,t_{i+1},t_{i+2}\}$得到用户匿名集序列分别是$\{A,B,C,D\}$,$\{A,B,G,H\}$和$\{A,C,E,Y\}$。对于每次的查询,攻击都不能确认是匿名集中哪个用户进行的查询。若攻击者为持续监测,可以通过观察不同时刻的匿名集内的用户组合来推测是哪个用户发送了查询消息,攻击者可按时间序列将相邻的匿名集做取交处理,得到结果序列是$\{A,B,C,D\}$,$\{A,B\}$和$\{A\}$,由此可推测出发送消息的用户。

4.2 连续查询轨迹匿名算法

文献[23]提出并实现了一个基于移动终端的用户移动轨迹保护方法,该方法采用独立式系统结构,全部匿名保护计算均在移动端即手机上进行,不借助于可信的第三方匿名服务器。根据用户对匿名安全性高低不同的需求,用户采用不同的频度来发送虚假查询,安全需求高则发送虚假查询的频度也相应地提高。为了生成合理的不容易被识别出来的虚假查询位置坐标,根据用户的运动方式,选择一个合理的运动速度 v,如骑自行车的最大速度为 25 km/h,然后根据上次查询的时间与现在的时间差 Δt,以上次查询点坐标为圆心,以 v 乘以 Δt 的积

为半径画一个圆,根据路面信息在圆内随机选择一个点作为虚假查询的查询点坐标。通过不断地发送虚假查询从而实现隐匿用户运动轨迹的目的。

但这种发送虚假查询的方法并不能起到 k-匿名的效果,并且如果攻击者可以依据移动蜂窝定位的信息,假设所生成的虚假查询的位置所属的移动蜂窝区内正好没有一个移动用户,这时攻击者就很容易判别出这是一个虚假的查询。不断地发送虚假查询,不仅加重了 LBS 服务器的负担,还增加了移动网络负载。

文献[24]利用移动用户的历史轨迹进行连续 k-匿名保护。其基本思想为:a)系统收集系统中移动用户的位置轨迹,保存到轨迹数据库中,并随着系统中每一位移动用户的移动,不断地将新的运动轨迹添加轨迹数据库中,将时间过长的运动轨迹删除以节省空间;b)当移动用户进行匿名查询时,匿名服务器将从轨迹数据库中搜索从过去某个时间起与查询用户的历史运动轨迹大致相同的 $k-1$ 个用户的历史运动轨迹,并通过一定的算法对这 $k-1$ 个轨迹进行拟合处理,在拟合过程中由于轨迹数据库中存储的是用户某个时间点的位置,当某个相近的轨迹不能与查询用户的轨迹进行拟合时,还要对该轨迹进行时间上的相应的插值处理,如图 7 所示,图中黑色脚印为查询发起的移动用户的历史运动轨迹,查询的匿名度为 $k=3$,白色和灰色脚印为搜索到的与查询发起用户运动轨迹相似的移动用户的运动轨迹,大圆圈为生成的历史匿名区 ASR;c)将找到的能与查询用户进行轨迹拟合的 $k-1$ 个用户作为查询用户的匿名集 *AS*,在该移动用户以后的所有匿名查询过程中全部采用该匿名集 *AS*,根据匿名集 *AS* 中的用户的现在的位置坐标生成初步的匿名区 ASR,由于匿名集 *AS* 中的移动用户的运动方向和速度是不确定的,生成的匿名区 ASR 可能会因面积相当大而不能用,这时采用 2.3 节中的匿名区分割算法将初步生成的匿名区 ASR 进行分割,从而减小匿名区 ASR 的总面积,根据修改后的匿名区 ASR 进行匿名查询。

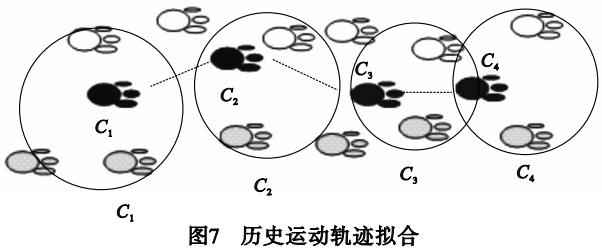

图7　历史运动轨迹拟合

该算法采用的思想是基于过去运动轨迹相近的用户,现在的位置和未来的位置也相近的概率很大的现象,如在同一长途汽车上或同一公交车上等。在移动用户进行匿名查询时始终采用同一个匿名集 *AS* 来生成最新的匿名区 ASR,从而从根本上解决了一个移动用户在运动过程中采用不同匿名集 *AS* 所导致的不同匿名集 *AS* 的链接攻击。然而,通过历史运动轨迹所求出的匿名集中的移动用户,在用户进行匿名查询时,可能断网或者没电关机,从而匿名集 *AS* 中的用户可能少于 k 个,最终导致算法的失败。

文献[25]提出了 silent period 方法,通过在邻近的用户之间构造 mix-zone,通过降低用户不同位置信息之间的可连续性来保护用户的运动轨迹。文献[26~29]都对连续 k-匿名保护进行了研究,限于篇幅不再对其算法进行详细阐述。关于连续查询轨迹匿名保护方面,到目前为止还没有一个真正完美的解决方案和算法。

5　存在的问题和未来的发展方向

虽然已有大量的文献对 k-匿名位置隐私查询技术进行了研究,但是目前还存在着一些没能完全解决的问题。

a)集中模式下,需要引入完全可信的第三方平台充当匿名器。当查询高峰时,匿名器成了系统处理的瓶颈。并且第三方匿名器也会成为攻击者的首选攻击目标,如果匿名器被攻破,系统内所有的用户位置信息将全部泄露。

b)P2P 模式下,由于对等网络分裂所导致的用户数量不足的问题(两个位置很近的无线网络之间没有中继导致不能相互通信),从而使查询要么降低匿名度 k,要么等待新对等点邻居的加盟或停止查询。由于 P2P 方式会产生大量的网络流量,所以一般用于 P2P 的网络都为 WLAN,而手机和 PDA 在启用 WiFi 通信后,会非常费电,并且由于不存在第三方匿名器,移动端的计算量也会加大,从而大大缩短移动端电池的使用时间。由于移动计算的先天性弱点,电池电量不足[30]问题导致移动终端关机,从而严重影响用户的使用效果。WiFi 的覆盖距离一般不超过 300 m,在城市的很多地方还存在盲区,在远郊区和农村地区基本上是不能使用的。基于 P2P 的 k-匿名位置隐私查询非常耗费移动端资源,所以会有一部分搭便车用户在自己查询后就关掉网络,或者退出软件和相关的服务,导致对等点邻居数量不足,从而无法完成匿名查询任务。最后,如果攻击者就是查询用户的发起用户的对等点邻居,并且还充当了代理的角色,那么查询用户的位置信息将完全泄露。

c)混合模式下,虽然兼有集中模式和 P2P 模式两者的优点,解决了匿名器的处理瓶颈和对等点邻居不足问题,但是,如果其集中的匿名器被攻击者攻破,同样会导致隐私泄露。

连续查询 k-匿名轨迹保护,同样存在上述问题,于是现在一部分学者又开始了几年前的非 k-匿名算法的研究,但目前非 k-匿名的算法在安全性上还不能与 k-匿名算法相比。

未来,位置隐私保护技术会主要向基于 k-匿名算法和连续查询 k-匿名轨迹保护的算法方向发展,其匿名区生成算法将由基于欧氏平面二维空间的算法向基于具体的道路交通网络[31]的算法方向发展。集中模式下的匿名器的安全问题和 P2P 模式下的对等点邻居的可信问题以及非 k-匿名算法和 k-匿名算法两者的融合算法很可能会成为研究的热点。

6　结束语

国内外大量文献对位置隐私保护技术进行了研究,提出了很多有用的方案和算法。为了提高安全性,有关学者把用于数据库信息发布的 k-匿名成功地移植到了位置隐私保护。为了不依赖于可信的第三方,有关学者又将无线 P2P 技术与 k-匿名位置查询技术相结合,提出了一系列基于无线 P2P 架构的 k-匿名位置查询算法。由于移动用户在运动过程中连续性地使用 k-匿名位置查询,从而导致了链接攻击问题,于是又有很多学者对连续运动轨迹匿名保护进行了大量的研究。本文以上述内容为线索,通过对大量相关文献的阅读,对相关问题和算法进行了综述和评价,并分析了目前所存在的问题和未来的发展方向。

参考文献:

[1] KIDO H, YANAGISAWA Y, SATOH T. An anonymous communication technique using dummies for location-based services[C]//Proc of International Conference on Pervasive Services. [S. l.]: IEEE Press,2005: 88-97.

[2] YIU M L, JENSEN C S, HUANG Xue-gang, *et al*. SpaceTwist:managing the trade-offs among location privacy, query performance, and query accuracy in mobile services[C]//Proc of the 24th International Conference on Data Engineering. [S. l.]: IEEE Press, 2008: 366-375.

[3] GHINITA G, KALNIS P, KHOSHGOZARAN A, *et al*. Private queries in location based services: anonymizers are not necessary[C]// Proc of ACM SIGMOD International Conference on Management of Data. New York: ACM Press,2008:121-132.

[4] MOKBEL M F, CHOW C Y, AREF W G. The new Casper:a privacy-aware location-based database server [C]//Proc of the 23rd IEEE International Conference on Data Engineering. [S. l.]: IEEE Press, 2007:1499-1500.

[5] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//Proc of the 1st International Conference on Mobile System, Applications and Service. New York:ACM Press,2003:81-84.

[6] MOKBEL M F, CHOW C Y, AREF W G. Casper*:query procesing for location services without compromising privacy[J]. ACM Trans on Database Systems,2009,34(4):24-48.

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

[8] GEDIK B, LIU Ling. Protecting location privacy with personalized *k*-anonymity: architecture and algorithms[J]. IEEE Trans on Mobile Computing,2008,7(1):620-629.

[9] BUTZ A R. Alternative algorithm for Hilbert's space-filling curve [J]. IEEE Trans on Computers,1971,20(4):424-426.

[10] JANG M Y, CHANG J W. A new K-NN query processing algorithm enhancing privacy protection in location-based services[C]//Proc of the 11th IEEE International Conference on Computer and Information Technology. Washington DC: IEEE Computer Society, 2011: 421-428.

[11] TAN K W, LIN Yi-min, MOURATIDIS K. Spatial cloaking revisited:distin guishing information leakage from anonymity[C]//Proc of the 11th International Symposium on Advances in Spatial and Temporal Databases. Berlin:Spinger-Verlag,2009:113-117.

[12] 杨朝晖,李善平,林欣. LBS 中面向 *k*-匿名服务资源约束的匿名度调节算法[J]. 浙江大学学报:工学版, 2011,45(7):1154-1160.

[13] CHOW C Y, MOKBEL M F, LIU Xuan. A peer-to-peer spatial cloaking algorithm for anonymous location-based services [C]//Proc of the 14th ACM International Symposium on Advances in Geographic Information Systems. New York:ACM Press,2006:171-178.

[14] 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.

[15] ZHANG Cheng-yang, HUANG Yan. Cloaking locations for anonymous location based services: a hybrid approach [J]. Geoinformatica, 2009,13(2):159-182.

[16] MOKBEL M F, CHOW C Y. Challenges in preserving location privacy in peer-to-peer environments[C]//Proc of the 7th International Conference on Web-Age Information Management. Washington DC: IEEE Computer Society,2006:56-64.

[17] GHINITAL G, KALNIS P, SKIADOPOULOS S. Mobiide:a mobile peer-to-peer system for anonymous location-based queries[C]//Proc of the 10th International Symposium on Advances in Spatial and Temporal Databases. Berlin:Springer-Verlag,2007:516-523.

[18] HASHEM T, KULIK L. Safeguarding location privacy in wireless Ad hoc networks[C]//Proc of the 9th International Conference on Ubiquitous Computing. Berlin:Springer-Verlag,2007:372-390.

[19] GUO Yan-hua, GUO Yi-min, GUO Ya-jun, *et al*. A trajectory privacy preserving method in peer-to-peer LBS system[J]. Advanced Materials Research,2010,34-35:571-575.

[20] THORPE C, MURPHY S, MURPHY L. IEEE 802.11 k-assisted load balancing (K-LB) for peer-to-peer content distribution over mobile SCTP [C]//Proc of International Wireless Communications and Mobile Computing Conference. New York:ACM Press,2011:683-688.

[21] TUNG Y C, LIN K C J. Location-assisted energy-efficient content search for mobile peer-to-peer networks[C]//Proc of IEEE International Conference on Pervasive Computing and Communications Workshops. [S. l.]: IEEE Press, 2011:477-482.

[22] TSUNODA T, SEZAKI K. Efficient discovery technique of distributed KML data in mobile Ad hoc network[C]//Proc of the 11th International Conference on Advanced Communication Technology. Piscataway:IEEE Press,2009:751-756.

[23] 陈洌,冯云霞,戴国强. LBS 中基于移动终端的连续查询用户轨迹隐匿方法[J]. 计算机应用研究,2011,28(12):4653-4656.

[24] XU T, CAI Ying. Exploring historical location data for anonymity preservation in location-based services[C]//Proc of the 27th International Conference on Computer Communications. [S. l.]: IEEE Press,2008:547-555.

[25] HUANG Le-ping, MATSUURA K, YAMANE H, *et al*. Enhancing wireless location privacy using silent period[C]//Proc of IEEE Wireless Communications and Networking Conference. [S. l.]: IEEE Press, 2005:1187-1192.

[26] MANWEILER J, SCUDELLARI R, COX L P, *et al*. SMILE: encounter-based trust for mobile social services[C]//Proc of the 16th ACM Conference on Computer an Communications Security. New York :ACM Press,2009:246-255.

[27] XIONG Xiao-peng, MOKBEL M F, AREF W G. SEA-CNN:scalable processing of continuous K-nearest neighbor queries in spatio-temporal databases[C]//Proc of International Conference on Data Engineering. Washington DC:IEEE Computer Society,2005: 643-654.

[28] XU T, CAI Ying. Location anonymity in continuous location-based services[C]//Proc of the 15th ACM International Symposium on Advances in Geographic Information Systems. New York: ACM Press, 2007:300-307.

[29] CHOW C, MOKBEL M. Enabling private continuous queries for revealed user locations[C]//Proc of the 10th International Symposium on Advances in Spatial and Temporal Databases. Berlin:Springer-Verlag,2007:258-273.

[30] MA R T B, LEE S C M, LUI J C S, *et al*. Incentive resource distribution in P2P networks[C]//Proc of the 24th IEEE International Conference on Distributed Computing Systems. [S. l.]: IEEE Press, 2004:516-523.

[31] MOURATIDIS K, YIU M L. Anonymous query processing in road networks[J]. IEEE Trans on Knowledge and Data Engineering, 2010,22(1):2-15.