文章编号：2096-1472(2017)-12-12-03

基于*k*-匿名的轨迹数据隐私发布研究综述

赵凯毅，朱 麟，路士兵
(公安海警学院电子技术系，浙江 宁波 315801)

摘 要：随着移动定位技术的发展，大量移动轨迹数据使信息泄露于公开的互联空间中，使攻击者可以通过计算推理挖掘轨迹信息。轨迹数据发布的隐私保护是近年来网络空间安全领域研究的热点问题。为了防止该类轨迹数据隐私的泄露，通常采用*k*−匿名技术实现轨迹的隐私保护。该技术在国内外研究中取得了一定的成果。本文阐述了轨迹隐私保护的相关定义及研究方法，对国内外移动轨迹数据k−匿名隐私保护研究的成果进行了总结，并介绍了国内外有关轨迹数据*k*−匿名隐私保护研究的相关技术。同时对国内外的技术进行了比较，详细叙述了国外与国内各自方法的优点，指出了研究中存在的不足与今后研究的大致方向。

关键词：轨迹数据；轨迹数据发布；隐私保护；*k*−匿名
中图分类号：T391 **文献标识码**：A

A Research Review on Privacy-Preserving of Trajectory Data Publishing Based on *K*-Anonymity

ZHAO Kaiyi,ZHU Lin,LU Shibing
(*Department of Electronics,China Maritime Police Academy,Ningbo* 315801,*China*)

Abstract: With the development of mobile localization technology,large amounts of mobile trajectory data expose information to open cyberspace,so that hackers can dig out trajectory data by computing and reasoning.Recently,privacy-preserving of trajectory data publishing is one of the hottest topics in Internet security.In order to prevent the disclosure of trajectory data,*k*-anonymous privacy preserving model is widely applied,on which some research achievements have been made both at home and abroad.This paper presents some related definitions and research methods in trajectory data privacy preserving,summarizes domestic and abroad research achievements on the *k*-anonymous privacy preserving of mobile trajectory data,introduces and compares related technology on the *k*-anonymous privacy preserving of mobile trajectory data both at home and abroad,elaborates on respective advantages of each method at home and abroad,and points out the limitation in previous studies and the direction in future studies.

Keywords:trajectory data;trajectory data publishing;privacy preserving;*k*-anonymity

1 引言(Introduction)

随着移动设备和定位技术的发展，移动轨迹数据的隐私保护也受到了人们的关注。轨迹数据不仅蕴含着丰富的个人、位置等显性信息，而且还可以通过推理计算轨迹的隐性信息，挖掘移动终端设备的轨迹行为特征、行为模式和行为习惯，从而获取设备的信息数据，导致设备对象的隐私泄露。为有效保证数据的安全性，移动轨迹数据的隐私保护成为迫切需要解决的问题。

当前，普遍采用的轨迹隐私方法是基于*k*−匿名模型[1,2]，它是由Gruteser等人提出的。该技术是指在相关空间领域中，给定该空间节点位置形成的轨迹。当任意一条轨迹*T*在任意采样时刻t_i时，当至少有*k*−1条轨迹在相应的采样位置上与*T*泛化为同一区域时，则满足轨迹的*k*−匿名。*k*−匿名的技术的核心思想是将敏感属性泛化，使得单条记录无法和其他*k*−1条记录区分开，进而实现数据的隐私保护。

本文阐述了轨迹隐私保护的相关定义及的研究方法，对国内外移动轨迹数据*k*−匿名隐私保护研究的成果进行了总结，并介绍了国内外有关轨迹数据*k*−匿名隐私保护研究的相

基金项目：本文由公安部技术研究计划项目(2016JSYJC60)，“十三五”浙江省一流学科建设信息与通信工程资助.

关技术。同时对国内外的技术进行了比较，详细叙述了国外与国内各自方法的优点，指出了研究中存在的不足与今后研究的大致方向。

2 轨迹隐私保护方法(Privacy preserving methods)

轨迹数据发布中的隐私保护方法主要是假数据法、抑制法和泛化法。其中，假数据法是在原数据的基础上，采用与原始轨迹不同的信息来抵御攻击，同时保证数据统计属性的真实性。该方法计算量较小，容易产生数据存储量过大而导致数据的有效性降低，一般用于数据量不大的情况，可以体现方法的可用性；抑制法可以有条件地选择发布的数据，对敏感数据和频繁访问数据则不再选择发布，但是计算方法时间性能太大，会导致数据信息的损失量过大，一般也用于训练计算数据量较小的轨迹集合；泛化法是通过获取轨迹上的位置采样点，将这些采样点泛化为匿名区域，从而实现隐私保护。轨迹k−匿名技术就是一种基于泛化方法的隐私保护技术，可以平衡隐私保护度量和增加服务质量。

3 轨迹*k*-匿名隐私保护技术(*K*-Anonymous privacy preserving technology)

3.1 (*k*,*δ*)-匿名技术

轨迹(k,δ)−匿名的隐私保护技术最先是由国外研究者Abul等人[3]提出的。经过大量研究与分析，移动轨迹数据抽样和定位系统存在不精确性，而Abul与其他几位研究者正是利用这一不确定性，提出了轨迹(k,δ)−匿名的隐私保护技术。该技术在基于簇的方法基础上，实现了基于簇的轨迹(k,δ)−匿名。

所谓(k,δ)−匿名模型，是指在不确定性阈值δ和匿名阈值k给出的情况下，当且仅当$|D|\geqslant k$时，轨迹集合D满足(k,δ)−匿名。同时D中任意两条轨迹tr1和tr2均满足Coloc(tr1,tr2)。如果对于每一条轨迹，在该轨迹距离δ的半径区域内，应当至少存在与当前轨迹相似度较高的k−1条轨迹，δ越大，则说明通过聚类的组就越大，信息发布的安全性就越高，但信息损失率越高。

为了在提高信息隐私保护安全性的同时降低信息的损失率，NWA(Never Walk Alone)方法[3]将分布于同一时间段内的轨迹存储到一个等价类中，通过聚类寻找空间距离相似的轨迹，并满足k−匿名模型。该方法的优点可以采用欧式距离计算轨迹之间的距离。但是当数据量较多时，聚类使得当前轨迹的离群点增加，数据有效性受到影响。因此，在该问题的基础上，W4M[4]在NWA的基础上做了优化，采用编辑距离[5]来度量轨迹上采样点间的距离，使轨迹上离群点的数量较之前的方法相比有了明显的减少，降低了所舍弃数据的数量，更好且更有效地预防了攻击者使用特定的位置信息对发布的轨迹数据隐私进行攻击，同时保证了所发布的轨迹数据具有较高的质量。

上述的几种方法都将整条轨迹上的所有的采样点进行了泛化的处理，但忽视了准标识符QI的属性与其他属性间所存在的区别。因此，Yarovoy等人[6]针对此问题，对动态QI的属性进行了处理，查找所有时刻内移动对象的聚集距离最小的k−1个对象，并将该k−1个对象匿名到同一匿名的区域内，更加全面地做到了在泛化处理整条轨迹上所有采样点的同时，兼顾准标识符属性与其他轨迹间属性所存在的区别，提高了数据发布的服务质量。

3.2 (*k*,*l*)-匿名技术

轨迹(k,l)−匿名技术[7]与上述国外轨迹(k,δ)−匿名的隐私保护技术有所不同，该技术由RASUAR等人提出，是指在满足轨迹的k−匿名的基础上，同时满足轨迹l多样性的轨迹的集合。从构建轨迹k−匿名集的原则中我们了解到，在构建轨迹k−匿名集的同时，要使集合内的k条轨迹具有相似性，且转化后的数据库与原始数据库的信息扭曲度要尽可能小，这便是所说的NP难的问题。可是，假如轨迹在集合内的相似性过高，同样会导致隐私的泄露，因为在集合中轨迹都是相邻且靠近的，如果每条轨迹之间的相似性过高，所有移动对象的运动方式与路径便会很容易被攻击者分析了解到，从而降低了隐私的保护程度。所以，在构造轨迹k−匿名集时，要尽量避免出现相似的轨迹形状。

为了避免这一问题，文献[8]提出了一种面向个体的、个性化扩展的l−多样性隐私匿名模型。该模型虽然对传统的匿名模型进行了优化，但却只考虑到了轨迹的时间与空间属性。基于轨迹形状多样性的轨迹(k,l)−匿名技术，在考虑了轨迹的时间与空间属性的同时，还考虑到了轨迹的形状属性，使匿名集内的轨迹在保持相近距离与相似关联的基础上，依然保持了一定的形状多样性，避免了因高度的相似性而导致隐私泄露的可能。

轨迹(k,l)−匿名技术中所运用的SP算法是以文献[3]中的NWA算法的框架为基础所提出的，分为轨迹数据预处理、数据聚类、数据分组和数据(k,l)−匿名处理。数据预处理时，对所有的轨迹进行遍历，分配并选择时间段相等的轨迹，放置于同一个等价类中。如果某个等价类中的轨迹不足k条，为使损失的信息降低至最小，则通过贪心算法查找另一个等价类。通过重复同步化处理，将两个等价类合并为一个新的等价类，以此类推，最后获得一系列轨迹数不小于k的等价类。

轨迹(k,l)−匿名技术中的SP算法与NWA类似，但相比之

下，SP算法将轨迹间的多样性考虑了进去，即在相似复杂度得以保证的前提下，更有效地实现了隐私的保护。并且该算法使大量高度相似的轨迹存在于同一集合中的情况得到了避免，很好地实现了轨迹的(k,l)匿名，提升了数据的可用性。虽然该技术考虑到了轨迹的多样性，却未详细考虑轨迹集合经过较大敏感区的情况，故也存在一定不足。

3.3 (k,δ,Δ)-匿名技术

通过以上几种基于k匿名隐私保护技术的介绍，我们大致可以了解到：传统的关于轨迹数据发布的隐私保护的研究多数都将聚类的方法运用在其中。在很多情况下，其相关算法都只注重对于每条轨迹的隐私保护，而忽略了对轨迹聚类组特征的保护。通过相关理论研究与实验的论证，一些学者发现，在用聚类技术产生相应的轨迹数据后，对该轨迹数据进行二次聚类，可得到一系列特征，而该特征与在发布之前的原始轨迹数据所拥有的聚类组特征高度相似，从而可能导致隐私泄露。因此，为了有效地预防这种轨迹聚类的二次攻击，福州大学吴英杰等人[9]提出了一种(k,δ,Δ)-匿名模型和基于该模型的聚类杂交隐私保护轨迹数据发布算法CH-TDP。它是在(k,δ)-匿名模型的基础上，以(k,δ)-匿名模型为切入点，对聚类分组进行杂交，再进行组内扰乱。通过控制组间扰乱的程度，达到保护聚类组的共同特征。(k,δ,Δ)-匿名模型的目的是在抵御发布轨迹数据遭受二次聚类攻击的前提下，保证发布轨迹数据的质量不低于质量阈值Δ。

通过了大量的仿真实验和数据分析验证了轨迹(k,δ,Δ)-匿名的隐私保护技术中(k,δ,Δ)-匿名模型及CH-TDP算法的有效性。实验的相关结果证明，CH-TDP算法与(k,δ,Δ)-匿名模型不仅可以有效地抵御轨迹聚类过程中所产生的二次聚类攻击，同时，该匿名技术还很好地保证了匿名数据的质量，确保了轨迹的相似度，控制了区域查询结果的误差率，在k匿名隐私保护研究方面起到了至关重要的作用，更全面地实现了移动轨迹数据的隐私保护这一根本目的。

4 基于k-匿名技术的轨迹隐私保护技术的比较与分析(Comparison and analysis of trajectory data privacy preserving technology based on k-anonymity)

分析了三种较为典型的基于k-匿名的轨迹隐私保护技术，分别详细叙述了各自技术的主要原理、所用方法、重要用途，以及各自存在的优点及其所存在的一定局限性与不足。本文以表格的形式进行比较。

表1 基于k-匿名轨迹隐私保护技术比较

Tab.1 Comparison of privacy preserving techniques based on k-anonymity

技术名称	主要算法	用途	优点	局限与不足
(k,δ)-匿名	NWA[3]	寻找分布于相同时间段的轨迹，通过聚类计算空间距离相似的集合	能够结合k-匿名模型与轨迹聚类方法，较好地达到隐私保护效果	可能会产生大量离群点，导致无效数据增多
	W4M[4]	通过采样点的距离实现匿名集	在NWA技术上改进，减少了轨迹上离群点的数量，降低了数据舍弃的可能性，同时保证了所发布的轨迹数据具有较高的质量	没有考虑准标识符QI属性与其他属性的区别
	Min-MOB[6]	查找在所有时刻内，移动对象的聚集距离最小的k-1个对象，从而进行匿名	很好地考虑到了准标识符QI的属性与其他属性间所存在的区别，并且针对动态QI的属性进行了一定的处理，进一步提高了所发布数据的质量与完整性	解决了QI问题，但忽略了二次聚类所产生的聚类攻击
(k,l)-匿名	SP	以NWA为框架	以NWA为框架，保证匿名集内轨迹相似性外，同时保持轨迹形状多样性	仅考虑轨迹形状间的多样性，忽略了轨迹集合经过较大敏感区域的情况
(k,δ,Δ)-匿名	CH-TDP	用于抵御聚类过程中因二次聚类而产生的二次攻击	有效地抵御了轨迹聚类过程中所产生的二次聚类攻击，降低了二次聚类攻击的风险，还可以更好地保证匿名数据的质量和相似度，控制区域查询结果的误差率，更全面地做到了对于轨迹数据发布的隐私保护	未考虑聚类过程中轨迹间可能存在的高度相似性而导致的隐私泄露的可能性

5 结论(Conclusion)

在当前数据隐私保护研究领域，对于轨迹数据发布的隐私保护研究很广泛。本文中所提到的几种轨迹k-匿名技术都是比较成功的应用案例。轨迹(k,δ)-技术模型、轨迹(k,l)-技术模型、轨迹(k,δ,Δ)-匿名技术都是在k-匿名模型的基础上所提出来的。这些模型都有各自新颖之处，但其算法的精确性还有待改进。因此，虽然移动轨迹数据隐私保护是一个空间安全领域的研究热点，但其技术还不够成熟。目前，关于轨迹数据发布的隐私保护对模型和算法的研究相对较多，且有不少的学者进行了研究并提出了新的技术与算法。未来若能成功且全面地将轨迹数据隐私保护的研究成果进行实际应用，将会更好地促进信息共享和融合。

参考文献(References)

[1] Marco Gruteser,Dirk Grunwald.Anonymous Usage of Location-Based Services through Spatial and Temporal Cloaking[C].Proceedings of the 1st International Conference on Mobile Systems,Applications and Services,San Francisco,2003:31-42.

[2] 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.

[3] O.Abul,F.Bonchi and M.Nanni.Never walk alone:Uncertainty for anonymity in moving objects databases[C].Proceedings of the IEEE 24th International Conference on Data Engineering.IEEE,2008:376-385.

[4] O.Abul,F.Bonchi,M.Nanni.Anonymization of Moving Objects Databases by Clustering and Perturbation[J].Information Systems,2010,35(8):884-910.

[5] CHEN L,OZSU M T,ORIA V.Robust and fast similarity search for moving object trajectories[C].Proceedings of the 2005 ACM SIGMOD international conference on Management of data,2005:491-502.

[6] Yarovoy,R.,Bonchi,F.,Lakshmanan,S..Anonymizing Moving Objects:How to Hide a MOB in a Crowd?[C].In:12th International Conference on Extending Database Technology,2009:72-83.

[7] Trujillo-R,Domingo-Ferrer J.On the privacy offered by(k,δ)anonymity[J].Information Systems,2014,38(4):491-494.

[8] 孙丹丹,罗永龙,范国婷,等.基于轨迹形状多样性的隐私保护算法[J].计算机应用,2016,36(6):1544-1551.

[9] 吴英杰,唐庆明,倪巍伟,等.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.

作者简介：

赵凯毅(1997–)，男，本科生.研究领域：信息安全，数据挖掘.

朱　麟(1984–)，男，硕士，讲师.研究领域：智能计算，数据挖掘，信息安全.

路士兵(1978–)，男，硕士，副教授.研究领域：信息安全，嵌入式系统.

(上接第28页)

该功能基于Freemaker模板引擎，将编写好的实践报告存储为.flt,通过setClassForTemplateLoading方法加载文档模板，然后将存储的数据装载到定义出的map中，通过Freemaker技术编写工具类对文件里的标记位置如${velue}进行替换，将获取来的数据存储在定义好的键值对map中。最终将修改后的.flt文件以.doc文档格式生成并下载到根目录。

5　结论(Conclusion)

系统针对高校实践课程的特点和管理中的实际需要而设计，能够有效地实现课程管理的信息化，减轻老师的工作负担，高效率、规范化地管理大量的学生程序设计课程信息，并避免人为操作的错误和不规范行为。系统使用目前较成熟的开发工具及框架，使其对后期维护及其继续开发都提供了便利的条件。利用Ajax动态加载菜单是一种新的实现用户权限登录的方式，其能有效地提高程序运行的效率，减少了响应时间。在实践报告模板方面，利用Freemaker直接对模板进行修改，避免了传统技术上利用各种插件进行Word文件生成所带来的使用上的不便，从实用性和功能的完善上都有了不小的提升。该系统实现的自动生成实践报告模板功能是现存一些系统所没有的。该功能的实现使学生能够更好地规范实践报告的格式要求，能够使老师更加便利地管理学生所提交的实践报告。

参考文献(References)

[1] 杜文艺.基于Struts和Hibernate的web应用设计与实现中[M].北京:电子工业出版社,2014:4-10.

[2] Nicholas S.Williams.Professional Java for Web Applications:Featuring Websockets,Spring Framework,JPA Hibernate,and Spring Security[M].USA:Wrox Press,2014:295-314.

[3] 李刚.轻量级Java EE企业应用实战(第3版):Struts 2+Spring 3+Hibernate整合开发[M].北京:电子工业出版社,2012:362-429.

[4] Apace Maven.Welcome to Apache Maven[EB/OL].http//maven.apache.org/[2013-06-16].

[5] 李俊杰.Maven在企业Java软件产品中的应用[J].电脑知识与技术,2011,7(7):1562-1565.

[6] 李洋,孙永维,许冰,等.基于Ajax,Struts,Hibernate和Spring的J2EE架构[J].吉林大学学报(信息科学版),2011(06):577-578.

[7] 屈展,李婵.JSON在Ajax数据交换中的应用研究[J].西安石油大学学报(自然科学版),2011,1:96-98.

[8] 甘文丽,刘为超.基于Struts2和Ajax的企业级Web应用开发[J].工矿自动化,2013(02):24-26.

[9] Paul DuBois.MySQL[M].USA:Addison-Wesley Educational Publishers,2013:609-695.

作者简介：

杨佳绩(1997–)，男，本科生.研究领域：软件开发.

孟艳红(1973–)，女，硕士，讲师.研究领域：计算机网络与信息安全技术.