

2017 届研究生硕士学位论文

分类号: _____

学校代码: 10269

密 级: _____

学 号: 51141500081



华东师范大学

East China Normal University

硕 士 学 位 论 文

MASTER'S DISSERTATION

**论文题目: 基于位置服务的隐私保护
关键技术研究**

院 系: 软件学院
专业名称: 软件工程
研究方向: 密码与网络安全
指导教师: 董晓蕾 教授
学位申请人: 孙浩

2017 年 1 月

Dissertation for master degree in 2017

University Code: 10269

Student ID: 51141500081

EAST CHINA NORMAL UNIVERSITY

**THE RESEARCH KEY TECHNOLOGY OF
PRIVACY PROTECTIONS FOR
LOCATION-BASED SERVICES**

Department: Software Engineering Institute

Major: Software Engineering

Research direction: Cryptography and Network Security

Supervisor: Prof. Dong Xiaolei

Candidate: Sun Hao

2017.01

华东师范大学学位论文原创性声明

郑重声明：本人呈交的学位论文《基于位置服务的隐私保护关键技术研究》，是在华东师范大学攻读硕士/博士（请勾选）学位期间，在导师的指导下进行的研究工作及取得的研究成果。除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确说明并表示谢意。

作者签名:_____

日期: 年 月 日

华东师范大学学位论文著作权使用声明

《基于位置服务的隐私保护关键技术研究》系本人在华东师范大学攻读学位期间在导师指导下完成的硕士/博士（请勾选）学位论文，本论文的研究成果归华东师范大学所有。本人同意华东师范大学根据相关规定保留和使用此学位论文，并向主管部门和相关机构如国家图书馆、中信所和“知网”送交学位论文的印刷版和电子版；允许学位论文进入华东师范大学图书馆及数据库被查阅、借阅；同意学校将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于（请勾选）

- () 1. 经华东师范大学相关部门审查核定的“内部”或“涉密”学位论文*，于年月日解密，解密后适用上述授权。
- () 2. 不保密，适用上述授权。

导师签名:_____

本人签名:_____

年 月 日

* “涉密”学位论文应是已经华东师范大学学位评定委员会办公室或保密委员会审定过的学位论文（需附获批的《华东师范大学研究生申请学位论文“涉密”审批表》方为有效），未经上述部门审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权）。

孙浩 硕士学位论文答辩委员会成员名单

姓名	职称	单位	备注
曹珍富	教授	华东师范大学	主席
董晓蕾	教授	华东师范大学	
张磊	研究员	华东师范大学	
何道敬	教授	华东师范大学	
周俊	副教授	华东师范大学	

摘要

智能终端的普及推动了移动互联网进入一个高度发展的时代。GPS 定位技术以及无线通信的日益完善也给人类的衣食住行提供了极大的便利，如今交通、医疗、教育等这些和民众生活息息相关的服务都离不开智能终端、无线通信等技术的支持。基于位置的服务（LBS; Location-Based Service）是地理位置和移动互联网结合，是当下信息化时代比较耀眼的模式之一。利用 LBS，用户可以获得当前位置下的兴趣点（餐厅、KTV、影院...），但同时也暴露出一些隐私问题。用户和 LBS 服务提供商提出查询服务的时候，首先需要通过 GPS 获取到自己当前的位置，然后通过基站将位置发送给 LBS 服务提供商。由于用户的位置也许会揭露出一些敏感个人信息，而用户的当前位置完全暴露在 LBS 服务提供商面前，因此对用户的位置隐私构成了威胁。

本文主要工作包括以下几个方面：

- **位置隐私保护方法的对比分析** 提出了海量轨迹数据的分布式处理框架，分别讨论通用轨迹数据处理中的噪声过滤、路网匹配和特征抽取三个阶段利用 Map-Reduce 的计算方案，并实现了本文的路网匹配 RouteFit 算法。
- **针对 KNN 查询的隐私保护算法** 提出了海量轨迹数据的分布式处理框架，分别讨论通用轨迹数据处理中的噪声过滤、路网匹配和特征抽取三个阶段利用 Map-Reduce 的计算方案，并实现了本文的路网匹配 RouteFit 算法。
- **时空数据聚合隐私保护算法** 采用基于密度的聚类方法来发现位置点数据中的兴趣点和兴趣区域，通过实现 Pick-up DBScan 算法来完成对出租车轨迹数据中具有语义特征的上下客位置点的聚类，生成候选出租车扬招 POI 和热门目的地 ROI，为推荐提供重要数据集。
- **基于差分隐私的位置隐私保护算法** 介绍了利用海量出租车轨迹数据来优化出行的位置推荐服务，提出了出租扬招位置查询和候车时间预测系统，以推荐合理的出租车扬招位置点和预测准确的候车时间为目地，离线处理部分通过分布式轨迹处理框架完成轨迹预处理和特征抽取工作，以路段聚类的方法来划分模型粒度，设计多种空车等候时间的预测模型并进行评估和选择，在线查询部分利用空间索引技术和 Web 服务技术实现对输入查询点的实时位

置推荐服务，最后实现了基于上海市区大规模出租车轨迹历史数据的处理和分析预测的原型系统，提供对出租车扬招点得位置推荐服务。

关键词: 移动推荐, LBS, 轨迹挖掘

ABSTRACT

Being equipped with built-in GPS devices, thousands of taxis generate a large-scale collection of trajectory data in metropolitan areas everyday. Such data plays an essential role in variety of well-established location-based service (LBS) applications. So far taxi GPS data has been used for traffic modeling and urban computing. Examples include congestion prediction, itinerary planning, convoy detection, amongst others. Related LBS research works face challenges due to large data scale and low data quality. Gaining insights about the prevalence of distributed platforms, i.e. Hadoop, can provide useful tools to process and analyze large scale data. Thus, we introduce a framework to solve modeling and analysis of massive trajectory using distributed platform. Based on the processing framework, we implement a recommendation system which will answer queries of recommended pick-up points and predict vacant taxi waiting time for passengers.

Main contributions of this paper are as follows:

- **Distributed Big Trajectory Data Analysis Framework** Proposed a distributed framework for massive trajectory processing. We split the general trajectory process into three phrases: noise filtering, map matching and feature extraction, then present the Map-Reduce compute paradigm for each of them. RouteFit algorithm is implemented for map matching in our work.
- **Clustering-based Region of Interest Discovery** Using density-based clustering algorithms to discover Point of Interest or Region of Interest in location data. Pick-up DBScan algorithm is developed in this paper to generate pick-up clustering from the pick-up points in trajectory data. Then, a candidate set of taxi pick-up POI, which will be the recommendation items, is generated using the clustering results.

- **Pick-up Point Recommendation and Waiting Time Prediction** We introduce the location-based service for travel optimization using large-scale taxi trajectories. Taxi pick-up points recommendation and waiting-time prediction system aims to recommend efficient locations for taxi hailing and give precise waiting-time prediction for passengers. The offline part re-builds the traffic prediction models using recent data periodically. Firstly, in the preprocessing module, we filter the raw G-PS data with noise and errors. Then, the road segments are clustered into groups to reflect different traffic situations and generate ST-unit as modeling granularity. Finally, we build regression models and Poisson process models and then choose the best models for each ST-unit by evaluating sampled test set. The online part processes queries and gives real time pick-up points recommendations taking advantage of spatial indexing and web service techniques. Finally, we implement the prototype system to provide taxi pick-up points recommendation based on Shanghai taxi trajectory data.

Keywords: *Mobile Recommendation; Location-based service; Trajectory mining.*

目录

第一章 绪 论	1
1.1 研究背景	1
1.2 隐私保护研究现状	3
1.3 本文工作与主要贡献	5
1.4 组织结构	7
第二章 位置隐私保护技术	8
2.1 LBS 应用模式	8
2.1.1 用户提问 - 服务器应答	8
2.1.2 服务器提问 - 用户应答	9
2.2 隐私保护系统结构	10
2.2.1 独立式结构	10
2.2.2 分布式点对点结构	11
2.2.3 中心服务器结构	13
2.3 隐私保护技术	14
2.3.1 假名技术	14
2.3.2 假位置技术	16
2.3.3 区域覆盖技术	19
2.3.4 密码学技术	22
2.4 本章小结	22
参考文献	23

第一章 緒論

1.1 研究背景

随着移动定位技术、无线通信技术、普适计算技术的快速发展，以及智能手机的大量普及，人类进入了一个新的信息化时代——移动互联网时代。移动互联网已经融入到了交通、教育、娱乐、金融等人类生活的各个领域，对人类的发展有着极为深刻的意义。在早期的普通互联网时代，人们只能通过固定的 PC 终端连接到互联网，由于 PC 终端体积庞大，不具有便携性，极大的限制了互联网的发展。而在如今的移动互联网时代、智能手机的普及，使得人们可以不受时间和地点的限制，无论是在逛街还是旅途中都可以方便的连接到互联网。移动互联网的重要特性之一就是与地理位置的结合，这一特性使得移动互联网服务和应用更加丰富多彩，影响更加深远。

地理位置和移动互联网的结合促进了基于位置服务（Location-Based Service,LBS）相关应用的产生及快速发展，Foursquare、Loopt 和新浪微博等在国内取得的成功足以彰显出这种新兴的 LBS 隐藏着巨大的商业价值。在 LBS 应用中，用户通过定位设备（如 GPS, 传感器, RFID 等）可以随时随地的获取到自身的当前位置。当用户需要 LBS 服务器（如大众点评）提供某种服务时，例如，娱乐信息服务（如查询距离我当前位置最近 KTV、游戏厅或商店），导航服务（如查询当前位置到达华东师范大学的最短路径），交通服务（如查询“距离我最近的服务站”）。如图1.1所示：① 移动用户首先通过定位设备（卫星）获取到自己当前的位置、② 用户将自己的当前位置和查询需求（查找当前位置周围的 KTV）发送给位置服务提供商、③ 位置服务提供商在收到用户的请求消息后，会根据相应的算法

(如 k 邻近算法, KNN)结合数据库选出和用户请求最为匹配的结果、④服务提供商将匹配结果返回给移动用户。简而言之, LBS 应用就是位置服务器根据用户提供的自身位置信息为用户提供的增值服务 [1]。

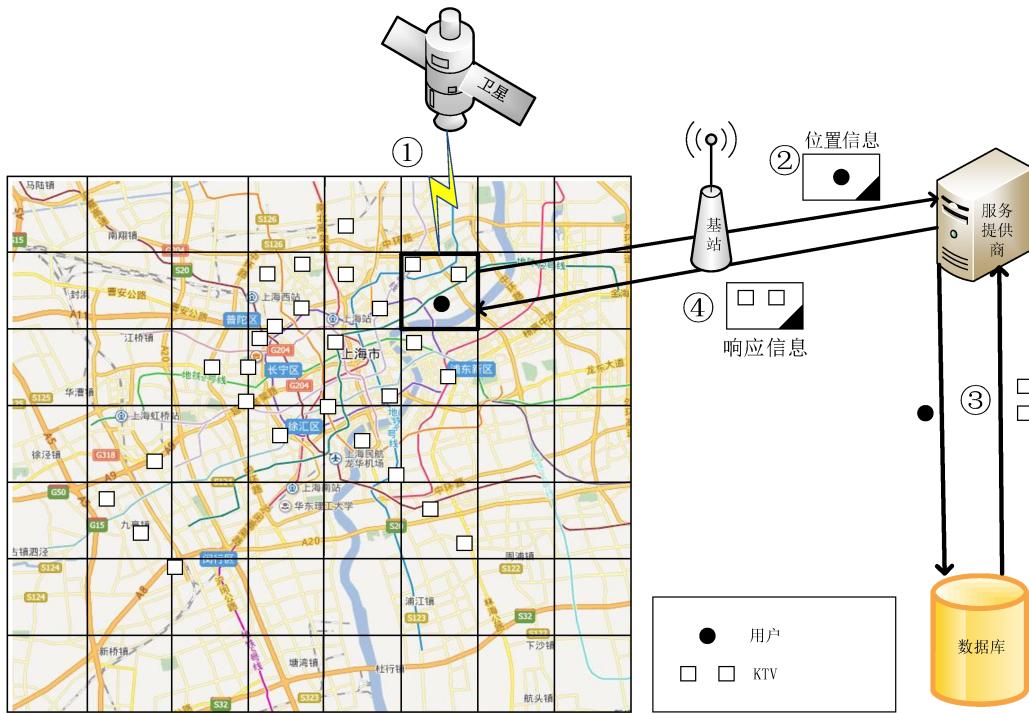


图 1.1: 基于位置服务通信模拟图

在 LBS 应用中用户向 LBS 服务提供商提供自己的位置以及个人信息, 服务器可以对这些数据进行统计归纳。这些统计信息无论是对个人还是对社会都有着极为重要的意义, 个人可以通过这些信息挑选出最合适的餐馆就餐, 到最合适的商场进行购物, 利用位置的分布信息, 可以避免上下班拥堵路段等。商户可以利用这些信息指定出更合适的销售计划, 为不同的群体用户定制不同方案从而提供销量。政府可以利用这些信息进行道路交通的疏导, 合理的部署公共基础设施等。可见, 合法的利用位置信息对人们的生活, 社会的发展有着重要的影响。

然而 LBS 应用的发展并非一帆风顺, LBS 应用在给人们生活带来极大便利的同时, 也对人们的隐私造成了一定的威胁。LBS 服务提供商收集移动用户当前的

位置信息，在此基础上执行空间查询，为用户提供位置服务。用户位置提供的越精确，LBS 服务提供商返回的服务信息也越精确，但使得用户的位置暴露的越详细。这些位置信息也许揭露的不仅仅是用户的经度和维度，知道移动用户的位置信息也许可以知道用户在干嘛：正在参加会议还是听一场音乐会又或者是在哪里度假。当用户的位置信息被收集后，通过数据挖掘或分析，也许会得出用户的日常出行习惯以及出行路线等等。通常情况下，LBS 服务提供商将收集到的用户位置信息以一定的格式存放在数据库中，一旦存放这些数据的服务器被敌手攻破，那无疑是为攻击者窃取用户的隐私提供了极大的便利。

2010 年微软在英国、德国、日本、美国以及加拿大进行了一份调查，调查结果显示 94% 的消费者在使用 LBS 应用时会考虑他们带来的价值，但在相同地调查中发现 52% 人会关注他们的隐私是否存在泄露 [2]。位置信息的泄露对个人名誉以及人生安全都有严重的威胁。据国外媒体统计，78% 的小偷使用 Facebook、Twitter 对目标进行定位，从而确定主人是否在家中 [3]。如何权衡服务质量与隐私保护之间的矛盾，已经成为 LBS 中位置隐私保护亟待解决的核心问题。

1.2 隐私保护研究现状

隐私保护一直是国内外学者研究的热点之一，研究内容主要有隐私保护数据的发布、用户空间位置的隐私保护。

数据发布是当前数据挖掘、数据分析到信息共享的一个重要环节。信息大爆炸以来，人们可以明显感受到大数据的来势凶猛。据相关调查显示，目前全球互联网每天的流量累计达 1EB(即 10 亿 GB 或 1000PB)，这意味着每天产生的信息量可刻满 1.88 亿张 DVD 光盘。海量数据如同一座未开采的金矿，里面包含着无尽的信息与财富。但与此同时，也给数据的隐私带来了威胁。例如，通过对超市顾客的购买商品的记录进行分析，可以发现各种商品之间的关联（如啤酒与尿布），从而更好的进行货架的物品整理。然而在挖掘和分析的过程当中，不可避免的会使得顾客的信息暴露，从而可能造成顾客敏感信息的泄露。在文献 [4] 中，通

通过对性别、出生日期、住址等属性对选民登记表和隐藏了唯一标识符的医疗信息表进行连接操作，发现超过 87% 的美国公民的身份可以被标识。因此，如何解决数据发布过程中存在的隐私泄露问题，已成为隐私保护研究的重点对象，也由此产生了一个新的研究领域——隐私保护数的发布。

数据发布的隐私保护技术主要有数据加密、数据匿名、数据扰乱等隐私保护技术。数据加密技术主要是基于密码学的隐私保护技术，文献 [5] 利用群签名短发自主生成一系列伪签名证书来达到隐私保护效果，文献 [6] 利用全同态加密方法使得服务器在不知道任何明文的内容的情况下可以在密文域上进行运算操作，得到的结果与在明文进行运算的结果相同。文献 [7] 利用安全多方计算，可以保证在他人无法获得个人数据内同的情况下，计算出想要的结果。数据扰乱是一种数据失真的技术，Dwork 等人提出了一种典型的数据扰乱的隐私保护模型——差分隐私模型 [8]，通过对发布的数据添加噪声进行随机扰动，使得在统计意义上攻击者无论具有何种背景知识，都不能判断一条记录是否存在原始数据表中。基于数据匿名的隐私保护技术主要是通过 k -匿名技术 [9]，在一个满足 k -匿名的数据库中，对于某一个准标识符 (QID , Quasi-Identifiers)，值相同的记录至少有 k 条记录，因此通过 QID 去推断某一个目标记录的概率最多为 $1/k$ 。

用户空间位置的隐私保护旨在保护用户当前的位置，近些年来，出现了很多位置隐私保护技术，在一定程度上保护了位置的隐私。这些技术主要包括：信息访问控制 (Information access control) [10][11]、混合区域 (Mix zone)[12]、 k -匿名技术 (k -anonymity)[13][14][15]、假地址技术 ("Dummy locations") [16][17]、地理数据转 (Geographic data transformation)[18][19]、私有信息检索 (Private Information Retrieval, PIR)[20][21]。

基于访问控制，混合区域以及 k -匿名 LBS 查询需要服务提供商或者中间件维护所有用户的位置。当服务器提供商/中间件由不可信方代理，受到的保护力度会相应的降低，因此容易受到第三方的攻击。在过去，私人数据无意间就暴露在互联网上。

k -匿名最初用在身份隐私保护。将 k -匿名用在位置的隐私保护有点不适当，在位置的隐私保护概念中位置之间的距离是最重要的（身份隐私保护中身份之间的间隔是重要因素）。基于 k -匿名的 LBS 查询精度很大程度上受到移动用户的密度和分布的影响，而这一影响因素已经超过了位置隐私保护技术所能控制的范围。

基于假地址的 LBS 查询需要移动用户随机选择一组虚假位置集合，通过移基站将虚假位置发送给 LBS 服务提供商并从服务商那里获得一份错误的报告。这将导致移动设备的通信和计算量过载。为了提高效率，移动用户也许减少集合中虚假位置的数量，但这将导致弱隐私性。

基于地理数据转换的 LBS 查询易受到访问模式攻击 [22]，因为相同的查询总是返回相同的加密结果。例如，LBS 服务提供商可以观察返回密文出现的频率，依靠数据库内容的相关背景知识，可以根据出现频率匹配出最有可能的明文结果，从而得到相关的查询信息。

基于 PIR 的 LBS 查询提供了很强的密码保障，通过数据加密使得服务器无法得到用户位置的信息，且能对用户的请求提供正常的服务。相比于之前的位置隐私保护技术，PIR 技术对位置隐私保护的更加的安全，从理论上完全杜绝了敌手的攻击。

1.3 本文工作与主要贡献

本文以海量出租车轨迹数据为研究对象，基于已有研究成果，以智能打车推荐为应用目标，建立对轨迹数据的分布式处理框架和挖掘分析系统，并实现在线的查询与推荐服务。解决的问题包括：轨迹预处理、轨迹数据聚类、轨迹数据查询、预测和推荐模型建立等多个方面。本文主要的研究工作内容如下：

对轨迹数据的分布式处理 在对轨迹数据进行挖掘和分析的之前，数据的预处理工作能够提高模型准确度并辅助模型抽取出所需数据，通常包括对数据的降噪处理和过滤、轨迹数据到路网的映射等。鉴于轨迹数据的数据规模，并行的

数据处理策略能够大大提高对批量历史数据处理的效率，本文工作中建立基于 Map-Reduce 的通用轨迹处理框架，实现在不同采样密度下优化的路网匹配算法，并将分布式处理框架应用于数据过滤、路网匹配、特征抽取等轨迹数据处理的多个关键阶段。

兴趣点和兴趣区域挖掘 兴趣点和兴趣区域通常作为推荐元素向用户推荐，在不同的挖掘任务中，根据推荐的目标不同采用的方法也不同，聚类是发现轨迹数据特征的最常用方法之一，而对于时空特性明显的地理位置数据，聚类算法的设计、度量方法的选择、数据查询结构等均是该部分的主要研究内容。本文以打车推荐为目的，重点讨论采用基于密度的聚类方法对候选扬招点和热门目的地的挖掘方法。

出租车扬招推荐服务和候车时间预测 基于位置的服务泛指一类利用定位技术获得当前位置信息，再通过无线网络得到某项服务的技术，能够为大量普通用户提供服务。利用历史出租车轨迹数据，我们可以为用户提供智能出行建议，减少用户行程中浪费不必要的时间。本文基于多种模型对不同路段出租车空车到达时间进行建模，利用兴趣点挖掘技术提供备选扬招点和目的地方案，建立预测准确、推荐合理的扬招点推荐系统，并提供查询应用服务。

本文主要贡献如下：

- 建立了基于 Map-Reduce 的分布式轨迹处理框架，并在此框架下实现了轨迹预处理和优化的路网匹配算法。
- 采用基于密度的聚类方法发现兴趣点和兴趣区域，从而找到备选扬招点和热门目的地。
- 实现了智能出租车出行推荐系统，该包含了数据预处理、路网匹配、特征抽取、路段聚类、在线预测、查询推荐等多个模块，完成了基于泊松过程、线

性回归等的出租车等待时间预测算法。

- 提供在线扬招点查询和推荐应用，能够支持多种客户端实时的请求相应。

1.4 组织结构

第 2 章中回顾基于位置的服务的发展以及与本文工作相关的研究进展；第 3 章中介绍了本文实现的分布式轨迹数据处理框架，并具体阐述了本文采用的路网匹配方法和实验效果；为了实现后续的基于位置的打车推荐服务，第 4 章中介绍了利用轨迹点的聚类方法进行兴趣点和兴趣区域的挖掘，包括关键辅助索引结构和优化的基于密度的聚类算法；第 5 章基于前面的轨迹数据和聚类结果，实现了对扬招点和目的地的推荐应用，其中，重点介绍了对候选扬招点候车时间的预测模型，以及在线的查询和推荐算法。第 6 章总结全文，对后续的研究工作进行展望。

第二章 位置隐私保护技术

本章首先介绍 LBS 应用模式的分类，主要分为两种：“用户提问 - 服务器应答”模式、“服务器提问 - 用户应答”模式 [23]。然后对位置隐私保护的系统结构进行了分类介绍，具体包括独立式结构、分布式点对点结构、中心服务器结构三个方面。之后对 LBS 应用中的隐私保护技术进行了介绍，最后对隐私保护技术进行对比和总结。

2.1 LBS 应用模式

移动互联网与地理维度的巧妙结合，推动了 LBS 的快速发展。LBS 应用已经渗透到了人们的衣食住行。交通部分利用 LBS 可以对交通拥挤路段进行分流，企业利用 LBS 可以更具有针对性的广告投放，民众可以利用 LBS 寻找距离自己最近且性价比更高的娱乐设施。通过对 LBS 应用模式的总结，可以将其分为“用户提问 - 服务器应答”模式、“服务器提问 - 用户应答”模式。

2.1.1 用户提问 - 服务器应答

在此模式中，用户是请求的发起者。当用户需要获得 LBS 提供的服务时，用户通过定位设备获取到当前所在位置，然后将自己当前位置以及需求发送给服务提供商，服务器收到用户的请求后，根据用户的位置以及需求，返回给用户相应的结果。在此过程根据用户的请求方式，又可以此模式分为以下两类：

单次查询应用 单次查询应用是当前应用最广泛，也是技术最成熟的模式。在此类场景中，用户发送给服务器一个当前时刻的位置信息和请求，服务器收到请求

后，根据位置信息向用户提供此时此地的个性化信息，用户得到了直接的服务。如同现在的美团、大众点评就是典型的单次查询应用，可以获得距离用户此时位置最近的影院（K 近邻问题），也可以查询 1 公里以内最便宜的 KTV（区域排序问题）。此模式中用户只提交一次或少许几次的请求查询服务，服务器得到是用户当前的静态位置。

连续查询查询应用 在此模式中，用户需要持续不断的向服务器发送自己当前的位置信息。典型的应用就是现在的导航系统，用户在使用导航服务的时候，需要时时的向服务器发送自己的当前所在位置，服务器收到用户的实时位置信息后，为用户推荐正确的路线，提醒用户哪里会出现监控以及哪条路段有速度限制。此外还有如今也有不少 APP 需要用户实时向服务器提供当前的位置，比如微米，他能够发现此时此刻周围的好友，当你在购物的时候，如果你的好友刚好也出现在商场周围，那么 APP 将会推送一条实时消息，通知你有好友也在逛街，与此同时，你的好友也会收到类似的消息。在此应用中，服务器不仅可以获得用户的静态位置，还能获得用户的运动轨迹信息。

2.1.2 服务器提问 - 用户应答

此模式中的角色与“用户提问 - 服务器应答”模式角色信息相反，此模式的服务器是请求的发起者，服务器会向用户请求一些特定的数据，用户接受到请求消息后将相应的个人数据发送给服务器。此模式中服务器可以收集到大量的用户信息，服务器可以对这些数据进行分析，挖掘出一些隐藏的有利信息。因此“服务器提问 - 用户应答”模式在数据统计场景方面用的比较广泛。例如可以实时手机公交车、出租车等交通工具的位置信息，等待时间，从而预测路段的拥挤情况，及时的进行拥堵路段的车辆分流。特别的，如今电子钱包都被植入手机（支付宝、Apple Pay），商家可以利用顾客在何地消费，挖掘出潜在的商业价值，此类模式在今后的发展中应该会越来越好，伴随的应用也会随之增多。

2.2 隐私保护系统结构

在 LBS 中，隐私保护技术主要分为三种类别：独立式结构（Non-cooperative Architecture）、分布式点对点结构（Peer-to-Peer Architecture）、中心服务器结构（Centralized Architecture）[24]。

2.2.1 独立式结构

独立式结构 [25] 是一个典型的客户端/服务器（Client/Server,C/S）结构，主要构成部分为移动用户（客户端）与位置服务提供商（服务器），如图2.1所示。独立式结构要求客户端具具有自身定位、数据存储、数据计算的能力。移动用户根据自身的隐私需求，设置合理的隐私保护方案，将自己当前位置进行匿名化处理。匿名处理完成后，用户将位置匿名结果和查询内容通过移动互联网一起发送给位置服务器；服务器在接收到请求后，根据匿名后的位臵进行查询处理，并将查询结果返回给移动用户；移动用户在收到位置服务器返回的结果后，根据自己当前的真实位臵选出正确的结果。

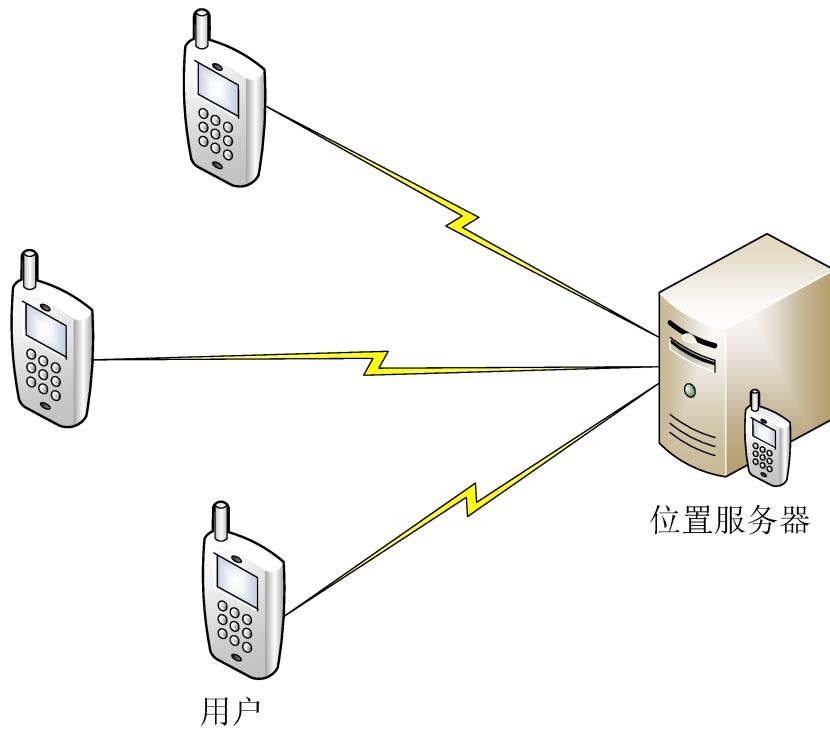


图 2.1: 独立式结构示意图

独立式结构主要有客户端和位置服务器两部分构成，结构简单，容易扩展。但由于客户端是独立存在的，不能达到负载均衡，因此要求客户端需要具有一定的数据计算与存储能力，而现在的可便携设备（如手机，智能手表，GPS 导航仪等）的计算和存储能力都比较有限。

2.2.2 分布式点对点结构

分布式点对点结构 [26] 同独立式结构一样，同为 C/S 结构，都是由移动用户和位置服务器两部分构成的。不同的是，独立式结构中的移动用户（客户端）是单独存在的个体，而分布式点对点结构中移动用户之间相互通信，构成一个群体，如图所示图2.2。群体中的每个移动用户都是平等的，都具有一定的数据存储和计算能力的通信设备。

分布式点对点的信息请求与查询处理过程主要分为两个步骤：I. 移动用户找

到当前位置周围的其他用户，根据自身的隐私需求，选择合适的匿名算法，将自己位置隐匿在用户位置组当中，并将匿名后的位置发送给位置服务器。在图2.2中，假如移动用户①为提出查询请求的用户，用户①的当前位置周围有用户②、用户③、用户④三个用户。用户①选择匿名算法，将自己位置隐匿（可以将位置转为周围任何一位用户的位置，此处假设隐匿为用户③的位置），用户③将用户①匿名后的位置发送给位置服务器。II. 位置服务器收到用户的位置查询请求后，作出响应，并将结果返回给用户③。用户③接收到服务器端的响应后，将处理请求结果发送给用户①。

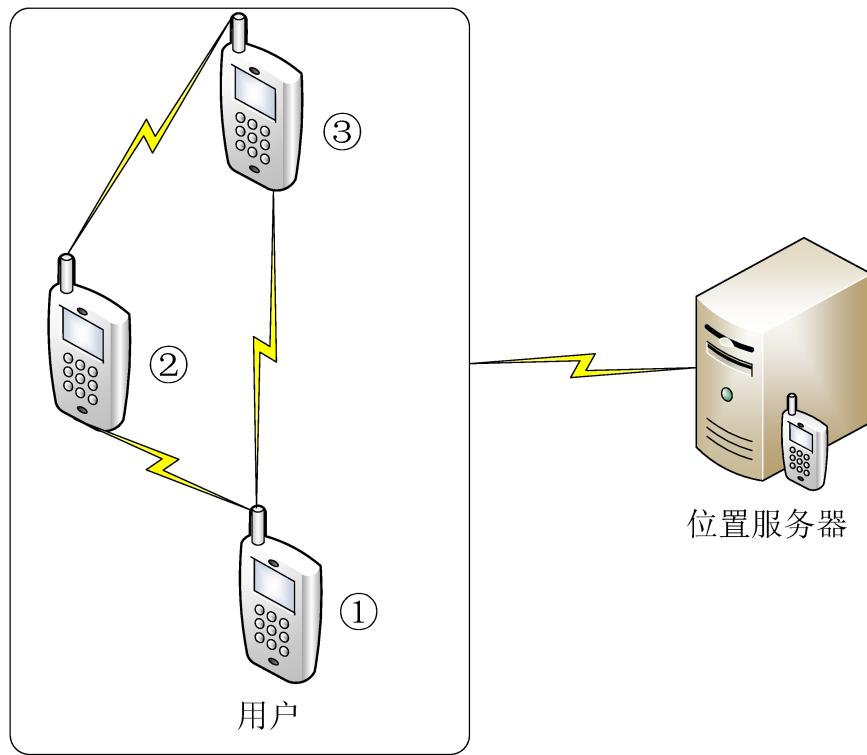


图 2.2: 独立式结构示意图

相对于独立式结构，分布式点对点结构通过分布式计算减轻了个体用户的计算和存储开销，达到了负载均衡，但是由于各个用户（客户端）之间需要互相通信，使得网络间的通信代价很高，存在通信延迟，一旦出现网络风暴，可能导致系统瘫痪，消息请求响应失败。

2.2.3 中心服务器结构

由于位置服务器通常是半可信甚至不可信的，因此中心服务器结构中移动用户与位置服务器之间不直接进行通信，而是在两者之间增加了一个可信第三方 - 匿名服务器，如图2.3所示。匿名服务器主要有如下作用：1. 移动用户将当前位置的确切信息发送给匿名服务器，匿名服务器负责收集用户的位置信息，当用户位置信息发生变化后，负责更新用户的位置信息。2. 匿名服务器在收到移动用户的精确位置信息后，根据匿名算法，将用户得精确位置信息转换为隐匿区域，并将匿名后的位置发送给位置服务器。3. 位置服务器对匿名后的区域进行查询处理，将查询处理结果发送给匿名服务器，匿名服务器收到候选结果后，选择正确的响应结果发送给相应的移动用户。

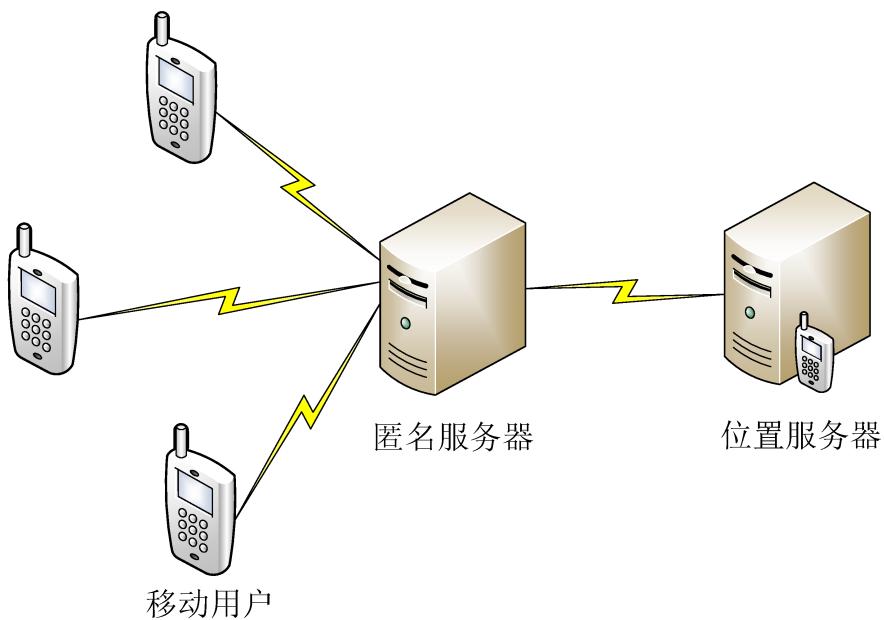


图 2.3: 独立式结构示意图

在独立式结构中位置匿名化由移动端进行处理，移动端通常是计算性能较低可便携或可穿戴设备，容易造成计算瓶颈。在中心服务器结构中，降低了移动客

客户端的计算和存储开销，同时还能满足用户的隐私需求。但是由于匿名服务器需要对用户的位置更新，位置的匿名进行处理，一旦操作频繁，这些请求容易成为匿名服务器的处理瓶颈。另外，匿名服务器掌握着移动用户的精确位置信息，一旦匿名服务器被敌手攻破，将会使得用户资料的泄漏。

2.3 隐私保护技术

基于位置服务的隐私保护的主要目的是用户在享受高质量的服务的同时又不会将自己的个人信息以及位置信息暴露给服务提供商。最近几年，国家大力提倡网络安全，网络空间安全也被列为教育部一级学科。与此同时，人们对 LBS 应用的隐私保护问题关注度也日益高涨，国内外学者对此已有大量的研究总结 [5][12][25]。当前 LBS 应用的隐私保护研究重点集中在位置信息的隐私保护上面，我们首先给出用户发送给服务提供商的服务请求定义：

定义 2.3.1 (：LBS 服务请求) 移动用户向位置服务提供商提出查询请求 R ，可以抽象定义为下面的三元组： $R = (UserID, Position, Query)$

其中， $userID$ 代表移动用户的唯一身份标识符，如身份证号、手机号码；而 $Position$ 可以代表一个真实的精确位置，如 GPS 定位中的经度和纬度，也可以代表用户所在位置的一个模糊区域。 $Query$ 代表用户查询的内容，如“查询附近的影院”。

由定义 2.3.1 可知，通过对 $UserID$ 进行保护处理，让服务提供商无法识别用户的 ID。也可以对 $Position$ 进行变换处理，让服务提供商无法通过 $Position$ 获得用户的真实位置。综合这两方面，下面将详细介绍四种类型的隐私保护技术：(1) 假名技术；(2) 假位置技术；(3) 区域覆盖技术；(4) 密码学技术。

2.3.1 假名技术

假名技术属于对用户 $userID$ 进行保护的一种方法，当用户向位置服务提供商发送请求的时候，采用虚假的 $userID$ 代替用户的真实 $userID$ 。这样位置服务提供

商就无法收集 userID 与位置信息的对应信息。即使存在某个攻击者获得了用户名和对应的位置信息，由于用户名是伪造的，因此不能对用户的隐私造成危害。

混淆的概念早期被应用于网络间的通讯，如今在隐私保护的应用方面也得到了各位学者的青睐。假名技术的代表就是混淆区域（Mix-Zone）技术。Mix-Zone 将地图划分为两个区域：应用区域和混淆区域 [27]。在应用区域中，用户无需做任何操作，可以正常的享受位置服务提供商所提供的服务。当用户从应用区域进入混淆区域后，用户将不能向位置服务提供商发送自己的位置信息。另外在用户离开混淆区域之前，用户将同步更新自己的身份信息，并且使用是一个之前未曾使用过的假名代替现有的名字。如图2.4，当一个用户从混淆区域出来后，服务提供商无法将用户和当前混淆区域中的其他用户区分开来，从而实现了混淆区域中的用户的 k -匿名保护，在攻击者看来目标用户和其他 $k-1$ 个用户在准标识符 QI 上相一致。由于用户在经过不同混淆区域的时候，都会生成新的且从未使用过的假名代替当前的名字，这样使得用户信息的隐私保护程度得到了增强。

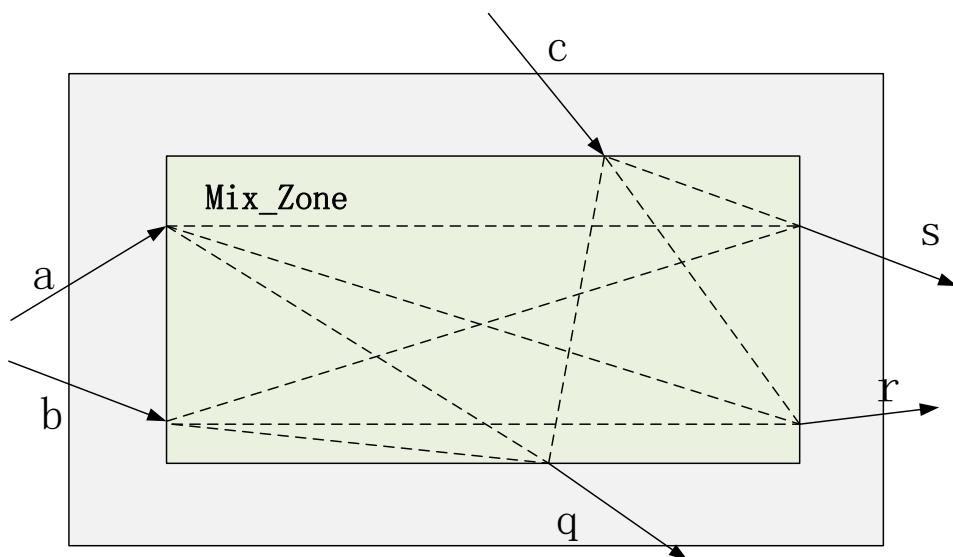


图 2.4: 独立式结构示意图

表给出了拥有三个用户的混淆区域的实例，在表中，三个用户的真实身份分

真实身份	假名 1	时间 1	假名 2	时间 2	停留时间
甲	A	1	X	5	4
乙	B	3	Y	6	3
丙	C	5	Z	8	3

表 2.1: 混淆区域实例

别为甲、乙、丙。三个用户分别在第 2、5、8 个时间点进入混淆区域，且以 A 、 B 、 C 作为其各自的假名。之后进入混合区域 2，此时每位用户又分别以新的假名 X, Y, Z 作为各自的新身份，在第 7、8、12 个时间点用户分别走出混淆区域。综上，我们可以得出每位用户在混淆区域内的时间分别为 5、3、5。由于攻击者无法预测用户在混淆区域内停留时间的长短，当混淆区域用户人数数量较多的时候想要关联用户的身份信息难度很大。

由于 Mix-Zone 技术限制了用户在混淆区域时任何用户都不能将自己的位置发送给位置服务提供商，当用户进行持续位置服务请求的时候（例如导航），混淆区域技术就不能满足这方面需求，因为用户会有一段时间进入“盲区”。另外混淆区域的大小，以及混淆区域内用户的数量，也会对隐私保护的质量带来一定的影响。

2.3.2 假位置技术

在发布假位置技术中，用户将自己当前的真实位置以及生成的假位置发送给位服务提供商。提供商根据每个位置信息作出响应，并将响应消息发送给用户，用户收到反馈信息后，仅从中抽取真实信息。图2.5描述了假位置的 LBS 服务过程，其步骤主要如下：

- ① 用户通过定位设备获得位置信息 P_1 。
- ② 生成假位置 D_1 和假位置 D_2 。
- ③ 用户将请求消息 $S(P_1, D_1, D_2)$ 发送给服务提供商。
- ④ 服务提供商对所有位置信息进行查询，并将响应消息 R 发送给用户。
- ⑤ 用户从 R 中选出正确的消息 T 。

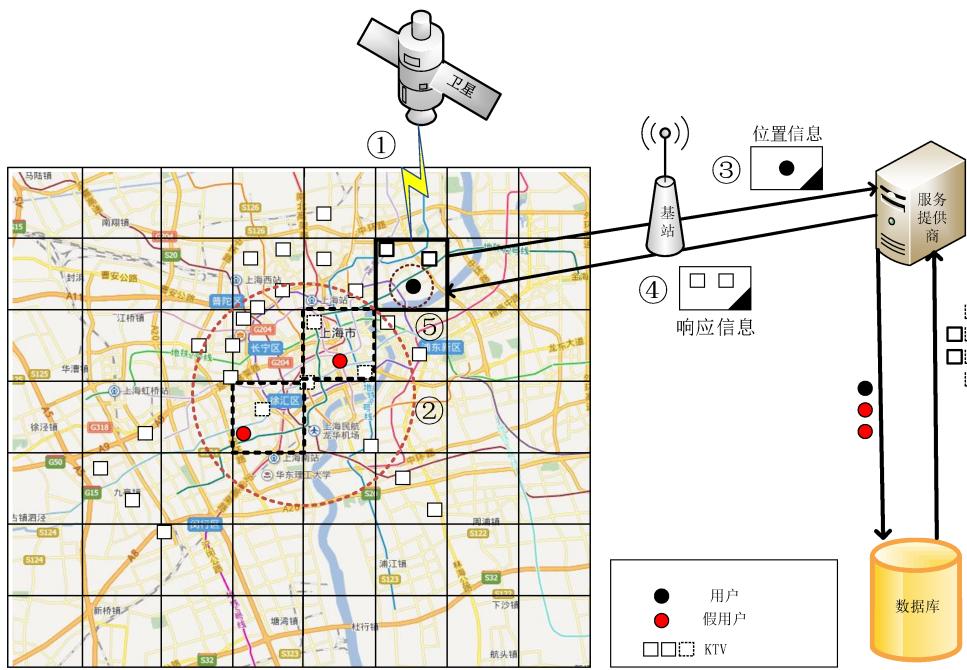


图 2.5: 假位置通信图

在这当中核心部分就是如何生成假位置集，通常情况下考虑位置匿名性的程度从两方面考虑：

普遍性 普遍性是指分布在每个区域中，如图2.6(a)。当所有用户当住在同一个区域中，那服务器很容易列举出个一些户，而当用户分布各个区域后，服务器就很难列举出。因此普遍性可以提高整个区域内的用户位置匿名。

密集性 密集性是指大量用户处在同一个区域，如图2.6(b)。这思想主要来源于 K 匿名，用户在某个区域中发送位置信息给服务器，由于此区域存在大量的用户，服务器很难指定出某个用户，因此密集性提高了某个区域用户位置的匿名。

由于普遍性考虑的是整个区域内用户的位置隐私，而密集型考虑的是局部用户的位置隐私，因此本节将主要围绕普遍性进行叙说。

在 [28] 种，Hidetoshi Kido 提出的 Dummys 的算法中，用户每次讲将自己的真实位置和随机生成的假位置发送给服务提供商，这其中并没有考虑到重复单次查询下的聚合攻击。比如用户通过手机向大众点评请求相关服务，第一次用户查询

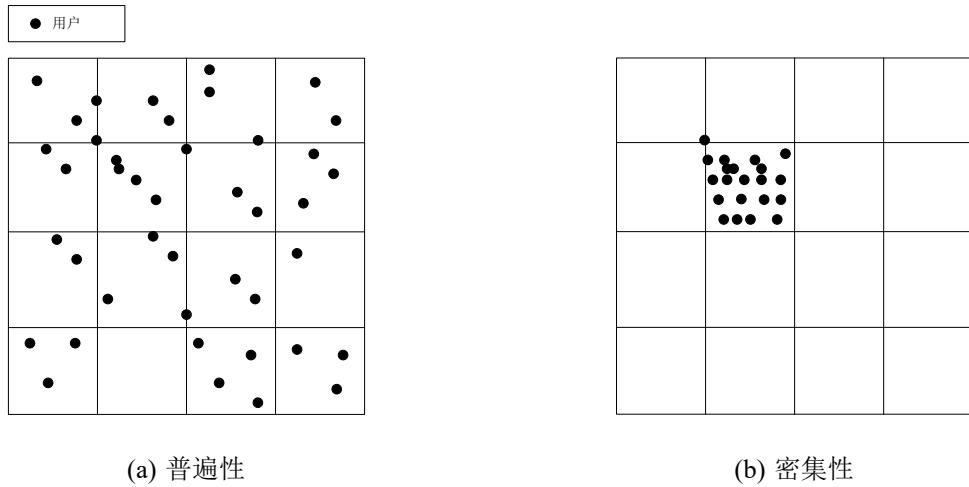


图 2.6:普遍性和密集性示意图

当前位置 P_1 周围的影院，用户将真实位置 P_1 以及两个假位置 D_2, D_3 发送给服务器。第二次用户查询当前位置 P_1 周围的 KTV，同样将 P_1 以及假位置 D_3, D_4 发送给服务器。由于用户每次发送的位置集合中必有一个是用户的真实位置，因此服务器可以通过对两次请求进行求交集，即求 $(P_1, D_1, D_2) \cap (P_1, D_3, D_4)$ ，结果显然为 P_1 ，即为用户的真实位置，因此用户的位置隐私得不到保护。

同样，如果用户每次都使用相同的假位置，而不是进行随机的选择，此方法可以避免用户在相同位置下的聚合攻击，但一旦用户的位置发生变化，那么无法抵抗连续查询下的聚合攻击。例如用户在位置 P_1 时向服务器提出查询，此时将位置集合 (P_1, D_1, D_2) 发送给服务器。当用户的位置发生变化，用户在位置 P_2 时，向服务器提出查询，并将位置集合 (P_2, D_1, D_2) 发送给服务器。服务器通过求后一个状态和前一个状态集合的差集，即 $(P_2, D_1, D_2) - (P_1, D_1, D_2)$ ，结果为 P_2 ，因此服务器便能得出用户当前时空下的位置信息。

针对上述的情况，本节将在假位置的基础上提出一种新的位置匿名算法，此算法即能抵抗用户静止状态下的连续查询聚合攻击，又能抵抗用户在非静止状态下的连续查询聚合攻击。算法主要由一下几部分组成：基于上述描述过程，伪代码如代码1所示：

- ① 假位置生成，用户随机生成一定数量的假位置集合 D 。

- ② 位置获取，用户通过定位设备获取自己当前位置 P_c 。
- ③ 状态获取，获取用户的上一个位置状态信息 P 。
- ④ 状态判断，判断当前位置是否和当前位置存在偏差，若偏差达到一定的阈值 m ，则转至①，获取新的假位置集合 D_N 。
- ⑤ 请求发送，将真实位置和假位置组成的位置集合发送至服务器。
- ⑥ 响应反馈，服务器将响应信息返回至用户。
- ⑦ 信息筛选，用户从响应消息中筛选出正确的信息。

Algorithm 1 生成假位置集合

Input: 用户上一个查询位置 P , 预先设定假位置集合 $D=D_1,D_2,D_3,...,D_k$, 区域 R , 阈值 m

Output: 位置集合 $D_N=(D_{N_1},D_{N_2},D_{N_3},...,D_{N_K})$

```

1:  $D_{N_1} \leftarrow \text{Random}(0, R)$ 
2: for  $i \leftarrow 2$  to  $K$  do
3:    $D_{N_i} \leftarrow \text{Random}(D_{N_{i-1}} - m, D_{N_{i-1}} + m)$ 
4:   if  $D_{N_i}$  equals  $D_{N_{i-1}}$  then
5:      $D_{N_i} \leftarrow \text{Random}(D_{N_i}, D_{N_i} + m)$ 
6:   end if
7: end for
8:  $P_c \leftarrow$  用户获取当前实时位置
9: if  $P_c - P < m$  then
10:   $M \leftarrow (D_1, D_2, D_3, \dots, D_k)$ 
11: else
12:   $M \leftarrow (D_{N_1}, D_{N_2}, D_{N_3}, \dots, D_{N_k})$ 
13: end if
```

2.3.3 区域覆盖技术

区域覆盖技术是位置隐私保护中常见的方法之一 [15][29][30]。该方法的主要思想是将用户精准的位置信息用一个模糊的区域代替，用户在发送请求时，并不将自己的位置发送给服务器，而是所在区域的某一模糊区域发送给服务器。区域覆盖技术将用户隐藏在一定大小的区域内，使得他人无法获得目标的准确位置。

覆盖区域根据实际情况进行选择，一般有圆形覆盖区域和矩形覆盖区域。圆形覆盖区域看起来最为直观和自然 [31][32]，然而将地图划分为圆形区域会产生大量的重叠，而且在计算和表示等方面都没有矩形区域方便。因此目前最为普遍的区域划分法为矩形区域划分，此划分法将地图划分为若干个互补相交的矩形区域，实现对区域更好的粒度控制。另外文献 [33][34] 给出了一些不规则的区域划分，例如结合道路的形状将用户的位置以星形区域进行覆盖。

在其中基于 k -匿名的保护方式最为广泛 [13][15][35][36]，很显然，如果一个区域中包含 k 个移动用户，那么当攻击者得到用户所在的区域时，攻击者无法区分出当前发出请求的为哪一个用户，从而这个区域实现了对移动用户位置隐私的 k -匿名保护。如图2.7，用户 A 使用模糊区域（虚线区域）代替自己当前的精确位置，模糊区域中包含其他两位用户，因此模糊区域实现了 $k=3$ 的匿名保护。 k -匿名亦存在若干的缺点 [37]。为解决这些缺点，出现了 k -匿名的加强版 $l - diversity$ [38], $t - closeness$ [39] 等技术，使得攻击者更难得到用户的信息。一般来说，构造的模糊区域越大，模糊区域中包含的用户就会越多， k -匿名保护中的 k 值也就越大，从而隐私保护程度就越高。但这样会给服务器带来更多的计算和通信开销，造成服务器响应延迟，导致服务器服务质量下降。因此区域覆盖技术实际上是通过消耗服务质量来提高隐私保护程度。如何在服务质量与隐私保护程度之间寻求一个平衡点，一直是国内外学者的研究热点。

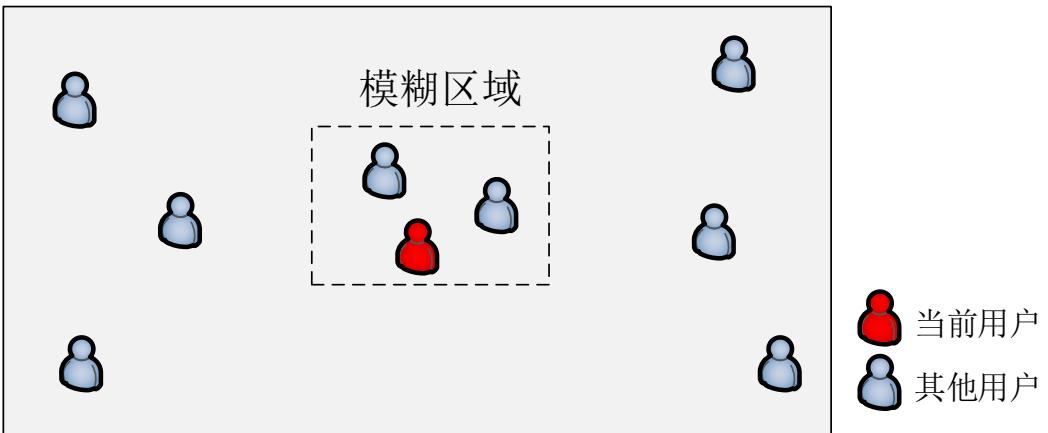


图 2.7: 空间覆盖实例

每个用户对 k -匿名中的 k 要求可能有所不同，文献 [15] 中，Casper 将区域划分为正方形网格，以金字塔的结构对区域进行管理，尽可能的生成用户要求的最小覆盖区域。Casper 假设存在一个可信第三方机构——中心代理，中心代理负责将区域划分为 $L+1$ 层（即金字塔的层数为 $L+1$ ），每一层都是对区域的一种划分方式，自上而下划分粒度越来越细。在第一层，只有一个正方形方格，代表将整个区域作为一块方格，第一层粒度最粗（相当于没有划分）。之后每一层的方格数都是由上一层的方格分裂为 4 个小方格组成的，例如第二层为第一层的 1 个方格分裂出来的 4 个方格，第三层为第二层的 4 个方格分裂出来的 16 个方格，图2.8给出了前三层的划分结构。当可信代理收到用户的请求后，采用自底向上的请求方式，先查看 L 层中用户位置所在的方格，然后查看所在方格中的其他用户数量 $USERSL$ 是否满足用户提出 k -匿名保护中 k 的值，如果满足，则返回 L 层的覆盖区域给用户，否则在同一层查看水平相邻的单元格中用户数量 $USERSL_{RL}$ 和竖直相邻单元格中用户数量 $USERSL_{CL}$ ，计算 $\max\{USERSL + USERSL_{RL}, USERSL + USERSL_{CL}\}$ ，若得到的值满足用户对 k 的要求则返回两个单元格，否则在上一层 $L-1$ 层中进行查找。

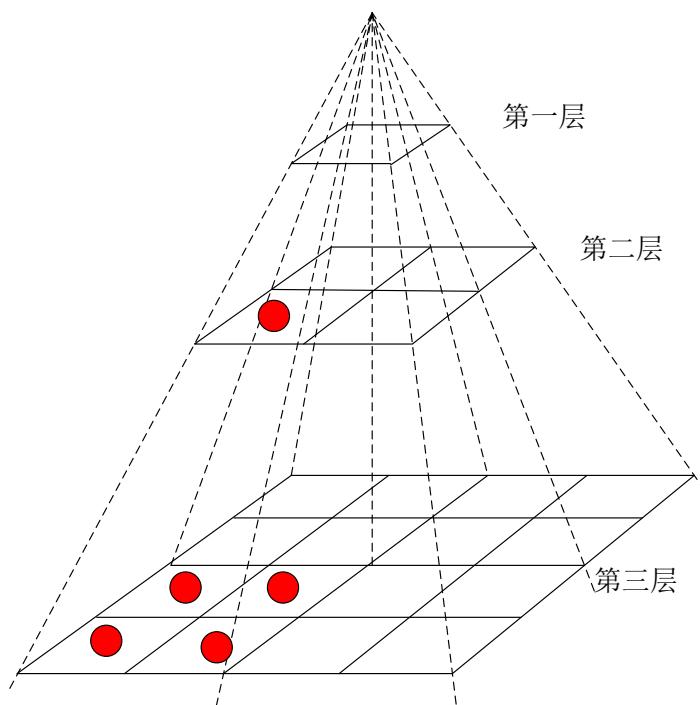


图 2.8: 金字塔划分实例

2.3.4 密码学技术

2.4 本章小结

兴趣点和兴趣区域挖掘 兴趣点和兴趣区域通常作为推荐元素向用户推荐，在不同的挖掘任务中，根据推荐的目标不同采用的方法也不同，聚类是发现轨迹数据特征的最常用方法之一，而对于时空特性明显的地理位置数据，聚类算法的设计、度量方法的选择、数据查询结构等均是该部分的主要研究内容。本文以打车推荐为目的，重点讨论采用基于密度的聚类方法对候选扬招点和热门目的地的挖掘方法。

参考文献

- [1] MOKBEL M F. Privacy in Location-Based Services:State-of-the-Art and Research Direction[C] // Proceedings of the 8th International Conference on Mobile Data Management. 2007.
- [2] XUN Y, RUSSELL P, ELISA B, et al. Practical k nearest neighbor queries with location privacy[C] // 2014 IEEE 30th International Conference on Data Engineering. [S.l.]: IEEE, 2014: 640 – 651.
- [3] 王丽娜. 于移动互联网络的位置服务隐私保护 [R]. [S.l.]: RSA 信息安大会, 2012.
- [4] SWEENEY L. k-anonymity: a model for protecting privacy[J]. International Journal on Uncertainty,Fuzziness and Knowledge-based System, 2002, 10(5): 557 – 570.
- [5] RONGXING L, XIAODONG L, TOM H L, et al. Pseudonym changing at social spots: An effective strategy for location privacy in vanets[J]. IEEE Transactions on Vehicular Technology, 2012, 61(1): 86 – 96.
- [6] GENTRY C. Fully homomorphic encryption using ideal lattices[C] // STOC. 2009 : 169 – 178.
- [7] GOLDWASSER S. Multi party computations: past and present[C] // Proceedings of 16th Annual ACM Symposium on Principles of Distributed Computing. [S.l.] : ACM, 1997.

- [8] DWORK C. Differential privacy[M]. [S.l.] : Automata, languages and programming. Springer Berlin Heidelberg, 2006 : 1 – 12.
- [9] SWEENEY L. Achieving k-anonymity privacy protection using generalization and suppression[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5) : 571 – 588.
- [10] MYLES G, FRIDAY A, DAVIES N. Preserving privacy in environments with location-based applications[J]. IEEE Pervasive Computing, 2003, 2(1) : 56 – 64.
- [11] YOUSSEF M, ATLURI V, ADAM R N. Preserving mobile customer privacy: An access control system for moving objects and customer profiles[C] // Proceedings of the 6th international conference on Mobile data management. [S.l.] : ACM, 2005 : 67 – 76.
- [12] BERESFORD R A, STAJANO F. Location privacy in pervasive computing[J]. IEEE Pervasive computing, 2003, 2(1) : 46 – 55.
- [13] BAMBA B, LIU L, PESTI P. Supporting anonymous location queries in mobile environments with privacygrid[C] // Proceedings of the 17th international conference on World Wide Web. [S.l.] : ACM, 2008 : 237 – 246.
- [14] CHOW Y C, MOKBEL F M, LIU X. A peer-to-peer spatial cloaking algorithm for anonymous location-based service[C] // Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems. [S.l.] : ACM, 2006 : 171 – 178.
- [15] MOKBEL F M, CHOW Y C, AREF W G. The new Casper: query processing for location services without compromising privacy[C] // Proceedings of the 32nd international conference on Very large data bases. VLDB Endowment. 2006 : 763 – 774.

- [16] YIU L M, JENSEN C S, HUANG X. Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services[C] // In Proc. ICDE. 2008 : 366 – 375.
- [17] SHANKAR P, GANAPATHY V, IFTODE L. Privately querying location-based services with SybilQuery[C] // Proceedings of the 11th international conference on Ubiquitous computing. [S.l.] : ACM, 2009 : 31 – 40.
- [18] HU H, XU J, REN C. Processing private queries over untrusted data cloud through privacy homomorphism[C] // Data Engineering (ICDE). [S.l.] : IEEE, 2011 : 601 – 612.
- [19] KHOSHGOZARAN A, SHAHABI C. Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy[R]. [S.l.] : Advances in Spatial and Temporal Databases. Springer Berlin Heidelberg, 2007 : 239 – 257.
- [20] GHINITA G, KALNIS P, KHOSHGOZARAN A. Private queries in location based services: anonymizers are not necessary[C] // Proceedings of the 2008 ACM SIGMOD international conference on Management of data. [S.l.] : ACM, 2008 : 121 – 132.
- [21] GHINITA G, KALNIS P, SKIADOPoulos S. PRIVE: anonymous location-based queries in distributed mobile systems[C] // Proceedings of the 16th international conference on World Wide Web. [S.l.] : ACM, 2007 : 371 – 380.
- [22] WILLIAMS P, SION R. Usable PIR[C] // NDSS. 2008.
- [23] 张浩. 基于位置服务的信息隐私保护技术研究 [M]. [S.l.] : 中国科学技术大学, 2014.
- [24] YINJIE W. Privacy Preserving Data Publishing: Models and Algorithms[M]. [S.l.] : Tsinghua University Press, 2015.

- [25] CHENG R, ZHANG Y, BERTINO E. Preserving user location privacy in mobile data management infrastructures[C] // Privacy Enhancing Technologies. [S.l.] : Springer Berlin Heidelberg, 2006 : 393 – 412.
- [26] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C] // Proceedings of the 1st international conference on Mobile systems, applications and services. [S.l.] : ACM, 2003 : 31 – 42.
- [27] BERESFORD R A, STAJANO F. Mix zones: User privacy in location-aware services[J]. IEEE Pervasive computing, 2004.
- [28] KIDO H, YANAGISAWA Y, SATOH T. An anonymous communication technique using dummies for location-based services[C] // ICPS. [S.l.] : IEEE, 2005 : 88 – 97.
- [29] XU J, TANG X, HU H, et al. Privacy-conscious location-based queries in mobile environments[J]. Parallel and Distributed Systems, IEEE Transactions on, 2010, 21(3) : 313 – 326.
- [30] XU T, CAI Y. Feeling-based location privacy protection for location-based services[C] // Proceedings of the 16th ACM conference on Computer and communications security. 2009 : 348 – 357.
- [31] ARDAGNA C A, CREMONINI M, DAMIANI E, et al. Location privacy protection through obfuscation-based techniques[G] // Data and Applications Security XXI. [S.l.] : Springer, 2007 : 47 – 60.
- [32] KALNIS P, GHINITA G, MOURATIDIS K, et al. Preventing location-based identity inference in anonymous spatial queries[J]. Knowledge and Data Engineering, IEEE Transactions on, 2007, 19(12) : 1719 – 1733.
- [33] WANG T, LIU L. Privacy-aware mobile services over road networks[J]. Proceedings of the VLDB Endowment, 2009, 2(1) : 1042 – 1053.

- [34] HOSSAIN A-A, HOSSAIN A, YOO H-K, et al. H-star: Hilbert-order based star network expansion cloaking algorithm in road networks[C] // Computational Science and Engineering (CSE), 2011 IEEE 14th International Conference on. 2011 : 81–88.
- [35] WANG S, WANG X S. In-device spatial cloaking for mobile user privacy assisted by the cloud[C] //Mobile Data Management (MDM), 2010 Eleventh International Conference on. 2010 : 381–386.
- [36] PAN X, XU J, MENG X. Protecting location privacy against location-dependent attacks in mobile services[J]. Knowledge and Data Engineering, IEEE Transactions on, 2012, 24(8) : 1506–1519.
- [37] JUNCHENG P, HUIMIN D, YINGHUI S, et al. Potential Attacks against k-Anonymity on LBS and Solutions for Defending the Attacks[G] // Advances in Computer Science and its Applications. [S.l.] : Springer, 2014 : 877 – 883.
- [38] MACHANAVAJHALA A, KIFER D, GEHRKE J, et al. l-diversity: Privacy beyond k-anonymity[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2007, 1(1) : 3.
- [39] LI N, LI T, VENKATASUBRAMANIAN S. t-closeness: Privacy beyond k-anonymity and l-diversity[C] // Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on. 2007 : 106 – 115.