HZ BOOKS
华章IT

大 数 据 管 理 丛 书

位置大数据隐私管理

潘晓 霍峥 孟小峰 编著

机械工业出版社
China Machine Press

大/数/据/管/理/丛/书

位置大数据隐私管理

潘晓　霍峥　孟小峰　编著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

位置大数据隐私管理 / 潘晓，霍峥，孟小峰编著 . —北京：机械工业出版社，2017.3
（大数据管理丛书）

ISBN 978-7-111-56213-9

I. 位…　II. ① 潘…　② 霍…　③ 孟…　III. 数据管理　IV. TP274

中国版本图书馆 CIP 数据核字（2017）第 040463 号

本书在介绍了位置大数据等基本概念的基础上，总结归纳了传统位置隐私保护研究中经典的攻击模型和保护模型，详细介绍了若干基于数据失真的保护方法和基于数据加密的方法。全书共 6 章，内容包括位置隐私与隐私保护、典型攻击模型和隐私保护模型、快照位置隐私保护方法、动态位置隐私保护、连续轨迹数据隐私保护和面向隐私的查询处理技术。

本书可作为普通高等院校计算机和信息技术相关专业的大数据研究生课程的教材使用，也可供从事计算机相关的科研人员和学者作为技术参考。

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码：100037）
责任编辑：余　洁　　责任校对：李秋荣
印　　刷：北京诚信伟业印刷有限公司　　版　　次：2017 年 5 月第 1 版第 1 次印刷
开　　本：170mm × 242mm　1/16　　印　　张：11
书　　号：ISBN 978-7-111-56213-9　　定　　价：69.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88378991　88361066　　投稿热线：（010）88379604
购书热线：（010）68326294　88379649　68995259　　读者信箱：hzjsj@hzbook.com

当下大数据技术发展变化日新月异，大数据应用已经遍及工业和社会生活的方方面面，原有的数据管理理论体系与大数据产业应用之间的差距日益加大，而工业界对于大数据人才的需求却急剧增加。大数据专业人才的培养是新一轮科技较量的基础，高等院校承担着大数据人才培养的重任。因此大数据相关课程将逐渐成为国内高校计算机相关专业的重要课程。但纵观大数据人才培养课程体系尚不尽如人意，多是已有课程的“冷拼盘”，顶多是加点“调料”，原材料没有新鲜感。现阶段无论多么新多么好的人才培养计划，都只能在 20 世纪六七十年代编写的计算机知识体系上施教，无法把当下大数据带给我们的新思维、新知识传导给学生。

为此我们意识到，缺少基础性工作和原始积累，就难以培养符合工业界需要的大数据复合型和交叉型人才。因此急需在思维和理念方面进行转变，为现有的课程和知识体系按大数据应用需求进行延展和补充，加入新的可以因材施教的知识模块。我们肩负着大数据时代知识更新的使命，每一位学者都有责任和义务去为此“增砖添瓦”。

在此背景下，我们策划和组织了这套大数据管理丛书，希望能够培

养数据思维的理念，对原有数据管理知识体系进行完善和补充，面向新的技术热点，提出新的知识体系/知识点，拉近教材体系与大数据应用的距离，为受教者应对现代技术带来的大数据领域的新问题和挑战，扫除障碍。我们相信，假以时日，这些著作汇溪成河，必将对未来大数据人才培养起到“基石”的作用。

丛书定位：面向新形势下的大数据技术发展对人才培养提出的挑战，旨在为学术研究和人才培养提供可供参考的“基石”。虽然是一些不起眼的“砖头瓦块”，但可以为大数据人才培养积累可用的新模块（新素材），弥补原有知识体系与应用问题之前的鸿沟，力图为现有的数据管理知识查漏补缺，聚少成多，最终形成适应大数据技术发展和人才培养的知识体系和教材基础。

丛书特点：丛书借鉴 Morgan & Claypool Publishers 出版的 Synthesis Lectures on Data Management，特色在于选题新颖，短小精湛。选题新颖即面向技术热点，弥补现有知识体系的漏洞和不足（或延伸或补充），内容涵盖大数据管理的理论、方法、技术等诸多方面。短小精湛则不求系统性和完备性，但每本书要自成知识体系，重在阐述基本问题和方法，并辅以例题说明，便于施教。

丛书组织：丛书采用国际学术出版通行的主编负责制，为此特邀中国人民大学孟小峰教授（email：xfmeng@ruc.edu.cn）担任丛书主编，负责丛书的整体规划和选题。责任编辑为机械工业出版社华章分社姚蕾编辑（email：yaolei@hzbook.com）。

当今数据洪流席卷全球，而中国正在努力从数据大国走向数据强国，大数据时代的知识更新和人才培养刻不容缓，虽然我们的力量有限，但聚少成多，积小致巨。因此，我们在设计本套丛书封面的时候，特意选择了清代苏州籍宫廷画家徐扬描绘苏州风物的巨幅长卷画作《姑苏繁华图》（原名《盛世滋生图》）作为底图以表达我们的美好愿景，

每本书选取这幅巨卷的一部分，一步步见证和记录数据管理领域的学者在学术研究和工程应用中的探索和实践，最终形成适应大数据技术发展和人才培养的知识图谱，共同谱写出我们这个大数据时代的盛世华章。

在此期望有志于大数据人才培养并具有丰富理论和实践经验的学者和专业人员能够加入到这套书的编写工作中来，共同为中国大数据研究和人才培养贡献自己的智慧和力量，共筑属于我们自己的“时代记忆”。欢迎读者对我们的出版工作提出宝贵意见和建议。

大数据管理丛书

主编：孟小峰

大数据管理概论

孟小峰　编著

2017 年 5 月

异构信息网络挖掘：原理和方法

[美] 孙艺洲（Yizhou Sun）　韩家炜（Jiawei Han）　著

段磊　朱敏　唐常杰　译

2017 年 5 月

大规模元搜索引擎技术

[美] 孟卫一（Weiyi Meng）　於德（Clement T. Yu）　著

朱亮　译

2017 年 5 月

大数据集成

[美] 董欣（Xin Luna Dong）　戴夫士·斯里瓦斯塔瓦（Divesh Sriva-tava）　著

王秋月　杜治娟　王硕　译

2017 年 5 月

短文本数据理解

王仲远　编著

2017 年 5 月

个人数据管理

李玉坤　孟小峰　编著

2017 年 5 月

位置大数据隐私管理

潘晓　霍峥　孟小峰　编著

2017 年 5 月

移动数据挖掘

连德富　张富峥　王英子　袁晶　谢幸　编著

2017 年 5 月

云数据管理：挑战与机遇

[美] 迪卫艾肯特·阿格拉沃尔（Divyakant Agrawal）　苏迪皮托·达斯（Sudipto Das）　阿姆鲁·埃尔·阿巴迪（Amr El Abbadi）　著

马友忠　孟小峰　译

2017 年 5 月

大数据时代，移动通信和传感设备等位置感知技术的发展将人和事物的地理位置数据化，与用户位置相关的数据通过各种各样的服务以多种形式产生。例如，用户通过“签到”等移动社交网络服务（如Foursquare、Yelp、Flicker等）以文本、图片形式主动发布时空的行为。再如，通过用户手机通话、短信等记录，个人位置数据由基站自动隐式收集。无论自动发布还是被动收集的位置数据均具有规模大、产生速度快、蕴含价值高等特点。瑞典市场研究公司Berg Insight发布的最新报告预测，全球基于位置服务的市场规模到2020年将达到348亿欧元。位置大数据中蕴含人类行为的特征，在疾病传播、贫困消除、城市规划等重大社会科学问题以及路线推荐、乘车出行等重要生活应用中发挥了关键作用。

然而，位置大数据在带给人们巨大收益的同时，也带来了个人信息泄露的危害。这是因为位置大数据直接或间接包含了个人身份、行动目的、健康状况、兴趣爱好等多方面的敏感隐私信息。位置大数据的不当使用会给用户各方面的隐私带来严重威胁。已有的一些案例说明了隐私泄露的危害，如：某知名移动应用由于不注意保护位置数据，导致根据

三角测量方法可以推断出用户的家庭住址等敏感位置，引发多起犯罪案件；某著名移动设备厂商曾在未获得用户允许的情况下大量收集用户的位置数据，攻击者可以通过这些位置数据推测用户的身体状况等个人敏感信息。我国在十一届全国人大常委会第三十次会议上审议了《关于加强网络信息保护的决定草案》的议案，将个人信息保护纳入国家战略资源的保护和规划范畴，体现了国家对个人隐私保护问题的重视。随着个人隐私观念的增强以及相关法律法规的健全，如何在大数据多源数据融合的环境下既不泄露用户隐私又能提高位置大数据的利用率，如何保证在牺牲最小代价的前提下既满足服务质量要求又保护个人隐私，成为位置大数据隐私保护的研究重点。

本书的内容和组织结构

本书系统地介绍了位置大数据、基于位置服务、位置隐私等相关概念，总结归纳了传统位置隐私保护研究中经典的攻击模型和隐私保护模型，并举例说明了不同攻击模型的经典保护方法。随后分别针对用户静态快照位置、动态位置、连续轨迹介绍了相应的隐私保护方法，以及面向隐私的查询处理技术。

本书共分为 6 章，具体如下所示。

第 1 章介绍了位置大数据相关的基本概念、LBS 中的个人隐私保护问题所面临的主要挑战，以及典型的隐私保护技术。

第 2 章对典型攻击模型和相应的隐私保护模型进行了说明。

第 3 章针对用户的快照位置，分别介绍感知服务质量、无精确位置和无匿名区域的位置隐私保护方法。

第 4 章针对用户的动态位置，介绍了 3 种位置隐私保护技术，不仅考虑了移动用户的当前位置，同时顾及了用户的运动模式或未来位置。

第 5 章针对用户的历史位置数据，分别介绍了基于图划分的轨迹隐私保护技术、区分位置敏感度的轨迹隐私保护技术和基于前缀树的轨迹隐私保护方法。

第 6 章介绍一类在完全不泄露用户敏感查询信息的前提下，针对常见移动查询类型的面向隐私的查询处理技术。

致谢

孟小峰教授领导的中国人民大学网络与移动数据管理实验室自 2006 年即开始关注隐私保护这一领域的研究，先后针对位置数据隐私、轨迹数据隐私和位置大数据隐私等问题展开研究，取得了一系列研究成果，先后培养了多位隐私保护方面的博士。本书即是作者在多年研究成果的基础之上总结整理而成的。

首先感谢国家基金委和国家 863 计划的一贯支持，在连续十年间的研究中得到如下项目的资助：

2016～2020 年，国家自然基金重点项目“大规模关联数据管理的关键技术研究”，编号：61532010。国家自然科学基金重大研究计划“大数据驱动的管理与决策研究”重点项目“大数据开放与治理中的隐私保护关键技术研究”，编号：91646203。

2014～2017 年，国家自然基金面上项目“面向移动用户的 Web 集成技术研究”，编号：61379050。

2014～2016 年，国家自然基金青年项目“基于位置服务在受限网络中的个人隐私保护技术研究”，编号：61303017。

2011～2013 年，国家自然基金面上项目“Web 信息可信性研究”，编号：61070055。

2009～2011 年，国家 863 计划重点项目“普适计算基础软硬件关键技术及系统”课题“隐私保护技术”，编号：2009AA011904。

2014～2016 年，河北省自然科学基金面上项目和青年项目“基于位置服务中的隐私保护技术研究”，编号：F2014210068；“道路网络中轨迹隐私保护技术研究”，编号：F2015207009；“基于大数据的移动商务隐私感知推荐技术研究”，编号：F2015210106。

本书的形成凝聚了实验室的集体智慧。特别感谢实验室的博士生和硕士生们的工作，其中包括硕士生尹少宜、肖珍、谢敏、黄毅，以及博士生潘晓、霍峥、张啸剑、王璐等。潘晓和霍峥博士直接参与本书的写作，孟小峰教授负责审阅全书。

本书可作为普通高等院校计算机和信息技术相关专业的大数据研究生课程的教材，也可供从事计算机相关专业的技术人员和学者作为参考书。

感谢机械工业出版社华章公司的编辑们，他们在全文的校对和编辑出版过程中付出了巨大的努力。因作者水平有限，书中错误在所难免，恳请批评指正。

作　者

2016 年 10 月

潘　晓，石家庄铁道大学经济管理学院，副教授，商务信息系主任，中国人民大学计算机应用专业博士，师从孟小峰教授。曾在美国伊利诺伊大学芝加哥分校访学一年（2015-2016）。主要研究兴趣包括：数据管理，移动计算、隐私保护等。主持和参加了国家和省部级科研项目 4 项；在国际顶级或国内重要学术期刊和会议上发表学术论文近 20 篇；获国家专利 3 项；获北京市科技进步奖二等奖（排名第四）；2014 年被评为石家庄市青年拔尖人才，2015 年入选河北省“三三三人才工程”(第三层)，2016 年入选石家庄铁道大学第四届优秀青年科学基金项目。

霍　峥　河北经贸大学讲师，中国人民大学计算机软件与理论专业博士，师从孟小峰教授。目前从事计算机软件与理论方向的教学与研究。主要讲授的课程包括：数据库原理、数据结构、离散数学等。主要研究方向：移动对象数据管理、位置与轨迹隐私保护技术等。主持和参与了多项国家级科研项目的研究工作，发表论文 10 余篇，获省部级奖励 1 项。

孟小峰　中国人民大学信息学院教授，博士生导师。现为中国计算机学会会士、中国保密协会隐私保护专业委员会副主任、《Journal of Computer Science and Technology》、《Frontiers of Computer Science》、《软件学报》、《计算机研究与发展》等编委。先后获中国计算机学会“王选奖”一等奖(2009)，北京市科学技术奖二等奖（ 2011 ）等奖励，入选“第三届北京市高校名师奖”（ 2005)。发表论文 200 余篇，获得国家专利授权 12 项。近期主要研究领域为网络与移动大数据管理，包括 Web 数据管理、云数据管理、面向新型存储器的数据库系统、大数据隐私管理、社会计算等。