

面向关联攻击的轨迹匿名方法

张磊^{1,2}, 马春光¹, 杨松涛^{1,2}, 李增鹏¹

(1. 哈尔滨工程大学计算机科学与技术学院, 黑龙江 哈尔滨 150001;
2. 佳木斯大学信息电子技术学院, 黑龙江 佳木斯 154007)

摘要: 针对用户在申请位置连续查询服务时, 不同移动类型产生的轨迹差异泄露用户位置隐私的情况, 提出一种相似轨迹实时生成方法。该方法实时计算可产生相似轨迹的连续位置, 通过在生成的位置上添加虚假用户与真实用户建立匿名组, 满足在连续查询过程中实时轨迹匿名的需求, 弥补了当前轨迹隐私保护方法都针对已形成的轨迹进行匿名、无法满足实时位置更新下用户位置隐私保护方面的不足。在这种方法下, 提出了一种根据用户移动方向、速度等移动类型一致性计算的轨迹匿名方法, 提高了生成轨迹的相似程度。同时, 该方法充分考虑生成位置与真实用户位置之间存在的差异, 在保障轨迹相似的基础上, 进行了位置筛选, 降低了攻击者通过生成轨迹的不可到达性辨识出真实轨迹的概率。实验表明, 该方法在有效保护连续查询用户位置隐私的前提下, 具有很好的执行效率。

关键词: 隐私保护; 连续查询; 位置匿名; 相似轨迹

中图分类号: TP311

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2017.00130

Trajectories anonymous algorithm for association attack

ZHANG Lei^{1,2}, MA Chun-guang¹, YANG Song-tao^{1,2}, LI Zeng-peng¹

(1. College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China;
2. College of Information and Electronic Technology, Jiamusi University, Jiamusi 154007, China)

Abstract: During the location based service of continuous query, the trajectories which were produced by different movement types may reveal the privacy of user's location. To solve this problem, a mechanism was proposed to put forward the real-time similar trajectories. With this mechanism, the subsequent position was calculated and which the dummy one was generated to establish an anonymous group, the demands of real-time trajectory anonymity were met, compensating the shortage of real-time trajectories of anonymity in other similar kinds of algorithm. Under this mechanism, a method for calculating subsequent anonymous position was proposed, which used the user's movement patterns, such as mobile direction, speed. The dummy trajectories were made very similar with the trajectory produced by the continuous query user. At the same time, the position of reachable difference in real areas was put forward, and on the basis of position reachable, this mechanism also filtrates out the unreachable candidate locations which had been calculated in the phase of dummy position generates, then decrease the probability of adversary identifies the real trajectory from the location of unreachable. Experimental results show that this method is effective for protecting the location privacy of continuous query user, and provides a good working efficiency.

Key words: privacy protection, continuous query, trajectory anonymous, similar trajectories

收稿日期: 2016-11-08; 修回日期: 2017-02-29. 通信作者: 马春光, machunguang@hrbeu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61472097); 高等学校博士学科点专项科研基金资助项目 (No.20132304110017); 黑龙江省自然科学基金项目 (No.F2015022)

Foundation Items: The National Natural Science Foundation of China (No.61472097), Specialized Research Fund for the Doctoral Program of Higher Education (No.20132304110017), The Natural Science Foundation of Heilongjiang Province (No.F2015022)

1 引言

随着移动通信和定位技术的不断发展与完善,基于位置服务(LBS, location based service)正逐渐深入人们日常生活中。人们在享受这种服务带来便利的同时,不可避免地要面对隐私泄露问题。在众多隐私信息泄露问题中,位置隐私最为突出。当前以 k -匿名^[1]、 l -多样性^[2]、用户协作^[3]、锚点^[4]和PIR^[5]为代表的基于快照(snapshot)服务的方法已取得较好的隐私保护效果。然而,在实际使用中,用户通常会在不同位置提出连续查询请求,这些位置可被攻击者通过关联的方法构成位置轨迹。位置轨迹通常表现出比单个位置更多的时空关联信息,这些信息又进一步增加了用户隐私泄露的风险。如何对已产生和即将产生的位置轨迹信息加以保护,降低用户隐私泄露风险,成为当前研究者关注的重要问题。

位置轨迹隐私保护最早起源于数据发布,基于 k -匿名的思想分别从单元抑制^[6]、单元泛化^[7]、轨迹聚类^[8]、转化泛化^[9]和全局轨迹相似^[10]等方面被提出。Bonchi等^[11, 12]详细分析并总结了这些基于数据发布的隐私保护方法。然而,通过分析可以看出,以上方法并不能保护实时查询服务下的用户隐私。因此,针对实时产生位置轨迹的情况,研究者提出了匿名^[13-21]和mix-zone^[22-26]两类主要的隐私保护方法^[27]。其中,匿名方法较mix-zone更易部署,且易于优化调整,因而被广泛接受和使用。在实际部署过程中,由于匿名区域的限定,以及查询次数、查询距离与服务质量平衡等问题,一般使用连续匿名框进行匿名。这

种方法在每个连续的匿名框内寻找至少 $k-1$ 个相似用户,使攻击者无法确定 k 个用户中真实用户的所处位置。这种连续匿名的方法可以确保攻击者无法将连续的两次查询位置相关联,进而保护用户的位置隐私。例如,图1(a)表示的是用户A、B和C经过匿名后扰乱的位置轨迹,其中,实线表示真实轨迹,虚线表示模糊轨迹。由于匿名用户可产生多个不确定的后续位置,使攻击者无法确定用户产生的真实轨迹,在一定程度上保护了用户的隐私。然而,这种方法存在被移动类型关联用户轨迹的可能。假设攻击者通过观察或其他方式获取用户的真实移动类型,那么即使经过匿名方法模糊了用户的位置轨迹,攻击者仍然可以通过轨迹差异获得不同用户的真实位置轨迹如图1(b)所示。若A是待攻击用户,则根据3个用户不同的移动类型,攻击者可以正确识别出用户A。因此,需要建立一种轨迹匿名方法,使用户能够产生如图1(c)所示的位置轨迹,令攻击者无法分辨出真实的待攻击用户,进而保护用户的轨迹隐私。

全局轨迹相似^[10]可以实现如图1(c)所示的轨迹匿名,但其实时性较差,不适合在时效性要求较高的位置服务中使用。因此,需要针对全局轨迹相似的隐私保护思想,设计一种相似轨迹实时生成方法,最大限度地模糊攻击者可获取的用户位置轨迹,保护连续查询过程中用户的位置隐私。

2 相关工作

2003年,Gruteser等^[1]针对位置隐私保护问题,从数据发布的隐私保护中引入了 k -匿名方法,

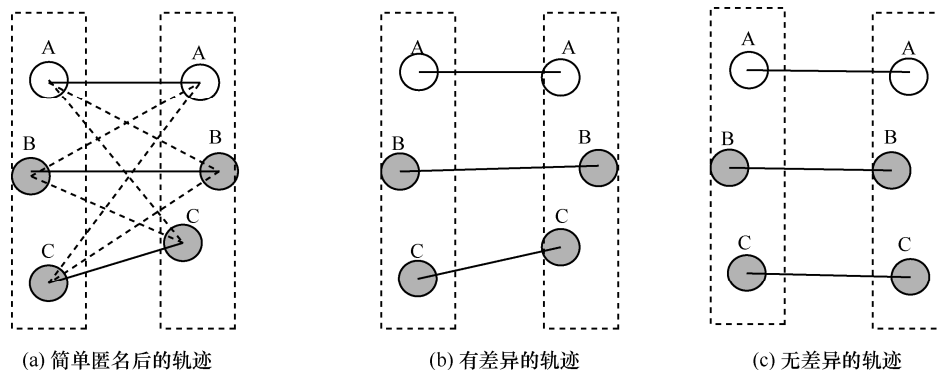


图1 用户轨迹示意

通过与 $k-1$ 个其他用户同时提出服务请求, 使攻击者无法从 k 个用户中准确分辨出待攻击用户所处的真实位置。Bamba 等^[2]针对可能存在的同质攻击, 提出了 l -多样性的概念, 丰富了 k -匿名技术。由于在用户分布稀疏的情况下, 寻找足够的匿名参与用户需要较大的匿名框, Wang^[19]和 Jia^[20]等分别提出了缩小匿名区域的解决方法, 并且通过单一匿名区域实现了连续查询的位置隐私保护。针对潜在的关联概率推测方法, 张磊等^[21]提出了关联概率相似性的 k -匿名方法。针对连续查询中的假名关联和内容关联攻击, Beresford 等^[22]提出了 mix-zone 技术。mix-zone 可以看作是一个固定的黑盒区域, 用户驶入该区域时, 可以与该区域中其他用户进行身份交换。由于在该区域中攻击者无法获知用户身份标识的更改过程, 从而无法获知更改后的身份标识与用户之间的关联情况。Freudiger 等^[23]优化了这种技术, 使在实际部署中降低了对 mix-zone 的部署个数要求。同时, Freudiger 等^[24]又将 mix-zone 技术引入路网环境中。针对矩形 mix-zone 无法有效模糊速度和方向等移动类型的情况, Palanisamy 等^[25]提出了非矩形 mix-zone 区域的概念, 并针对时间关联攻击提出了时空耽搁的 mix-zone^[26], 以此降低时空关联特性推测待攻击用户位置的概率。同样针对假名

关联攻击, 潘晓等^[13]出于对静态匿名的方法不足以保护连续查询位置隐私的考虑, 提出了利用移动对象时序相似性计算来保护用户位置隐私的方法。Hashem 等^[14]针对连续匿名区域重叠可被攻击者根据查询内容关联用户真实位置的情况, 提出了将连续的匿名区域分离, 并对可能的服务请求进行预检索的方式, 降低匿名区域重叠产生的频率, 防止匿名差异导致的位置信息泄露。张磊等^[28]针对轮廓关联问题提出了轮廓泛化的位置隐私保护模型。然而, 这些方法只能保护假名关联和查询内容关联攻击下的位置隐私, 并不能有效地防止攻击者通过时间间隔关联获取用户的位置轨迹。例如, 在图 2 中, 攻击者可以根据用户 A 在连续查询过程中的时间间隔差异, 辨识出 A 的连续查询位置, 进而获得 A 位置轨迹隐私。

针对这种时间间隔关联攻击, Hwang 等^[15]提出了一种时间模糊的方法。该方法基于匿名服务器中保存的历史数据, 使用历史查询数据混淆当前查询时间, 进而模糊潜在的时间间隔关联特性, 以此保护用户在时间间隔关联攻击下的轨迹隐私。但是, 该方法存在一定的局限性, 其模糊过程需要匿名服务器具有较高的存储能力和数据计算能力。因此, 在提供隐私保护的过程中, 可

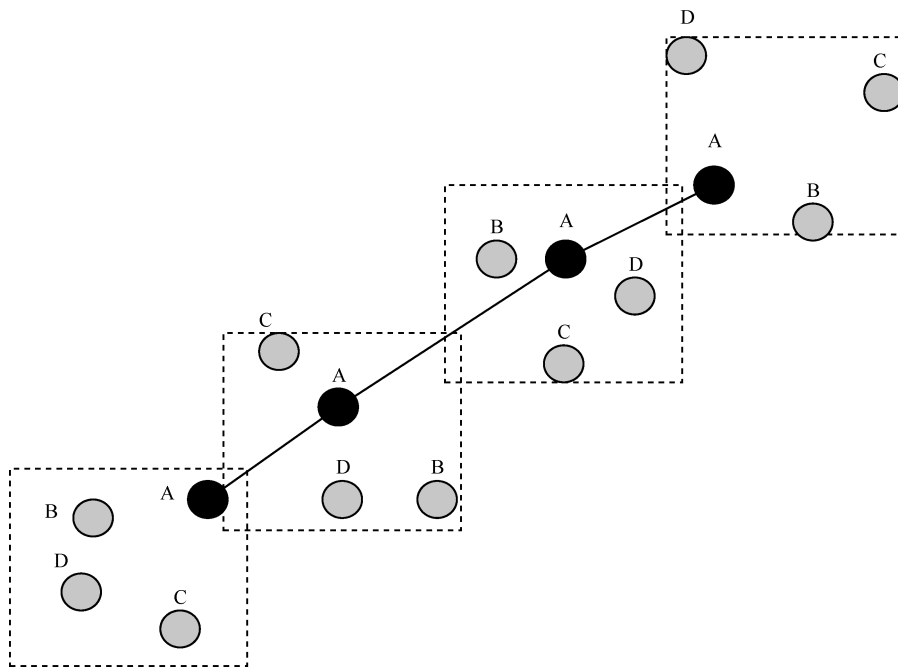


图2 时间间隔差异导致的位置轨迹泄露

能会出现服务瓶颈问题,同时这种方法并没有考虑攻击者根据移动类型进行关联攻击的情况。对于潜在的移动类型关联攻击,Gao等^[16]通过一种用户协作的方法,实现了用户轨迹的移动类型相似,其过程是:首先,协作用户统一使用与基用户相同的查询假名;然后,协作用户选择与基用户相同的移动方向;最后,在相同的时间间隔内提出位置服务请求。然而,这种方法仅能保护有限的查询次数,在多次查询的情况下,很难保证协作用户与基用户产生轨迹的移动类型相似程度。Wang等^[17]同样基于移动类型相似进行隐私保护,但其方法未能有效地对用户移动方向进行量化,导致所提出的算法存在轨迹差异性的攻击漏洞。针对这一系列的攻击问题,本文借鉴文献[18]关于轨迹相似性计算的思想,通过计算相似轨迹位置的方法,实时地获取相似轨迹,进而完成轨迹匿名。这种相似轨迹实时生成方法,模糊了用户产生的位置轨迹,降低了攻击者获取用户位置轨迹隐私的概率,较为有效地保护了用户的个人隐私。

3 预备知识

本节主要介绍相似轨迹实时生成方法采用的系统架构、基本概念以及潜在的攻击手段。

3.1 系统架构

基于位置服务隐私保护的系统架构主要分为2种:双层架构和3层架构。双层架构由用户和位置服务提供商(LSP, location service provider)

组成。该结构具有简单、无需可依赖第三方等特点,但存在用户自身计算负载过重以及协作用户不可信或协作程度不高等缺点。3层架构是在用户和LSP之间添加一个可信的第三方匿名服务器(AS, anonymous server)。该架构能够提供较好的隐私保护效果,同时降低用户的自身负载。本文所提出的隐私保护方法需要计算产生相似位置轨迹的后续匿名位置,若采用双层模式,则用户自身的计算负载过大,不宜在移动终端完成计算。因此,本文采用3层的位置隐私保护系统架构,该架构如图3所示。

从图3中可以看出,用户首先将位置查询发送给AS。经过AS中的位置计算模块生成可产生相似轨迹的匿名位置,再由位置匿名模块完成匿名后发送给LSP。LSP使用查询模块进行查询后,同时将查询结果保存到历史数据中,最后将结果返回给AS。AS将获得的匿名查询结果加以过滤后发送给用户,从而完成整个查询过程。

3.2 基本概念

定义 1 3层系统架构的网络实体可用三元组(User, AS, LSP)表示。其中,User代表当前网络中存在的所有移动用户,这些用户可通过智能移动设备获取位置服务,但其通信和计算能力有限;AS代表3层结构中的匿名服务器组,用来计算能够生成相似轨迹的查询位置,以此完成位置匿名,AS具有较强的计算和数据处理能力;LSP为不同的位置服务提供商,能够完成User或AS提出的基于位置服务,同时能够保存位置服务

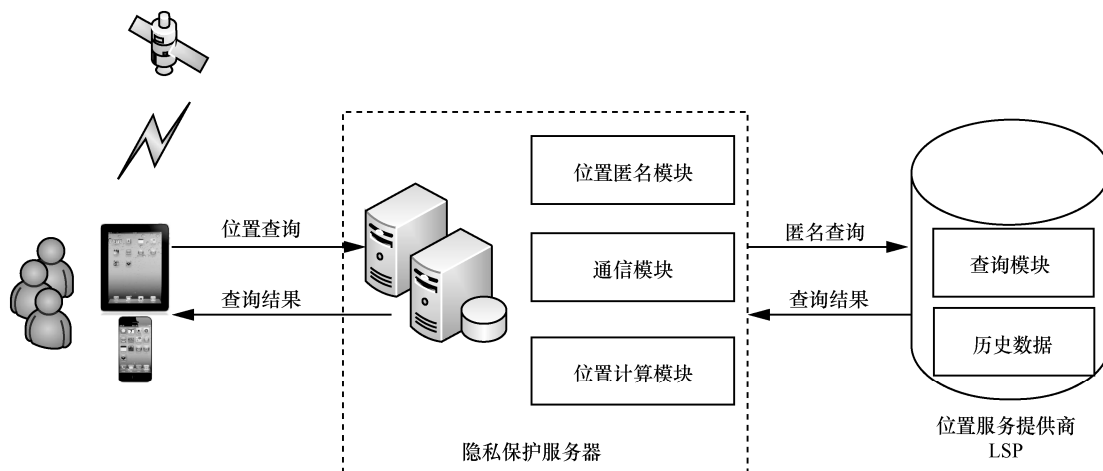


图3 3层系统架构示意

的历史数据,其数据分析和处理的能力极强。

在本文中,假设 AS 是可信的,其部署数量能够满足当前所有 User 的隐私保护需求;而 LSP 是半可信的,在其服务的过程中,对用户的位置隐私感兴趣,但能够按照协议完成服务,即 LSP 是好奇而非恶意的。

定义 2 连续查询服务是指 User 向 LSP 发起多次请求,该请求可表示为 $Q=(ID, L_i, t, P)$ 。其中, ID 表示用户的身份标识或假名; $L_i=(x_i, y_i)$ 表示当前查询所处的位置; t 表示时间间隔; P 表示查询的兴趣点,如餐厅、加油站等。

本文中 User 首先将 Q 发送给 AS,由 AS 完成匿名位置计算后,生成满足轨迹匿名的虚假用户,将 User 与虚假用户建立匿名组,AS 将匿名组作为请求用户发送给 LSP。LSP 在完成查询后,将查询结果返还给 AS。AS 在过滤掉虚假查询结果后,将真实的查询结果返还给 User,完成当前查询。

定义 3 位置轨迹匿名是指 LSP 同时获得包括真实用户轨迹 T 在内的至少 r 条轨迹,这些轨迹使 LSP 无法通过轨迹长度和轨迹方向等相似性准确辨识出 T 的隐私保护过程。

本文中的轨迹由于是在连续查询过程中产生的,因此与以往的轨迹匿名不同,需要对尚未完全生成的轨迹完成匿名操作,防止攻击者实时获取真实用户的位置隐私。

3.3 攻击手段

针对连续查询的位置轨迹隐私,攻击者主要采用假名关联、内容关联、时间间隔关联和移动类型关联等攻击手段。假名关联就是通过用户在连续查询过程中每个位置提交的假名信息,按照相同假名的连续关系,获取连续查询的位置,进而关联分析出用户的隐私信息;内容关联是根据用户所提交的查询内容,针对每个查询位置的相同查询内容关联具体的用户;时间间隔关联是针对连续查询过程中,User 所选择的查询时间间隔差异(如每隔 2 min 返回结果或每隔 4 min 返回查询结果),获取用户的位置隐私;移动类型关联(movement types correlation)是指攻击者通过观测等手段获取待攻击用户的运动速度和运动方向,以此进行用户比对,刻画出用户的行进轨迹,

推测出用户在后续查询中所处的位置,进而达到对用户位置隐私攻击的目的。下面以假名关联攻击为例,通过如图 4 所示的二分图来刻画这个攻击过程。

用户的假名经过泛化处理存在如图 4(a)所示的假名与位置的二分图刻画,任意一个假名都存在多个位置与之匹配,使攻击者无法获得真实用户假名与位置之间的关联。由于在连续查询的过程,同一假名连续出现在可相互连接的位置上,由此攻击者可以将假名与位置唯一匹配,获得如图 4(b)所示的真实匹配二分图刻画。其他 3 种关联攻击与之相似。

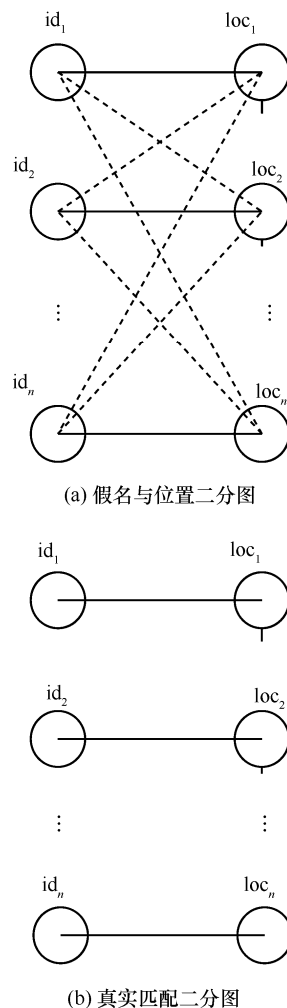


图4 假名关联攻击的二分图刻画

若要在这些关联攻击下保护用户的位置隐私,当前主要的方法是模糊这些关联关系。在二分图中可表示为扰乱位置与假名、内容、时间间隔和移动类型之间的匹配关系。在实时的位置服

务中,针对假名关联,可以在每次查询过程中进行假名的随机更换;针对内容关联,可以按照 I -多样性的标准进行查询兴趣点泛化;针对查询时间间隔关联,可以通过选择具有相同时间间隔的其他用户,或选择当前正在进行查询的用户进行泛化;而针对移动类型的关联攻击,则需要使用轨迹匿名的隐私保护技术。

已有的轨迹隐私保护方法主要基于历史轨迹数据^[15]或当前其他用户产生的轨迹^[16],将这些与基轨迹混合完成匿名,但这些方法均存在一定的不足。前者需要 AS 具有较好的存储能力,能保存大量的历史轨迹数据;后者需要 AS 具有高速的筛选计算能力,能够快速地在大量提交数据或候选数据中,查找相似的轨迹数据。因此,这些方法都不适合在连续查询中进行位置轨迹匿名。基于这个问题,本文设计了一种实时相似轨迹生成方法,通过这种方法计算并找到能够生成相似轨迹的查询位置,在 AS 的协作下构建匿名组并实时完成轨迹匿名,保护连续查询下的用户位置隐私。

4 相似轨迹实时生成方法

为实现连续查询过程中的实时轨迹相似,需要在 3 个不同阶段对真实用户的查询信息加以处理。这 3 个阶段分别为:初始位置匿名阶段、相似轨迹位置生成阶段和生成位置筛选阶段。初始位置匿名主要通过 AS 生成满足一定条件初始虚假位置,并与基用户首次查询的真实位置建立有效的匿名组,在满足 k -匿名的前提下发送给 LSP。相似轨迹位置生成阶段则要求 AS 根据初始匿名位置或连续查询过程中的前次匿名位置,通过与基位置完成轨迹相似计算后,获取满足 r -匿名的后续位置,并生成位于该位置的虚假用户,建立覆盖基用户当前查询位置的匿名组。在生成位置

筛选阶段,AS 需要按照基用户制定的轨迹相似度和轨迹夹角最大阈值,通过相似轨迹位置计算,得出 4 个不同的位置坐标,将这些坐标相互连接构成四边形区域,在该区域内根据实际的地理信息情况,剔除掉真实用户不可到达的位置,并选择区域中剩余的任意位置生成虚假用户,进而建立当前查询的匿名组。

4.1 初始位置匿名阶段

满足轨迹匿名最好的结果,是找到连续查询位置能够生成长度相等、方向相同的位置轨迹。然而,由于实际路段等具体地理情况的影响,在生成的轨迹中可能会存在轨迹之间间隔较小的情况,本文称这种轨迹为伴随轨迹。伴随轨迹可认为是相似轨迹中的一种特例,这种轨迹在发布的隐私保护中能达到较好的轨迹匿名效果,但在位置服务尤其是连续的位置查询服务下,这种伴随轨迹可能存在共同的查询位置和兴趣点,使攻击者能够根据任意其他轨迹获得基用户所在的最小区域,进而获取其位置轨迹隐私。如图 5 所示,3 条匿名轨迹之间相距较近,虚线构成的方框代表相同查询位置的最小区域。在连续匿名的过程中,3 条轨迹均经过相近或相同的查询位置,可以按照同一轨迹看待。此时若攻击者获知任一用户的任一查询位置,则可获知基用户的位置隐私。在本文所提方法中,后续阶段使用轨迹方向最小阈值的限定,要消除这种伴随轨迹带来的隐患,需在初始位置匿名阶段,扩大初始匿名虚假用户与基用户之间的间隔距离。因此,在初始位置匿名阶段,需加入对用户位置之间的距离限定,以防止后续匿名操作完成后产生轨迹伴随现象,影响用户的位置隐私。

4.2 相似轨迹位置生成阶段

相似轨迹位置是 AS 通过前次匿名中所有虚假用户位置和当前用户的基位置,按照轨迹相似

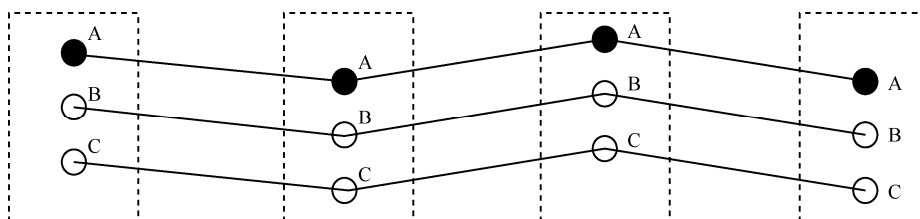


图5 伴随轨迹示意

的思想计算而来。在轨迹相似的思想中,认为 2 条轨迹之间的距离越短,则轨迹越相似,这里借用 Lee^[8]关于相似轨迹区段计算的基本思想,以轨迹距离来量化两轨迹之间的相似程度。其中,使用平行距离量化轨迹长度之间的差异、夹角距离量化轨迹方向之间的差异这两者之和来量化轨迹相似程度。

假设用户 A 和虚假用户 B 分别以 p_a 和 p_b 为初始查询位置建立匿名组,若要找到 B 的后续查询位置 p_{b+1} ,且连接用户 A、B 的相邻查询位置能够产生相似轨迹,需要通过 p_a 、 p_b 和 A 的后续查询位置 p_{a+1} 使用平行距离和夹角距离建立方程,进而解出 p_{b+1} 的位置坐标。令 L_a 和 L_b 分别以 p_a 、 p_{a+1} 和 p_b 、 p_{b+1} 为两端位置的轨迹序列, θ 表示用户制定的 2 条轨迹之间存在的夹角阈值。由于 p_{a+1} 已知,通过式(3)、式(4)联立建立关于 p_{b+1} 的二元二次方程,从而获取 p_{b+1} 的位置坐标,并以该坐标生成虚假用户 B 的后续查询位置,建立连续查询过程中的后续匿名组。其中,2 条轨迹之间存在的相似程度可表示为

$$\text{Sim}(A,B) = \frac{1}{1 + \text{Dist}(L_a, L_b)} \quad (1)$$

其中, $0 < \text{Sim}(A,B) \leq 1$, 2 个轨迹区段之间的轨迹距离可表示为

$$\text{Dist}(L_a, L_b) = \omega_{\parallel} d_{\parallel} + \omega_{\theta} d_{\theta} \quad (2)$$

获取相似轨迹后续位置计算所需要的平行距离和夹角距离公式可分别表示为

$$d_{\parallel} = \|p_{a+1} - p_a\| - \|p_{b+1} - p_b\| \quad (3)$$

$$d_{\theta} = \sin \left(\arccos \frac{\text{dot} \vec{A}, \vec{B}}{\|p_{a+1} - p_a\| - \|p_{b+1} - p_b\|} \right) \quad (4)$$

式(4)中的 dot 表示 L_a 和 L_b 的长度之积,轨迹距离计算中的 $\omega_{\parallel}$ 、 ω_{θ} 分别代表平行距离和夹角距离的权重, $\vec{A}$ 、 $\vec{B}$ 分别表示用户 A、B 在当前方向的移动距离,由于 2 个距离公式均对轨迹相似程度产生影响,因此其默认值均设定为 1。这种相似轨迹的计算方法按照平行距离和夹角距离的数值确定两轨迹之间的相似性,可以看出,在完全相似和轨迹平行的情况下,能够获取最好的轨迹相似程度,此时匿名程度最好。对于上文的

相似轨迹位置计算,可以用如下算法加以描述。

算法 1 相似轨迹位置计算

输入 用户当前查询位置 p_{i+1} 与前次查询位置 p_i , 以及生成虚假用户前次查询位置 p_j 。

输出 位置 p_{j+1} 集合, 该集合中位置满足轨迹匿名

程序 AS 收到用户提交的查询请求 Q ;

let threshold_Sim=1, threshold_θ=0; // 设定相似性和轨迹夹角阈值满足完全平行和长度相等

A=position of query user;

P_s =null; // 目标位置集为空

for every B ∈ S // 对于每个初始匿名组中虚假用户位置

{

Dist(A,B)=(1-threshold_Sim)/threshold_Sim;

dot($\vec{A}$, $\vec{B}$) = dot($[p_{i+1} - p_i]$, $[p_{j+1} - p_j]$);

$d_{\theta} = \text{threshold}_{\theta} = \sin \left(\arccos \frac{\text{dot} \vec{A}, \vec{B}}{\|p_{a+1} - p_a\| - \|p_{b+1} - p_b\|} \right)$;

// 关于位置 θ 的夹角距离方程

$d_{\parallel} = \text{Dist}(A,B) - \text{threshold}_{\theta} = \|p_{i+1} - p_i\| -$

$\|p_{j+1} - p_j\|$; // 关于位置 p_j 的平行距离方程

$p_{j+1} = \text{solve}(d_{\parallel}, d_{\theta})$; // 解关于 p_{j+1} 二元二次方程

if all p_{j+1} equals to each other

$P_s = \text{one of } p_{j+1}$; // 将结果存入目标位置集

else

check the threshold_Sim and threshold_θ
or break;

// 如果方程结果不同,则需检查输入的相似度和夹角阈值。

end if

}

算法 1 对每一个参与前次匿名的位置进行相似轨迹位置计算, 所得结果保存在 P_s 中, 连接 2 次邻近查询中的匿名位置, 可获得与真实用户 2 次查询形成的基轨迹相似的其他 $r-1$ 条轨迹。由于后续轨迹的生成数量就是首次匿名过程中的匿名用户数量, 此时轨迹数 r 与位置数 k 相等。此种情况下, 生成的虚假轨迹在方向、长度等方面与用户真实轨迹相似, 攻击者无法通过移动类型差异准确地识别出基轨迹, 从而保护用户的位置隐私。然而, 算法 1 得出的位置是一种理想位置, 在实际情况下可能会由于该位置

的地理特性，使所生成的虚假位置与前次匿名的虚假位置之间不可连接，进而攻击者通过位置的可达性剔除匿名轨迹。针对这一问题，本文根据真实用户制定的轨迹相似度和轨迹夹角最大阈值，计算获得满足最小轨迹相似度的位置区域，并在该区域内根据实际的地理信息情况，剔除用户不可到达的位置。

4.3 生成位置筛选阶段

为防止攻击者通过真实环境的不可到达性，剔除生成的虚假位置，需要对计算获取的轨迹相似位置进行筛选。相似轨迹位置区域是一个相似位置的集合，是在轨迹最大方向偏差 Max_θ 和最小相似度 $\text{min_Sim}(A,B)$ 规定范围内，通过轨迹相似位置计算生成获得的 4 个位置连接后建立的四边型区域。在这个区域内任一位置都满足用户规定的轨迹最小相似程度。由此，将前次查询的匿名用户位置与对应的该区域中任意位置相连接，生成的轨迹均满足最低相似度。在该区域中选择后续位置，需根据地图上的实际地理差别和相近查询位置的无障碍性予以选择。假设匿名用户 U，经过相似轨迹位置计算后获得如图 6 所示的由 A、B、C 这 3 个用户所在位置组成的相似位置区域。由于用户 U 与 A 所在位置之间存在建筑物阻挡，而用户 C 所在的位置位于河流当中，根据实际可到达情况，AS 只能选择 B 作为当前匿名用户的查询位置，建立匿名组。

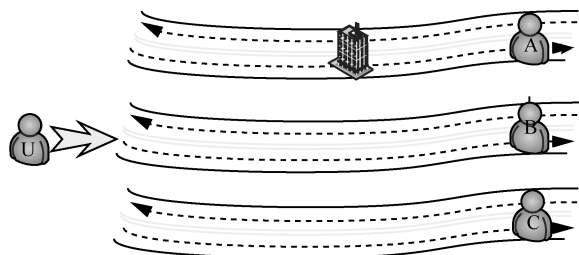


图6 生成位置筛选示意

在进行生成位置筛选前，需要首先计算出相似轨迹位置区域。与完全相似轨迹位置计算的方法类似，需要在最大方向偏差 Max_θ 和最小相似性 $\text{min_Sim}(A,B)$ 阈值的限定下，计算出该区域。对于轨迹相似性的判定，可以认为在可视范围下无明显差别即为相似。通过后续实验观测可以认定 $\text{min_Sim}(A,B)=0.9$ ，而 $\text{Max}_\theta=5$ 是满足轨迹相似情况下，可设定的最小相似程度和最大方向偏差。在算法 2 中将这 2 个阈值代入到位置计算中，

完成对相似轨迹区域的计算，同时对计算获得的位置进行筛选，进而获得相似轨迹位置。

算法 2 有筛选功能的相似轨迹位置计算

输入 用户当前查询位置 p_{i+1} 与前次查询位置 p_i 以及其他虚假用户前次查询位置 p_j 。

输出 位置 p_{j+1} 集合，该集合中位置满足轨迹匿名

程序 AS 收到用户提交的查询请求 Q ;

let $\text{threshold_Sim}=\text{min_Sim}(A,B)$; $\text{threshold}_\theta=\text{Max}_\theta$;

//设定相似性和轨迹夹角阈值满足完全平行和长度相等

$A=\text{position of query user}$;

$P_s=\text{null}$; //目标位置集为空

for every $B \in S$ //对于每个初始匿名组中其他虚假用户位置

{

$L_c=\text{null}$; //相似轨迹区域初始为空

$\text{Dist}(A,B)=(1-\text{threshold_Sim})/\text{threshold_Sim}$;

$\text{dot}(\vec{A}, \vec{B}) = \text{dot}([p_{i+1} - p_i], [p_{j+1} - p_j])$;

$d_\theta = \text{threshold}_\theta = \sin \left(\arccos \frac{\text{dot} \vec{A}, \vec{B}}{\|p_{i+1} - p_i\| - \|p_{j+1} - p_j\|} \right)$;

//关于位置 p_j 的夹角距离方程

$d_\parallel = \text{Dist}(A,B) - \text{threshold}_\theta = \|p_{i+1} - p_i\| - \|p_{j+1} - p_j\|$; //关于位置 p_j 的平行距离方程

$p_{j+1}=\text{solve}(d_\parallel, d_\theta)$; //解关于 p_{j+1} 二元二次方程

if all p_{j+1} equals to each other

execute algorithm 1;

else

draw quadrangle with p_{j+1} ; //建立位置相似

区域

end if

if the region reachable from p_j //判断位置可到达情况

random choose one point as p_c ;

$P_s=p_c$; //将结果存入目标位置集

else

check another region;

end if

if there is no reachable region in the qua-

drangle

break;

//如果当前计算出的可匿名的相似轨迹区域均不可到达, 则计算失败

```
end if
}
```

在算法 2 中对相似位置的求解方式与算法 1 相同, 但由于相似程度和夹角距离限制, 产生了二元二次方程的 4 个不同根, 将这些根表示的位置节点连接即可获得一个可产生相似轨迹的位置区域。在这个区域中的位置节点都可与前次查询的位置节点连接建立相似轨迹, 但该区域中的部分位置可能是不可到达的, 筛选掉这些不可到达的位置后, 在该区域中任选一位置, 即可作为当前匿名位置组建新的匿名组。若所有位置均不可到达, 则匿名失败。

5 性能分析

本节主要针对所提出的相似轨迹实时生成方法, 进行安全性和执行效率分析。

5.1 安全性分析

轨迹匿名方法的安全性取决于用户设定的匿名阈值, 本文将其与 Hwang^[15]提出的 r -匿名观点进行对比分析。 r -匿名是指在位置服务的隐私保护数据中, 其产生的轨迹数据至少存在 $r-1$ 条相似的历史轨迹, 使攻击者无法识别出基用户的真实轨迹。本文中的方法是在连续查询的过程中, 通过对后续位置的计算, 判定能够与真实用户产生相似轨迹的匿名位置, 通过在该位置生成虚假用户建立匿名组的方式, 实现实时的轨迹匿名。这种实时匿名与以往的轨迹匿名方法相比, 在 AS 提交位置给 LSP 之前实现轨迹匿名, 使攻击者无法通过实时轨迹差异获取用户隐私。若用户在查询过程中设定轨迹匿名阈值为 r , 并通过该方法生成相似轨迹, 则攻击者在已知真实用户移动类型的基础上, 通过轨迹方向和长度的差异, 正确识别出基用户轨迹的概率为 $\frac{1}{r}$ 。

另外, 在初始位置匿名阶段, 通过设定用户间间隔距离, 可降低生成的匿名相似轨迹出现伴随轨迹的情况, 阻止因生成轨迹存在伴随情况导致隐私泄露。同时, 根据位置轨迹匿名的实时性, 需要考虑生成位置是否可达, 本文对生成的相似轨迹位置进行了筛选, 通过建立满足最小相似度和最大轨迹夹角要求的位置区域, 在满足相似性

的前提下, 有效地阻止了攻击者利用用户不可到达性, 识别潜在的真实位置攻击行为。

在相似轨迹实时生成方法的运行过程中, 由于使用生成虚假用户的方式建立匿名组, 其用户假名、查询内容和查询时间间隔均与基用户一致。所以, 在满足轨迹匿名、有效抵抗移动类型关联攻击的基础上, 对假名关联、内容关联和时间间隔关联攻击也存在较好的抵抗能力。可以认为, 该方法能够很好地解决以上 4 种关联攻击对位置隐私的威胁, 保护用户在连续查询过程中的位置隐私。

5.2 效率分析

由于相似轨迹实时生成方法主要应用在 AS 端, 查询过程中基用户只需提交自身的查询, 生成虚假查询和基用户信息是在 AS 与 LSP 之间传递, 无需基用户进行二次通信, 因此基用户所产生的通信量与单纯查询所产生的通信量相同。在 AS 与 LSP 之间, 可以认为存在高速信息通道, 满足任何数量的通信需求。在整个连续查询过程中所有对轨迹匿名的计算都是通过 AS 来完成的, 大量的计算过程可能会导致 AS 出现瓶颈效应。因此, 这里通过对相似轨迹位置计算的执行效率加以分析, 以验证执行效率, 其运行结果在后面的实验中进行详细探讨。

在考虑位置可达性的情况下, 为实现位置轨迹匿名, 需要执行算法 2 来寻找匿名位置, 从而建立匿名组。按照安全性分析中采用的 r -匿名要求, 需要每个连续查询间隔中生成 r 条相似轨迹, 以满足轨迹 r -匿名。假设在特定的连续查询区域, 每次匿名过程均生成可到达的虚假用户, 即在最好的情况下, 算法 2 针对匿名阈值为 r 的匿名组进行相似轨迹位置计算, 其时间复杂度为 $O(r-1)$ 。针对获取的相似轨迹进行生成位置筛选, 其执行的时间复杂度同样为 $O(r-1)$, 由此可得出进行轨迹匿名的虚假位置并建立匿名组整个过程的时间复杂度为 $O(\text{sum})=O(r-1)+O(r-1)=2O(r-1)=O(r-1)$ 。

匿名过程中的最坏情况是所有计算获得的位置均不可到达, 此时算法 2 需计算相似轨迹位置区域, 并在相似轨迹位置区域中寻找匿名位置并建立匿名组。算法 2 中相似轨迹位置计算的时间复杂度与最好条件下同为 $O(r-1)$, 而对参与匿名虚假用户的查找需要在剔除不可到达位置之后随

机选取,此时的执行时间复杂度为 $2O(r-1)$ 。若在计算后的相似轨迹位置区域并不真实可达,则可认为轨迹匿名失败,此时无需计算算法的时间复杂度。由此得出,在最坏的情况下算法2执行的时间复杂度为 $O(\text{sum})=O(r-1)+2O(r-1)=3O(r-1)=O(r-1)$ 。

综上,可以认为,算法2的时间复杂度取决于用户设定的轨迹匿名阈值和后续相似位置是否可达,由于本文所提出的相似轨迹实时生成方法是基于欧氏空间提出的,可以认为在多数情况下能够找到可达位置上的匿名用户。因此,在一般情况下,算法2的时间复杂度位于最好情况与最坏情况之间,即算法2的时间复杂度为 $O(r-1)$ 。

6 实验验证

本节主要围绕相似轨迹实时生成方法产生的轨迹相似程度,以及相似轨迹位置计算方法的执行效率2个方面,通过模拟数据集加以实验分析,以验证该方法的可行性。

6.1 环境配置

本实验中所涉及的各种匿名方法是在 Windows 7 系统上使用 Matlab 7 来进行模拟的,运行环境为 1.70 GHz Intel Core i5,内存大小为 4 GB。实验数据集采用 Thomas Brinkhoff 路网数据生成器生成的连续位置数据。表1为实验中采用的相关参数阈值设定。

表1 实验参数阈值设定

参数名	取值范围	默认值
轨迹相似性 Sim	$0 < Sim \leq 1$	1
轨迹方向夹角 θ	$0 \leq \theta \leq 5$	0
距离权重 ω_1, ω_0	$0 \leq \omega_1, \omega_0 \leq 1$	1
连续查询次数 n	$2 \leq n \leq 10$	5
匿名阈值 r	$1 < r \leq 100$	10
区域内用户总数 U	$1 \leq U \leq 20$	6 000
初始匿名距离 L	$1 \leq L \leq 4$	1

6.2 生成连续位置轨迹相似性对比

这里通过实验获取在连续5次查询过程中,生成的位置轨迹与基轨迹之间的相似程度对比,来查看相似度阈值与夹角阈值对轨迹匿名所带来的影响。实验设定初始参与匿名用户距离为1,用以避

免在相似轨迹计算中形成伴随轨迹。下面从相似度阈值和夹角阈值变换方面,查看虚假位置参与匿名后所形成轨迹与基轨迹之间的相似程度。

对于生成的相似轨迹,在5次连续查询、匿名度 $r=2$ 、生成4段相似轨迹的情况下,分别以 $\min_Sim(A,B)=1$, $\text{Max_}\theta=0$ 、 $\min_Sim(A,B)=0.9$, $\text{Max_}\theta=0$ 以及 $\min_Sim(A,B)=0.9$, $\text{Max_}\theta=5$ 这3种情况运行算法2,获得如图7~图9所示的生成相似轨迹,其中,初始位置距离设定为1。

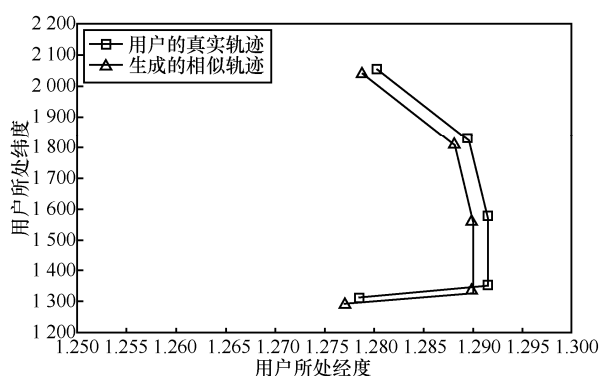


图7 相似性为1、夹角为0°时的轨迹对比

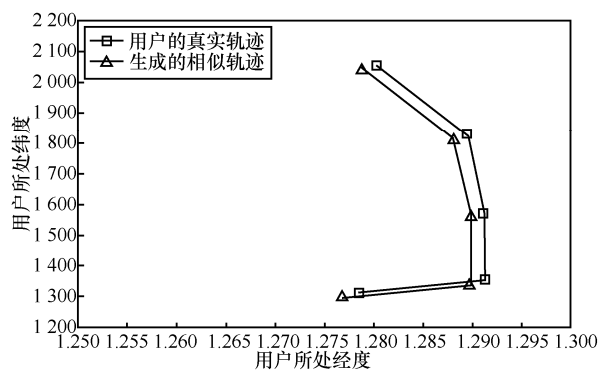


图8 相似性为0.9、夹角为0°时的轨迹对比

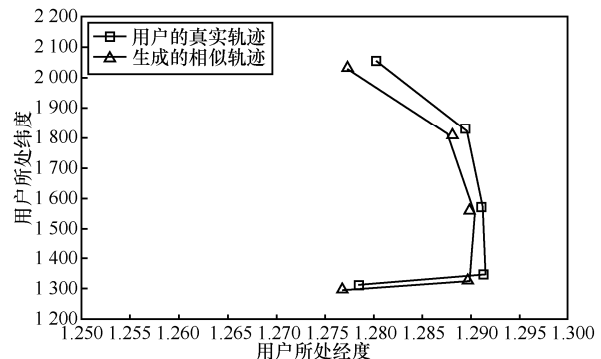


图9 相似性为0.9、夹角为5°时的轨迹对比

通过以上对比可以看出,单独在相似度要求降低的情况下,生成的相似轨迹会由于随机选择

的平行距离偏差,造成轨迹整体差异,但轨迹区段之间距离较远使这种差异并不明显。而在夹角度数变化的情况下,这种差异不仅体现在轨迹长度方面,更体现在方向偏移上。当查询次数增加较多时,连续的同向夹角偏差将会加大轨迹偏移,使最后产生的轨迹与真实轨迹之间偏差增大。因此,在连续查询次数较高的情况下,为保证生成轨迹的相似程度,需尽量选择较小的轨迹夹角阈值,以免生成轨迹与基轨迹之间差异过大。

6.3 执行效率对比

在执行效率的对比实验中,本文首先根据匿名参数的取值变化,查看在匿名参数逐渐增大的情况下,该方法与其他匿名方法之间在执行效率上的差异;然后,查看在连续多次执行过程中,在不同匿名值限定下的执行效率。为提高生成轨迹的真实程度,本文以 $\min_Sim(A,B)=0.9$, $\max_t=5$ 进行 2 次连续查询,计算随匿名值增加导致的时间变化情况。为检测该方法与其他同类方法在执行效率上的差异,本文加入 $C_{LAPPINGQ}$ ^[14]、EGCA^[19]和 LTPPM^[16]方法加以对比,其中,只有 LTPPM 与本文相似,主要针对移动类型关联攻击。

从图 10 中可以看出,相似轨迹实时生成方法由于需要计算产生轨迹的相似性,因此与直接选择用户进行位置匿名的 $C_{LAPPINGQ}$ 、EGCA 相比执行效率较差,而 $C_{LAPPINGQ}$ 由于采用预计算的方式,其位置选择效率更高,但这 2 种方法均无法抵抗移动类型关联攻击。针对移动类型关联攻击的保护方法有 LTPPM 和本文所提方法, LTPPM 通过用户协作建立相似轨迹,这种方法涉及产生相似轨迹的位置计算,同时需要对匿名轨迹进行筛选,因此其时间效

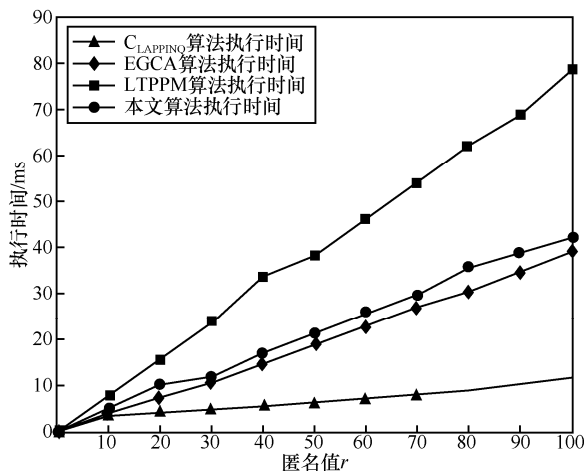


图 10 $U=60\,000$, $k(r)$ 值变化时各方法的执行效率

率随匿名值变化较大。本文所提方法由于采用计算后续位置的方法,无需进行大量的轨迹筛选,与 LTPPM 相比具有更好的执行效率。通过对比实验可以看出,本文所提出的方法在执行效率上介于直接位置选取和轨迹判断选取方法之间,在已有的基于实时轨迹匿名的隐私保护方法中效率较高。

同样在限定 $\min_Sim(A,B)=0.9$, $\max_t=5$ 的情况下,本文以 $r=10$ 进行连续重复的匿名操作,查看在连续运行过程中的执行效率变化,其中,使用的数据依旧采用 Thomas Brinkhoff 路网数据生成器生成。

从图 11 中可以看出,在匿名值固定的情况下,各方法进行连续隐私保护时,其执行时间随重复次数的增加而增大。本文所提方法在计算出匿名的位置之后,需进行相似轨迹位置筛选,其时间效率要低于直接选择用户进行匿名的方法,但高于查找相似轨迹的隐私保护方法。

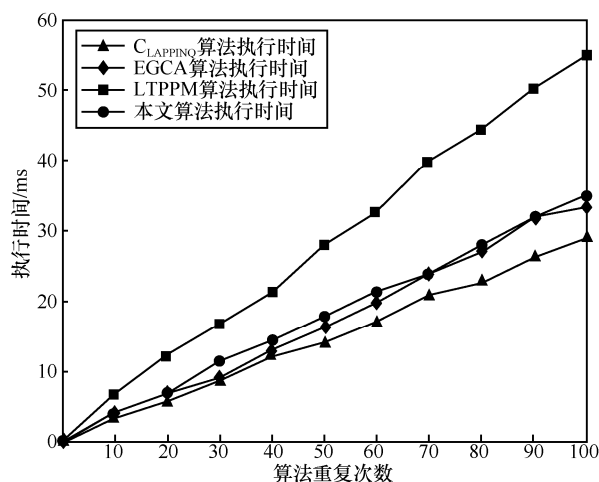


图 11 $r=10$ 时, 随重复计算次数变化的各方法的执行效率

从两组实验的对比中可以看出,本文所提出的相似轨迹实时生成方法尽管不如直接选择用户进行匿名的方法执行效率高,但是能够模糊因移动类型差异导致的轨迹差异,进而保护在连续查询过程中用户的位置隐私,并且与选择相似轨迹的方法相比,具有更好的执行效率,能够满足连续查询过程中的实时轨迹隐私保护。因此,本文所提方法具有较好的应用前景。

7 结束语

本文针对 LBS 连续查询过程中,攻击者可通过位置轨迹差异获取用户隐私的情况,从轨迹的

产生方式入手, 通过对轨迹方向、轨迹长度的量化, 引入轨迹相似性的计算思想, 提出了一种相似轨迹实时生成方法。在这种方法下建立通过相似轨迹计算, 获取后续位置的实时位置计算方法, 并通过相似轨迹位置筛选提高生成轨迹的相似程度。该方法能够有效地防止攻击者通过用户移动类型获取连续查询过程中的用户轨迹隐私。实验表明, 该方法在确保用户位置隐私的同时具有很好的执行效率。当然, 该方法也存在一定的不足, 由于相似轨迹位置是通过解二元二次方程的方式获得, 其执行效率受 AS 计算能力影响较大。同时, 该方法的提出主要基于欧氏空间, 不能有效地部署在实际路网中。因此, 在下一步的工作中, 将对以上 2 个方面的问题加以分析和研究, 尤其是加强在路网问题上的应用研究, 使该方法具备更好的实际部署价值。

参考文献:

- [1] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//The 1st International Conference on Mobile Systems, Applications and Services, ACM. 2003: 31-42.
- [2] BAMBA B, LIU L, PESTI P, et al. Supporting anonymous location queries in mobile environments with privacygrid[C]//The 17th International Conference on World Wide Web. 2008: 237-246.
- [3] REBOLLO-MONEDERO D, FORNE J, SOLANAS A, et al. Private location-based information retrieval through user collaboration[J]. Computer Communications, 2010, 33(6): 762-774.
- [4] YIU M L, JENSEN C S, MOLLER J, et al. Design and analysis of a ranking approach to private location-based services[J]. ACM Transactions on Database Systems, 2011, 36(2): 475-486.
- [5] 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. 2008: 121-132.
- [6] AGGARWAL G, FEDER T, KENTHAPADI K, et al. Anonymizing tables[M]. Berlin: Springer, 2005: 246-258.
- [7] WONG R C W, LI J, FU A W C, et al. (α , k)-anonymity: an enhanced k -anonymity model for privacy preserving data publishing[C]//The 12th ACM Sigkdd International Conference on Knowledge Discovery and Data Mining. 2006: 754-759.
- [8] LEE J G, HAN J, WHANG K Y. Trajectory clustering: a partition-and-group framework[C]//The 2007 ACM Sigmod International Conference on Management of Data. 2007: 593-604.
- [9] XU Y, WANG K, FU A W C, et al. Anonymizing transaction databases for publication[C]//The 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2008: 767-775.
- [10] ABUL O, BONCHI F, NANNI M, et al. Never walk alone: uncertainty for anonymity in moving objects databases[C]//The 24th IEEE International Conference on Data Engineering, Cancun. 2008: 376-385.
- [11] BONCHI F. Privacy preserving publication of moving object data[M]//Privacy in Location-Based Applications. Berlin: Springer, 2009: 190-215.
- [12] BONCHI F, LAKSHMANAN L V S, WANG H. Trajectory anonymity in publishing personal mobility data[J]. Sigkdd Explor News, 2011, 13(1): 30-42.
- [13] 潘晓, 郝兴, 孟小峰. 基于位置服务中的连续查询隐私保护研究[J]. 计算机研究与发展, 2010, 47(1): 121-129.
- [14] PAN X, HAO X, MENG X F. Privacy preserving towards continuous query in location-based services[J]. Journal of Computer Research and Development, 2010, 47(1): 121-129.
- [15] HASHEM T, KULIK L, ZHANG R. Countering overlapping rectangle privacy attack for moving KNN queries[J]. Information Systems, 2013, 38(3): 430-453.
- [16] HWANG R H, HSUEH Y L, CHUNG H W. A novel time-obfuscated algorithm for trajectory privacy protection[J]. IEEE Transactions on Services Computing, 2014, 7(2): 126-139.
- [17] GAO S, MA J F, SHI W S, et al. LTPPM: a location and trajectory privacy protection mechanism in participatory sensing[J]. Wireless Communications & Mobile Computing, 2015, 15(1): 155-169.
- [18] WANG Y, HE L P, PENG J, et al. Privacy preserving for continuous query in location based services[C]//The IEEE 18th International Conference on Parallel and Distributed System (ICPADS). 2012: 213-220.
- [19] SKOUMAS G, SKOUTAS D, VLACHAKI A. Efficient identification and approximation of k -nearest moving neighbors[C]//The 21st ACM Sigspatial International Conference on Advances in Geographic Information Systems. 2013: 264-273.
- [20] WANG E K, YE Y. A new privacy-preserving scheme for continuous query in location-based social networking services[J]. The International Journal of Distributed Sensor Networks, 2014(1): 836-839.
- [21] JIA J, ZHANG F. Nonexposure accurate location k -anonymity algorithm in LBS[J]. Scientific World Journal, 2014(1): 149-168.
- [22] 张磊, 马春光, 杨松涛. 基于位置关联相似性的匿名算法[J]. 中国科技论文, 2016(2): 197-201.
- [23] ZHANG L, MA C G, YANG S T. Anonymous algorithm based on position correlation similarity[J]. China Science and Technology, 2016 (2): 197-201.
- [24] BERESFORD A R, STAJANO F. Location privacy in pervasive computing[J]. Pervasive Computing, 2003, 2(1): 46-55.
- [25] FREUDIGER J, SHOKRI R, HUBAUX J P. On the optimal placement of mix zones[C]//The Privacy enhancing Technologies. 2009: 216-234.
- [26] FREUDIGER J, RAYA M, FÉLEGYHÁZI M, et al. Mix-zones for location privacy in vehicular networks[C]//The 1st International Workshop on Wireless Networking for Intelligent Transportation Systems (Win-ITS). 2007.

- [25] PALANISAMY B, LING L. MobiMix: protecting location privacy with mix-zones over road networks[C]//2011 IEEE 27th International Conference on Data Engineering (ICDE). 2011: 494-505.
- [26] PALANISAMY B, LIU L, LEE K, et al. Anonymizing continuous queries with delay-tolerant mix-zones over road networks[J]. Distributed and Parallel Databases, 2014, 32(1): 91-118.
- [27] 马春光, 张磊, 杨松涛. 位置轨迹隐私保护综述[J]. 信息网络安全, 2015(10): 24-31.
- MA C G, ZHANG L, YANG S T. Track location privacy protection review[J]. Information Network Security, 2015 (10): 24-31.
- [28] 张磊, 马春光, 杨松涛, 等. 基于轮廓泛化的位置隐私保护模型及方法[J]. 系统工程与电子技术, 2016,38(12): 2894-2900.
- MA C G, ZHANG L, YANG S T, et al. Location privacy protection model and method based on contour generalization[J]. Journal of Systems Engineering, and Electronics, 2016,38(12): 2894-2900.

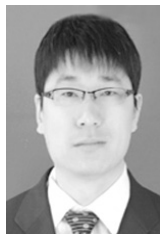


马春光 (1974-), 男, 黑龙江双城人, 博士, 哈尔滨工程大学教授、博士生导师, 主要研究方向为密码学、数据安全与隐私保护、无线自组织网络及安全。



杨松涛 (1972-), 男, 黑龙江鹤岗人, 博士, 佳木斯大学教授, 主要研究方向为信息安全、隐私保护。

作者简介:



张磊 (1982-), 男, 黑龙江绥化人, 哈尔滨工程大学博士生, 佳木斯大学讲师, 主要研究方向为信息安全、隐私保护。



李增鹏 (1989-), 男, 山东青岛人, 哈尔滨工程大学博士生, 主要研究方向为密码学、密码协议。