分类号 密级

UDC注1



**硕 士 学 位 论 文**

**位置隐私保护在移动定位社交服务中的研究与应用**

（题名和副题名）

**王敏君**

（作者姓名）

**指导教师姓名 李千目 教 授**

**学 位 类 别 工学硕士**

**学 科 名 称 计算机软件工程**

**研 究 方 向 信息安全**

**论文提交时间 2016.01**

注1：注明《国际十进分类法UDC》的类号。

**声 明**

本学位论文是我在导师的指导下取得的研究成果，尽我所知，在本学位论文中，除了加以标注和致谢的部分外，不包含其他人已经发表或公布过的研究成果，也不包含我为获得任何教育机构的学位或学历而使用过的材料。与我一同工作的同事对本学位论文做出的贡献均已在论文中作了明确的说明。

研究生签名： 年 月 日

**学位论文使用授权声明**

南京理工大学有权保存本学位论文的电子和纸质文档，可以借阅或上网公布本学位论文的部分或全部内容，可以向有关部门或机构送交并授权其保存、借阅或上网公布本学位论文的部分或全部内容。对于保密论文，按保密的有关规定和程序处理。

研究生签名： 年 月 日

# 摘 要

近年来，伴随着移动互联网的迅猛发展，移动定位社交服务快速兴起，它在传统的基于位置服务中加入了社交元素，向用户提供了形式多样的基于位置的社交服务，受到了用户的广泛喜爱。然而，在移动定位社交服务的发展过程中也引发了许多争议，其中关于如何使用用户的地理位置信息引起了研究人员的广泛关注。

本文深入研究了基于位置服务中的位置隐私保护方案，并结合移动定位社交服务中位置隐私保护的具体特点，对传统K-匿名模型进行改造，提出了一种新型的K-匿名位置隐私保护模型。同时，该模型针对实时查询匿名保护和位置历史记录匿名保护两种场景，分别提出了实时查询保护算法和位置历史记录保护算法。实验证明，两种算法都能够有效地对不同场景下的位置隐私进行保护。最后本文将保护后的位置信息应用到社交网络的好友推荐服务中，利用数据挖掘技术为用户推荐潜在的好友，实验发现匿名后的位置信息也具有很好的推荐效果，从而也验证了本论文为移动定位社交服务中的位置隐私保护问题提出了有效的解决方案。

**关键字：**基于位置服务，移动定位社交服务，位置隐私保护，K-匿名，好友推荐

# Abstract

In recent years, mobile intelligent terminal headed by Android is growing at a surprising speed. At the same time, short text message as the traditional means of communication got better inheritance and development. However, in the new circumstance, SMS and net-based SMS are facing more risks and challenges. With more incidents about privacy disclosure and viruses appeared to the mobile platform, it is necessary to design a more convenient and practical security mechanisms to protect the SMS system.

Finally, this paper will implement the proposed solutions on the Android platform, and test all the functions. After setting up the experimental environment, this paper compared each proposed algorithm with the traditional method. After using the elliptic curve encryption system, the system will increase the time overhead, but the encryption system can provide better security features; enhanced DEFLATE text compression algorithm is more suitable for compression of short text compared with traditional compressing algorithm. All the solutions proposed can be combined with each other effectively and provide a complete message protection for mobile device.

**Key word:** Location-based Services, Location-based Social Networking Services, Location Privacy Protection, K-Anonymity, Friend Recommendation

# 目 录

[摘 要 I](#_Toc414043019)

[Abstract II](#_Toc414043020)

[1 绪论 1](#_Toc414043022)

[1.1 课题研究背景 1](#_Toc414043023)

[1.2 国内外发展状况 2](#_Toc414043024)

[1.2.1 短信安全研究 2](#_Toc414043025)

[1.2.2 短信压缩技术 4](#_Toc414043026)

[1.3 论文的研究意义 6](#_Toc414043027)

[1.4 论文的主要研究工作 6](#_Toc414043028)

[1.5 论文的组织形式 7](#_Toc414043029)

[2 Android系统的短信安全的理论基础 8](#_Toc414043030)

[2.1 Android操作系统及其短信处理流程 8](#_Toc414043031)

[2.1.1 Android系统构成 8](#_Toc414043032)

[2.1.2 Android应用四大组件 10](#_Toc414043033)

[2.1.3 Android系统短信相关API 11](#_Toc414043034)

[2.1.4 Android短信相关风险 12](#_Toc414043035)

[2.2 对称密码学 13](#_Toc414043036)

[2.2.1 块加密算法 13](#_Toc414043037)

[2.2.2 流加密算法 15](#_Toc414043038)

[2.3 非对称加密技术及其应用 16](#_Toc414043039)

[2.3.1 非对称加密学的基本原理 16](#_Toc414043040)

[2.3.2 非对称加密算法及其安全机制 17](#_Toc414043041)

[2.3.3 非对称加密算法的应用 18](#_Toc414043042)

[2.3.4 非对称加密算法的缺陷 20](#_Toc414043043)

[2.4 本章小结 20](#_Toc414043044)

[3 基于身份验证和数据压缩的短信安全方案研究 21](#_Toc414043045)

[3.1 密钥的分发及管理 21](#_Toc414043046)

[3.1.1 私钥的管理方法 21](#_Toc414043047)

[3.1.2 公钥的管理方法 22](#_Toc414043048)

[3.1.3 基于非对称加密的对称加密算法密钥分配 24](#_Toc414043049)

[3.2 一次一密加密技术 24](#_Toc414043050)

[3.2.1 传统的一次一密加密技术 24](#_Toc414043051)

[3.2.2 基于密钥分发中心的一次一密加密体系 25](#_Toc414043052)

[3.2.3 随机动态口令 26](#_Toc414043053)

[3.2.4 基于时间序列的动态口令生成 27](#_Toc414043054)

[3.3 端到端非对称加密算法比较 28](#_Toc414043055)

[3.3.1 RSA加密算法系统 28](#_Toc414043056)

[3.3.2 椭圆曲线加密函数 29](#_Toc414043057)

[3.3.3 椭圆曲线的原理 29](#_Toc414043058)

[3.3.4 椭圆曲线的优势 30](#_Toc414043059)

[3.4 基于椭圆曲线和时间序列的身份验证协议 31](#_Toc414043060)

[3.4.1 客户端注册过程 32](#_Toc414043061)

[3.4.2 客户端登陆流程 33](#_Toc414043062)

[3.4.3 客户端间通讯验证 33](#_Toc414043063)

[3.4.4 证书更新及吊销 35](#_Toc414043064)

[3.4.5 协议安全性分析 36](#_Toc414043065)

[3.5 数据压缩技术 37](#_Toc414043066)

[3.5.1 LZ77编码理论 37](#_Toc414043067)

[3.5.2 霍夫曼编码理论 38](#_Toc414043068)

[3.5.3 改进的DEFLATE压缩算法 38](#_Toc414043069)

[3.5.4 改进的DEFLATE算法性能分析 40](#_Toc414043070)

[3.6 本章小结 43](#_Toc414043071)

[4 短信安全方案的实现及分析 44](#_Toc414043072)

[4.1 整体设计 44](#_Toc414043073)

[4.1.1 移动客户端平台选择 45](#_Toc414043074)

[4.2 手机客户端设计 45](#_Toc414043075)

[4.2.1 用户登录模块 46](#_Toc414043076)

[4.2.2 短信收发模块 47](#_Toc414043077)

[4.2.3 文本压缩模块和密钥及加解密模块 49](#_Toc414043078)

[4.3 服务器端设计 50](#_Toc414043079)

[4.3.1 公钥注册及更新模块 50](#_Toc414043080)

[4.3.2 公钥查询及分发模块 50](#_Toc414043081)

[4.3.3 基础支撑模块 51](#_Toc414043082)

[4.4 手机端及服务器端公共加密部分 51](#_Toc414043083)

[4.4.1 基于时间序列的动态口令生成算法 51](#_Toc414043084)

[4.4.2 椭圆曲线的加密算法参数选择 52](#_Toc414043085)

[4.4.3 椭圆曲线加密模块实现 53](#_Toc414043086)

[4.5 本章小结 57](#_Toc414043087)

[5 总结与展望 58](#_Toc414043088)

[5.1 工作总结 58](#_Toc414043089)

[5.2 未来展望 59](#_Toc414043090)

[致 谢 60](#_Toc414043091)

[参考文献 61](#_Toc414043092)

[附 录 65](#_Toc414043093)

# 绪论

## 研究背景

近年来，移动互联网快速发展，展现出了巨大的商业价值。而无线电通讯技术的不断进步以及智能手机的广泛普及，也使得移动互联网的发展前景更加光明。因此，越来越多的互联网公司逐渐将公司的业务重心从传统的互联网服务转移到移动互联网方向，其中就包括传统的社交网络服务，它正逐渐向基于位置的移动社交网络服务转变，为用户提供全新的具有社交性质的位置服务，从而吸引更多的用户。这正是本文即将讨论的移动定位社交服务（Location-Based Social Networking Services，LBSNS）。

LBSNS 是将基于位置服务（Location-Based Services，LBS）与传统社交服务（Social Networking Services，SNS）整合起来的一种新型的服务类型，它主要借助智能移动设备，将用户的行为活动和当前的空间地理位置紧密结合，增强社交网络与空间地理位置的关联性，增进用户与客观世界之间的联系，为用户提供以内容分享为核心的移动社交服务。移动定位社交服务的模式多种多样，主要包括有签到模式（Check-In）、地点交友即时通讯模式和基于地理位置的生活分享模式。第一种模式的基本特点是通过积分、勋章等虚拟荣誉激励用户主动 Check-In 来记录其当前所在的地理位置，比较出名的服务有Foursquare、Gowalla等，该模式的服务曾经大受追捧，但近几年已呈现出颓势。第二种模式的特点是在同一时间处于同一地理位置范围内的不同用户之间可以建立社交关系，如国内比较流行的陌陌和腾讯QQ中附近的人等服务。在第三种模式下，用户在发布分享内容的时候可以选择附加上当前事件发生的地理位置，这样不仅可以丰富分享内容的信息，还可以帮助其他用户更好地理解用户所要分享的内容，该模式下的服务有新浪微博和微信朋友圈等，目前正受到国人的大力追捧。这些形式丰富的移动定位社交服务还在不断的吸引更多的用户，LBSNS 已成为了移动互联网环境下传统社交网络服务发展的新趋势。

LBSNS 与传统社交网络服务最主要的区别在于地理位置信息的使用。LBSNS 通过将用户的空间位置信息和人类活动信息紧密的结合起来，从而以此作为基础向用户提供形式多样的基于位置的个性化服务。因此，用户的地理位置信息在LBSNS中扮演至关重要的角色，如何获取和利用地理位置信息就成为 LBSNS 的关键，而无线通信技术和定位技术的快速发展则为此解决了后顾之忧。现如今，智能移动设备可以采取多种方式来获取自身的物理位置信息。在这种前提下，传统网络社交的服务形式正逐渐向 LBSNS 转变。LBSNS 可以通过用户分享的地理位置信息分析出用户的生活偏好与行为模式，为用户提供多样化的个性服务，从而增加用户对服务的使用粘性。现在，常见的 LBSNS 应用包括预测用户的行为模式[1,2]以及好友推荐和位置推荐[3,4]等。

然而，在 LBSNS 不断发展的过程中，关于如何使用和保护用户的地理位置信息引起了广泛关注，越来越多的用户担心他们的位置隐私会在使用服务的时候被泄露，从而导致个人隐私和人身财产安全受到伤害。因为用户的位置信息与其日常生活习惯息息相关，对用户的位置信息历史记录进行分析，可以了解用户的行为轨迹，发现用户的行为模式，甚至可以推断出用户的家庭住址和宗教信仰等敏感信息。但用户想要使用移动定位社交服务，就不可避免要将自己当前的地理位置信息传递给位置服务提供商或其他第三方服务提供商，这就有可能导致个人的位置信息被泄露和滥用。授权访问个人位置信息的服务提供商越多，用户个人隐私泄露的可能性就越大，潜在的风险就越高。因此，保护用户的位置隐私安全显得尤为重要。

但是，就像鱼和熊掌不可兼得一样，用户的位置隐私安全和高质量的移动定位社交服务之间也存在着难以调和的矛盾。一方面，为了保护用户的位置隐私，就必须要将用户精确的位置坐标隐藏起来，以免被泄露；而另一方面，为了提供质量更高的移动定位社交服务，服务提供商就必须要收集用户更精确的位置信息。因此，如何在保护用户位置隐私和提供高质量服务之间找到一个平衡点就成为当前研究的重要内容。

## 研究现状

移动定位社交服务是在基于位置服务的基础上融入社交元素产生的一种新型的服务形式，现有的关于位置隐私保护的研究也大都针对的是基于位置服务。早期的位置隐私保护技术主要有假名技术、假位置技术、空间匿名技术和时空匿名技术。

Beresford[5]第一次提出了位置隐私的概念，并提出了一种经典的假名技术—— Mix Zones 技术来保护用户的位置隐私。假名技术要求用户使用一个临时的假名来代替用户向服务器发送请求，以此将服务请求与用户割裂开来。而在Mix Zones 技术中，要求用户在混合区内不能使用基于位置服务，并且在退出混合区时需要使用另一个假名，以此来保证攻击者无法追踪到用户。

Kido[6]提出了假位置技术。假位置技术要求用户在服务提供商提出查询请求时，除了提供自己的真实地理位置信息，同时也混入一些伪造的随机位置，这样可以保证用户无法从众多的位置信息中发现用户的真实位置。

Chow[7]提出了空间匿名技术，主要思想是在用户请求服务的时候，用一个匿名的位置区域来代替一个精确的地理坐标，这样在该匿名区域中就会出现多个用户而使得攻击者无法识别特定用户。Gruteser[8]则提出了时空匿名算法，它主要是通过服务提供商延长响应时间来实现保护用户位置的目的，因为在延长的时间区域内用户周围可能会产生其他用户，提交的服务请求也更多，对攻击者职别特定的用户造成干扰。

然而，上述这些保护技术并没有对匿名保护的效果有一个很好的度量，研究人员的研究重心开始逐渐转向 K-匿名模型。Sweeney L[9]首先提出了 K-匿名模型（K-anonymity），应用于关系型数据库中隐私数据的保护，它要求同时存在K-1条其他数据与需要保护的数据在敏感属性上的投影值相同，这样就无法将某一条数据对应到具体某一个用户，从而起到保护用户敏感属性的目的。将K-匿名模型引入到位置隐私保护研究中的是Gruteser[8]，他首先提出使用一个包含K个用户的匿名区域来表示这K个用户的位置，在该匿名区域中就无法区分他们彼此之间的位置信息，而匿名位置的保护效果就可以由 K 的大小来决定。在这之后，研究人员针对K-匿名模型在位置隐私保护方面的应用进行了大量研究。Chow[10]提出了 Interval Cloak 匿名算法，该算法主要基于四叉树。它将整个空间递归的划分成4个子空间，直到找到一个区域使得该区域中的用户数量大于等于k。Mokbel[11]在 Interval Cloak 算法的基础上提出了它的改进形式：Casper Cloak。该算法也是基于四叉树，只是在查找匿名空间的时候，它会首先合并邻居节点来查看当前区域用户数量是否满足要求，只有在不满足要求的情况下才会合并到父亲节点。Gedik[12]提出了Clique Cloak匿名算法，利用图论知识，将搜索匿名子集的问题转化成在图中查找 K-点团的问题。该算法中，用户可以根据自身不同的需求自定义 K 的大小，从而达到个性化隐私保护的目的。Kalnis[13]则提出了 Hilbert Cloak 算法，它利用 Hilbert 曲线将平面二维空间转换为一维数值，然后将这些值划分为 K 个用户的不同分组。

除了研究用户单个查询的位置保护外，研究人员还主要对移动物体连续查询的位置隐私保护进行研究，即对用户的轨迹隐私进行保护。在Pan等人[14]提出的方案中，匿名中心利用用户的移动方向和速度生成最小空间匿名区域，从而来保护用户的轨迹隐私。Xu[15]提出使用用户的历史轨迹信息生成匿名集合来实现轨迹 K-匿名，使用历史轨迹可以保证参与匿名的轨迹的真实性。Hoh [16]提出使用路径混淆的方法来保护用户的轨迹。You等人[17]分析现实生活中的客观因素并以此作为参考来生成虚拟轨迹保护用户的轨迹隐私。Kato[18]则主要考虑了用户的行为模式，以更加贴近用户真实生活的行为习惯来生成虚假轨迹。

国内的研究人员也在位置隐私保护方面做了许多工作。谈嵘等人[19] 对移动社交网络服务中的位置隐私保护问题提出了新的见解，并针对可能产生隐私泄露的风险提出了一种策略与算法相结合的隐私保护机制。黄毅等人[20]提出了一种用户协作无匿名区域的隐私保护方案 CoPrivacy。该方案主要通过用户点对点协作构成匿名组，组内用户使用匿名组的密度中心取代自身的地理位置向服务提供商发出查询请求，然后通过增量查询获得近邻查询结果，从而实现K-匿名效果。薛娇等人[21]面向公路网络提出一种隐匿环与森林的位置隐私保护方法，利用宽度优先搜索在图中查找满足要求的环和森林，来模糊移动物体在公路网络中的具体位置，从而保护用户的位置隐私。霍峥等人[22]提出 PrivateCheckIn 方法来保护用户的轨迹信息。该方法设计了缓存机制来缓存用户的签到序列，然后使用签到序列构建前缀树，并通过剪枝和重构生成 K-匿名前缀树，最后遍历该前缀树得到 K-匿名签到序列，实现轨迹 K-匿名保护。

## 研究目标和意义

移动定位社交服务既要求用户分享地理位置，也需要保护用户的位置隐私。一方面，用户在使用位置分享时担心个人隐私因泄露而受到伤害；另一方面，对地理位置过度保护会影响用户体验。基于对该问题的思考，本文将重点讨论如何保护 LBSNS 中的位置信息、以及如何将保护后的位置信息应用于 LBSNS 好友推荐服务中。

研究 LBSNS 中的位置隐私保护具有非常重要的意义，体现在以下两方面：

首先，当前移动互联网正在蓬勃发展，用户数量已经超过传统互联网的用户数量，而移动定位社交服务用户已经达到了一个相当庞大的规模。可以看到，国内 LBSNS 应用如新浪微博、腾讯微信的注册用户早已过亿，而国外 LBSNS 应用如 Twitter、Facebook 更是在全球范围内拥有数量庞大的用户。因此，保障他们的位置隐私安全、为他们提供更好的用户体验已经成为国内外学者迫切需要解决的问题。

另一方面，当前对于位置隐私保护的研究主要针对的是基于位置服务，而移动定位社交服务对于位置隐私保护的需求与基于位置服务并不一致，现有的位置隐私保护模型并不适用于移动定位社交服务中的位置隐私保护。本论文将对此做针对性的研究，保护 LBSNS 中的位置隐私，同时将保护后的位置信息应用到 LBSNS 中的好友推荐服务中。

## 主要工作

本论文基于 K-匿名技术研究移动定位社交服务中的位置隐私保护技术，研究内容主要分为以下方面：

（1）研究基于位置服务的相关基本概念，重点阐明了基于位置服务和移动定位服务的区别；研究现有的位置隐私保护模型体系结构和常用的位置隐私保护技术，为后文研究移动定位服务的位置隐私保护打下基础。

（2）在传统 K-匿名模型的基础上，结合移动定位服务位置隐私保护的特点，对 K-匿名模型进行改造，提出一种新型的K-匿名位置隐私保护模型。同时，该模型针对实时查询匿名保护和位置历史记录匿名保护两种场景，提出两种算法进行保护，并采用实验验证其具有良好的构建效率和保护效果。

（3）研究将保护后的位置信息应用到社交网络的好友推荐服务中，利用数据挖掘技术为用户推荐潜在的好友，实验发现匿名后的数据也有很好的推荐效果。

## 论文框架

本文共分为五个章节，具体内容安排如下：

第一章是绪论，主要介绍本论文的研究背景和研究现状，并说明本文研究的意义以及接下来具体的研究内容。

第二章介绍了相关的基本概念，包括基于位置服务和移动定位社交服务以及服务中的位置隐私保护等，阐述了现有位置隐私保护模型的体系结构，并总结了常用的位置隐私保护技术。

第三章主要描述了新型的 K-匿名位置隐私保护模型，以及该模型针对实时查询匿名保护场景和位置历史记录匿名保护场景提出的两种匿名保护算法。

第四章主要研究如何将保护后的位置信息应用到移动定位社交服务中的好友推荐服务中。

第五章是对全文工作的总结，并对未来的研究进行展望。

本文最后列出主要参考文献以及研究生期间作者的科研成果和参加的主要科研活动。

# 基于位置服务与位置隐私保护的理论基础

## 基本概念

### 基于位置服务

按照维基百科的定义[]，基于位置服务（Location-Based Service, LBS）是通过移动运营商的无线电通讯网络（如GSM网、CDMA网）或外部定位方式（如GPS）获取移动终端用户的地理位置信息（地理座标），在地理信息系统（Geographic Information System，GIS）平台的支持下，为用户提供与地理位置相关服务的一种增值业务。

现如今，基于位置服务在众多行业中己经有了十分广泛的应用。最常见的基于位置服务是用户主动根据所在的地理位置来查询特定目标的位置和相关信息，例如“离我最近的机在哪里”和“告诉我去某家超市的路线”。其次，基于位置服务可以满足商家有针对性投放广告的需求。即商家通过基于位置服务提供商查询到在一定距离范围内的所有用户，并向这些用户发送广告或者电子优惠券以吸引其来消费。第三，基于位置服务在交通运输和物流行业发挥了巨大的作用。很多运输商通过放置在车辆上的移动定位设备来获取车辆的地理位置信息，以此向用户提供车辆监视管理、人员跟踪以及防止车辆被窃等相比之下价格低廉的基于位置服务。第四，基于位置服务也在公共安全行业大显身手。美国推出的法律要求强制性构建一个公共安全网络，并规定了无论用户在何时何地，运营商必須为其提洪精准的定位服务。这种服务能力成为公共紧急救援服务中不可或缺的技术力量，例如在遭受地震或火灾等灾害时，救援人员可以通过基于位置服务迅速精准地定位受灾人员位置。又例如中国移动公司目前推出了专业的儿童老人定位服务，专门应对儿童老人走失或遭绑架等紧急情况的发生。

SNS，即社交网络服务(Social Networking Service)，是指旨在帮助人们建立社会性网络的互联网应用服务。近十年里，伴随着社交网络如火如荼地发展以及移动智能手机的大面积普及，基于位置的移动社交网络的数量开始激增，出现了一种将位置服务（LBS）与社交网络服务（SNS）整合起来的新型服务形式，我们称之为移动定位社交服务（LBSNS，Location-Based Social Networking Service）。LBSNS通过时间序列、行为轨迹和地理位置的信息标记组合，帮助用户与外部世界创建更加广泛和密切的联系，增强社交网络与地理位置的关联性。这种社交网络的特点是将社交活动于用户当前的位置紧密的联系起来，使得社交过程更加有针对性和趣味性。LBSNS用户能够通过在不同地理位置的签到记录和行为反应，与本地企业的商业行动产生对应关系。确立这种对应关系，能够帮助本地企业进一步明确目标客户范围，提高服务的关联度和精准性，从而提高本地服务的质量，产生更大的商业价值。伴随着近年来基于位置的移动社交网络发展规模遍布全球，这个领域快速的成为了基于位置服务提供商和应用开发商的创新阵地和商业战场。例如著名的移动社交网络“Facebook”和“Twitter”以及国内的“新浪微博和“微信”，基于位置的生活分享应用“Pinterest”和“Foursquare”以及国内的“大众点评”和“美团”等等。

### 移动定位技术

显而易见，快速而精准的定位技术是使用基于位置服务的前提条件，也是该服务最重要的保障技术之一。定位技术的概念是指通过移动设备及时确定该设备所在地理位置的技术，它结合了硬件（例如芯片）和软件（例如从多个基站信号中判断所在位置的程序）的技术目前常用的定位技术主要分为四种：基站定位技术、全球定位系统（Global Positioning System，GPS）与辅助全球定位系统（A-GPS）技术、小区识别码（CELL-ID）技术、无线定位技术。

基站定位技术是一种较为成熟的技术，一般应用于移动运营商给移动手机用户进行定位。如图所示，本技术首先计算移动设备与三个或以上基站的距离，在利用三角估算法，就可以推断出移动设备的位置坐标。一般来说，移动设备所测量的基站数目越多，位置计算的准确性越高，行为性能越强。

全球定位系统是由美国军方在二十世纪七十年代研发的，并于九十年代中期开始用于民用领域。相比基站定位技术来说，全球定位系统大大提高了定位的精度：当GPS设备接收到足够多的卫星数据时，其最高精度可达到米。但使用全球定位系统也有很多弊端，比如说从GPS设备冷启动到接收到足够的卫星数据而幵始定位，需要一段较长的预热时间，另外，设备较大的电量消耗也成为其一大弱点。而辅助全球定位系统技术，下称A-GPS，结合了蜂窝基站的GPS卫星信息和自己收到的GPS卫星信息从而缩短了冷启动时所需要的定位时间，并且节省了电量的消耗。总的来说GPS加上A-GPS的技术已经成为基于位置服务中最普遍的定位技术，并被手机制造商广泛采用。

小区识别码（CELL-ID）技术作为定位的一种有效的补充，其主要的优点就是减少定位过程的耗电量，并且支持室内定位。小区识别码定位技术的原理是给蜂窝网络运营商的所有的小区代码都建立一个关联到事先确定好的精确的位置的数据库，一旦移动终端接入了蜂窝网便得到其对应的小区码，就可以通过数据库的查询得到其所在的大概位置。其主要的缺点是定位精度过分依赖于基站的空间密度，在郊区等基站发射塔分布稀疏的地区，其定位精度就变得很差。除此之外，与定位系统无需支付額外位费相比，小识别码技术所需要一些优质的数据库是收费的，例如数据库。

无线定位技术是一种较新的定位技术，它支持室内定位，这一点是GPS系统无法实现的。原因是定位需要一定数量的可视卫星，一般只有在室外才可以正常工作。而无线定位技术是利用无线路由器发出的信号来定位，这一点与小区识别码定位技术类似，不过无线定位技术使用来自无线路由器的信号而不是来自移动运营商的信号。其主要优点是在无线路由信号密集的区域定位时间快、精度高，支持室内定位，而且使用无线热点数据库是免费的，其缺点是耗电量较高，并且要求手机有无线上网功能。

表2.1总结了上述四种定位技术的优缺点。

### 移动定位社交服务体系结构

如图所示，通常來说基于位置服务的体系结构包括四种类型的实体，分别是：定位系统、移动设备（用户）、网络服务提供商以及基于位置服务提供商。下面将对其进行详细介绍：

（1）定位系统

由2.1.2节可知，通常的定位系统包括基站定位系统、卫星定位系统、小区识别码定位系统以及无线定位系统。这些系统负责给移动定位设备提供当前的位置信息以供用户发起基于位置服务请求。本文假设定位系统提供给移动用户的位置信息是及时的、正确的并且是精准的。

（2）移动设备（用户）

在基于位置服务体系中，移动用户通常都持有具有定位功能的移动设备以获取自身的位置信息，同时，移动用户也可以通过蜂窝基站或者无线热点连接入互联网，并且发起基于位置的服务请求。另外，在移动点对点网络这种特殊的位置保护方案中，我们认为这些移动设备还具备另一块无线网卡，可以通过各种无线传输协议或者特定的自组织网络协议（例如协议或者协议自发的组织成一种移动点对点网络，并且互相之间交流信息。

（3）网络服务提供商

网络服务提供商的任务就是为移动用户和基于位置服务提供商之间的信息传输提供了承载网络。同络上传输的信息丨以是加密或者是非加密的，这取决于用户和基于位置服务提供商之间的协议。在互联网上可以部署各种的安全和隐私保护的协议，例如在前文中提到的洋葱路由技术。然而，本文并不强制性的在网络中部署任何额外的安全和隐私保护协议，为了更好的贴近现实网络环境，本文假设在互联网上传输的信息随时可以被恶意攻击者窃听或进行流量分析。

（4）基于位置服务提供商

提供商在网上部署了各种的基于位置服务器，用来为用户提供了各种类型的基于位置的服务。服务器接收到来自移动用户的服务请求，并同时接受用户的位置信息。用户的位置信息可能是精确的空间位置，或者是一块模糊的区域。服务器根据用户的位置信息对服务请求做出了计算，从而生成相应的一个（组）服务返回结果，并通过互联网传输给对应的移动用户。尽管用户和基于位置服务提供商之间可能签订了隐私保护的具体协议，本文仍然假设这类基于位置服务提供商是不可信的，即他们会通过各种手段来挖掘出移动用户的隐私，或者出于商业利益直接将用户的位置信息卖给某些恶意的第三方。

## 位置隐私保护技术

位置隐私保护的主要目的是让用户在使用基于位置服务的时候，既能保证用户享受高质量的服务，同时又让基于位置服务提供商本身不能或者无法轻易的识别出用户的真实的精准位置。我们首先给出用户发送给基于位置服务提供商的服务请求的定义：

定义基于位置服务请求）移动用户发起的基于位置服务的请求可以被抽象定义为如下的三元组：Req = (UserId, Location, Query)

其中UserId代表用户的唯一标识，Location表示的是用户的位置信息，该位置信息可以反应用户精确的位置，例如一个欧式空间中的二维坐标或者系统中的经纬度坐标也可以只是一个模糊的区域位置，例如一个长方形的区域范围；另外 Query代表的是用户服务请求的具体内容，例如“查询最近的五星级酒店”。

由定义可知，要保护用户的位置隐私，可以分别从对UserId和Location这两方面进行保护。目标是让基于位置服务提供商无法识别出用户的，或让其无法从信息中识别出用户真实的位置。下面将详细介绍三种类型的隐私保护技术：（1）假名技术；（2）假位置技术；（3）匿名空间保护技术。

### 假名技术

假名技术是属于对用户的进行保护的一种方法。该方法的主要思路是让用户在发送服务请求的时候采用一个虚假的用户身份来代替真实的用户身份，这样就会扰乱基于位置服务提供商收集的用户身份与位置的对应关系。即使恶意攻击者获得了某个身份信息对应的位置信息，但是由于身份信息是伪造了，所以也不能达到攻击的目的。

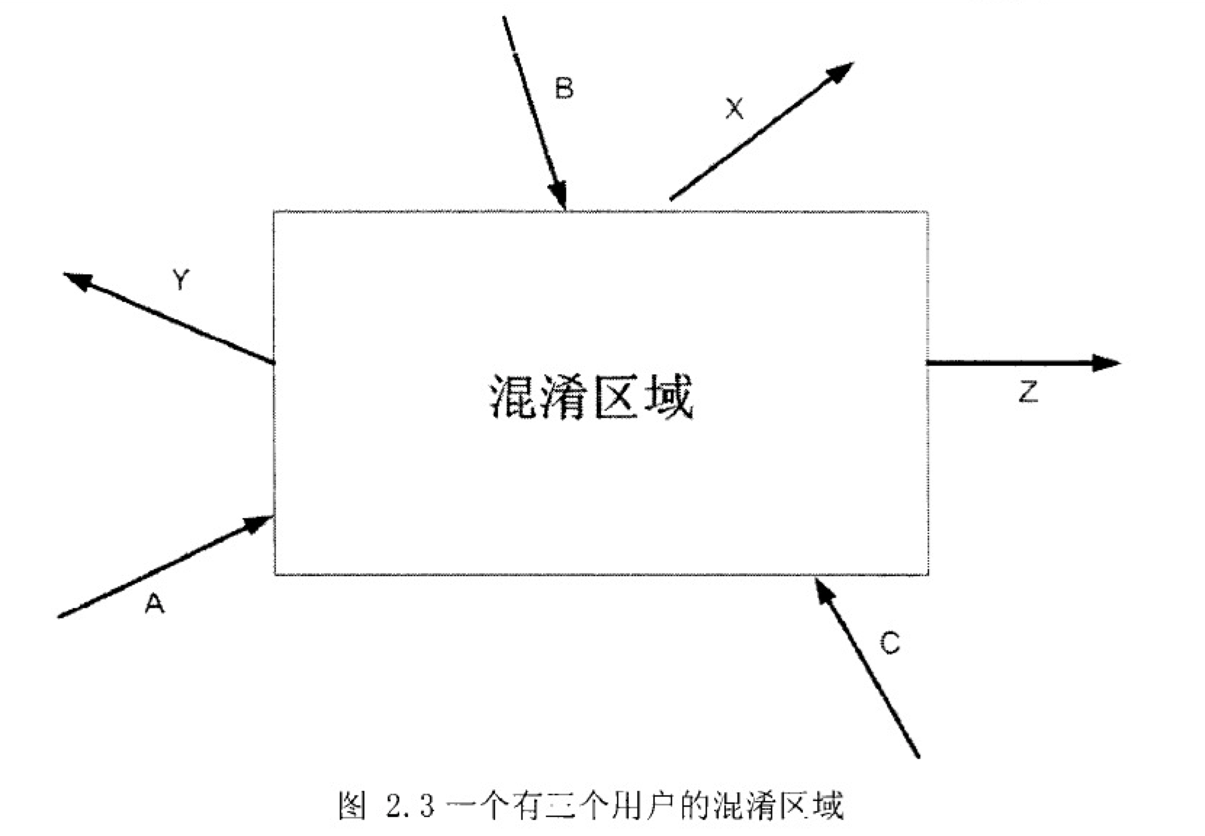
混淆(Mix)的概念很早就被应用在实现网络匿名通讯方案中并同样在位置隐私保护中得以应用。假名技术的代表就是混淆区域技术(Mix-Zone)。该技术把地图适当的划分为两种区域：应用区域与混淆区域。在应用区域，移动用户可以自由正常的使用基于位置服务。但一旦当用户进入一个混淆区域中，任何用户都不允许将自身的位置信息发送给基于位置服务提供商。除此之外，在离开混淆区域前，用户更换自己的用户身份信息，并开始使用一个新的以前没有用过的假名，如图2.3所示。当一个用户从混淆区域中出来的时候，攻击者是无法将其从另外的存在于混淆区域中的用户区分开来，从而实现了混淆区域中的用户的k-匿名保护，即攻击者无法把目标用户与其他k-1个用户区分开来。由于用户使用的假名每经过一个混淆区域都会更换，这也大大增加了用户隐私的保护程度。

表2.2给出了针对图2.3的一个拥有三个用户的混滑区域的实例。我们可知有三个用户（真实身份分别是甲、乙、丙）分别在第2、5、8个时间点进入了混滑区域，所使用的假名分别是A、B和C。进入了混靖区域后，三个用户都不在发布其地址信息，同时更换了假名。用户甲在第7个时间点以假名Z的身份出了混清区域，用户乙在第8个时间点以假Y的身份出了混淆区域，用户丙在第12个时间点以假名X的身份出了混清区域。因为攻击者无法预测三个用户在混淆区域里面停留了多长时间，所以要将用户的身份重新关联起来的难度很大，尤其是当混靖区域中的用户数量较多的时候。

然而混靖区域技术的最大的问题就是限制了用户在混淆区域内的基于位置服务的可用性，因为所有的用户在混淆区域内时都不允许使用该服务，这对于一些需要持续性提供用户地理位置信息的基于位置服务来说是不可接受的，会严重影响服务的质量，例如在线导航系统。另外，我们发现混清区域设置的位置和大小都会影响隐私保护的能力，对所有用户统一性的设置混滑区域显然是无法满足不同用户的个性化的隐私保护需求的，这也成为用户使用混淆区域技术需要考虑的另一个实际应用问题。

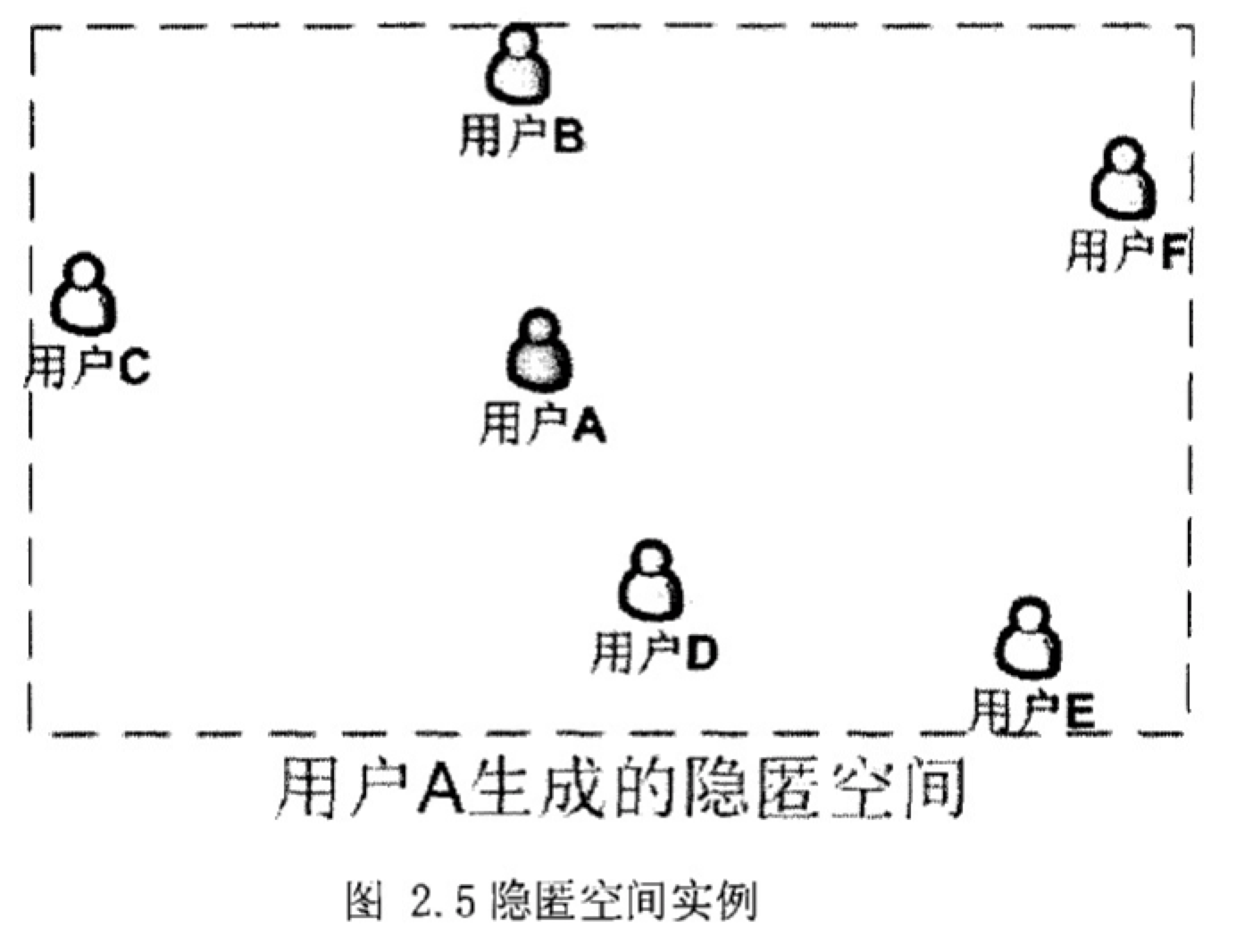
### 假位置技术

假位置技术是属于对用户的Location进行保护的一种方法。顾名思义，假位置技术是一种允许用户使用假的或者错误的位置来代替自己的真实位置来发送服务请求的技术。在这类方法中，用户除了向服务器传输自己真实的地址外，还同时以一定策略生成一组假位置信息作为掩护同时发送出去。例如使用分布在用户周围的一些人口密集的地点作为假地址。而这些真假信息对于基于位置服务提供商来说是完全无法区分的，所以即使攻击者窃听到了用户传输的所有位置信息，但仍然难以在众多的假地址中辨别出用户的真实地址。图是这个方法的一个实例。由图可见，用户额外发送了七个假位置信息作为其真实位置的掩护，这样就实现了k=8的匿名保护，即攻击者无法在八个位置中识别哪一个是用户的真正位置。

地标技术(Landmark)是假位置技术的另一种形式，其方法是让用户采用一个标志性的地理位置来代替其真实的位置传递给基于位置服务提供商，而基于位置服务提供商是通过这个标志性的地理位置来处理服务请求。由于攻击者只能推测出用户在标志性地理位置附近出现，而无法确定其精确的位置，因此用户的位置隐私得以受到保护。当地标位置离用户真实位置越远的时候，用户的位置隐私会有更好的保护效果，但是如果离得过于遥远，基于位置服务的质量也会严重的下降。

假位置技术的优点是技术简单而便于实现，适合在隐私保护要求不高的情况下使用。其缺点就是大量的假位置会大大增加位置服务器的计算开销及服务请求等待时间，并且传输大量的无用的服务应答数据也会耗费有限的移动设备带宽。另外，在使用地标技术时，基于位置服务的质量也会相应的受到影响。

### 空间匿名技术

匿名空间保护技术是属于对用户的进行保护的一种方法。该方法的主要思想是使用一个精心生成的模糊区域来代替用户真实的精确位置，从而将用户隐藏在一定面积大小的区域里。很显然，如果这个区域同时包含了数量为的移动用户，那么这个区域理论上便实现了对用户位置隐私的匿名保护。如图所示，用户A生成的隐匿空间（虚线框）中包含了五个其他的用户，这样就实现了k=6的匿名保护。一般来说，隐匿空间的面积越大，所包含的用户数量就越多，用户位置隐私的k-匿名保护程度就越高，但其副作用是给服务器带来更多的资源幵销，给网络传输带来更多的压力，从而导致服务的质量变得更差。所以，隐匿空间保护技术的实质是通过牺牲基于位置服务的质量来换取用户的位置隐私保护。

近些年来，学术界有众多的学者在隐匿空间保护技术方向上做出了大量的工作，从体系结构的角度来讲，隐匿空间技术主要有两种实现方式：（1）基于可信第三方的体系结构；（2）基于移动点对点的体系结构。

#### 基于可信第三方的体系结构

基于可信第三方的体系结构引入了一个可信第三方作为移动用户和基于位置服务提供商之间的中间人。这个可信第三方也叫做可信匿名位置服务器，它可以是中心服务器结构的或者是分布式的。它的作用主要有以下三点：（1）收集移动用户的位置服务请求，并将其中用户的位置信息记录下来；（2）根据用户个性化的匿名保护需求，结合已保存的大量用户的位置信息，釆用不同的隐匿空间生成算法将用户的精碗位置泛化为个区域；（3）代替用户向位置服务商发出服务请求，并返回给用户响应的服务结果。该体系下的代表技术是由Mokbel等人提出的时空匿名保护技术：New Casper算法，以及Ghinita等人提出的时空加密保护技术：PRIVE算法等。

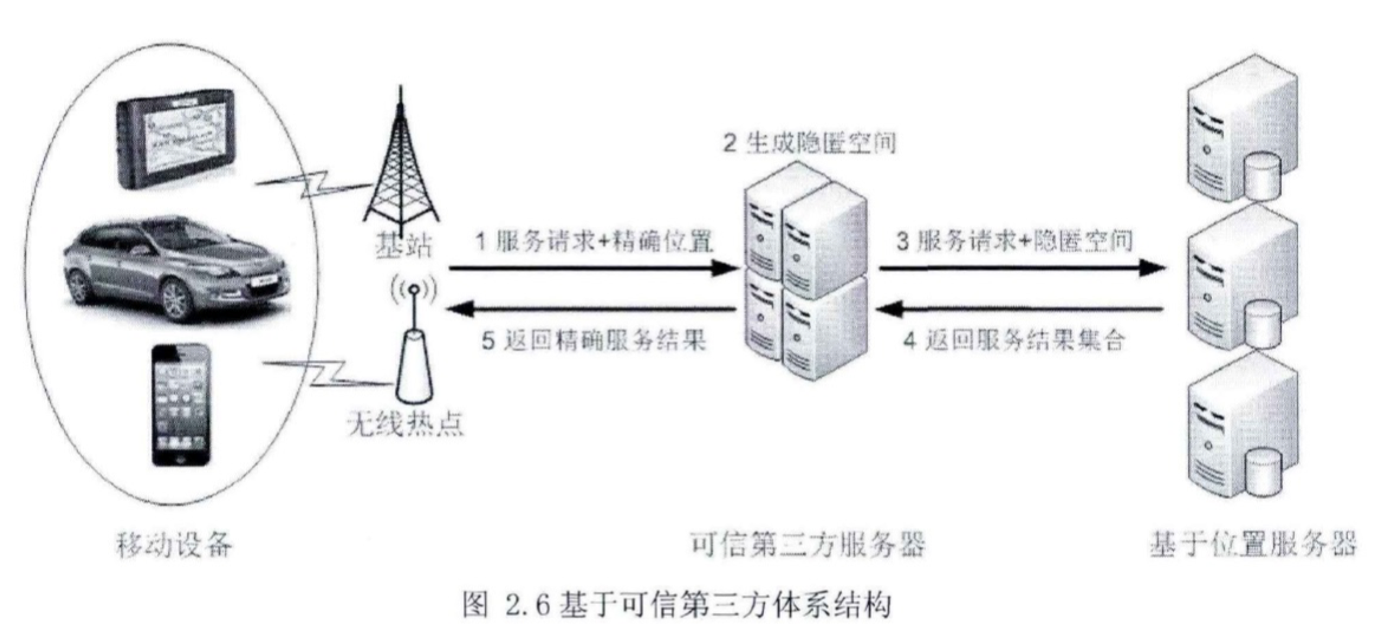
如图，一般来说，该体系结构下一次成功的服务请求过程如下：

（1）移动用户通过不同的移动设备向可信第三方服务器发出基于位置服务请求，该请求信息中包括了用户本身的精确位置信息。

（2）可信第三方服务器将用户的精确位置信息进行匿名化处理，生成一个隐匿空间。

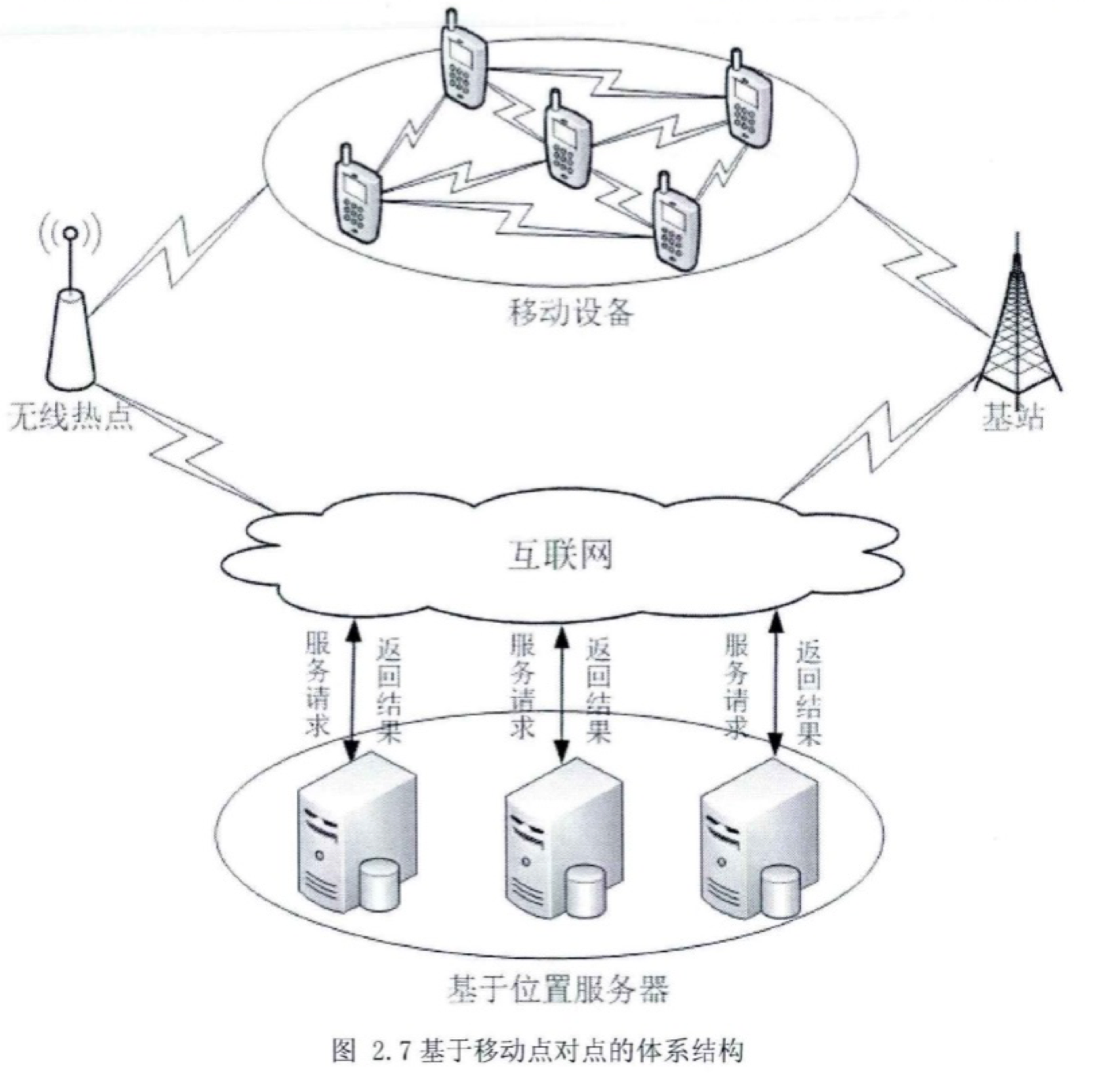
（3）可信第三方服务器将用户的服务请求中的精确位置替换为隐匿空间并发送到对应的基于位置服务器。

（4）基于位置服务器针对收到的隐匿空间生成了一组服务结果的集合，并将其返回给可信第三方服务器。

（5）可信第三方服务器根据用户的精确位置从服务结果集合中挑选出最精确的结果，返回给移动用户。

这种体系结构好处是将本来应该由用户设备完成的隐匿空间计算任务交给了第三方来完成，减少了用户对其设备计算能力和电量方面的高要求。但该体系的缺点也同样十分明显：（1）当大量的用户同时发起服务请求时，可信第三方服务器将成为整个系统计算和通讯的瓶颈。（2）可信第三方是整个系统的关键，这导致对其服务器的可用性要求是非常高的。（3）用户必须无条件信任这个第三方，当第三方被恶意攻击者入侵时会有严重的隐私泄露风险。

#### 基于移动点对点的体系结构

基于移动点对点体系结构没有依赖于任何第三方服务器，移动用户之间是平等的合作关系，用户通过点对点通讯技术自发的组织成一个移动点对点网络，用户之间可以互相分享位置信息。一旦某一用户准备发起位置服务请求，它会根据已收集到的其他用户位置信息生成一个隐匿空间来代替自身的精确位置，并发送给基于位置服务提供商。图2.7描述了基于移动点对点的体系结构。该体系下的代表性算法是由Chow等人提出的，该算法有两个模式，分别是On-demand（按需）模式和Proactive（主动）模式。

该体系的优点是不需要一个可信第三方，也就同时解决了第三方作为系统瓶颈的问题，也无需担心自己的位置信息被第三方泄露的风险。其缺点是对用户的移动设备的计算能力要求较高，网络通讯量较大。同时，由于移动设备收集位置信息的数量和质量的限制，生成隐匿空间的算法效率会相对较低，而且生成的隐匿空间的位置隐私保护能力相对也较差。

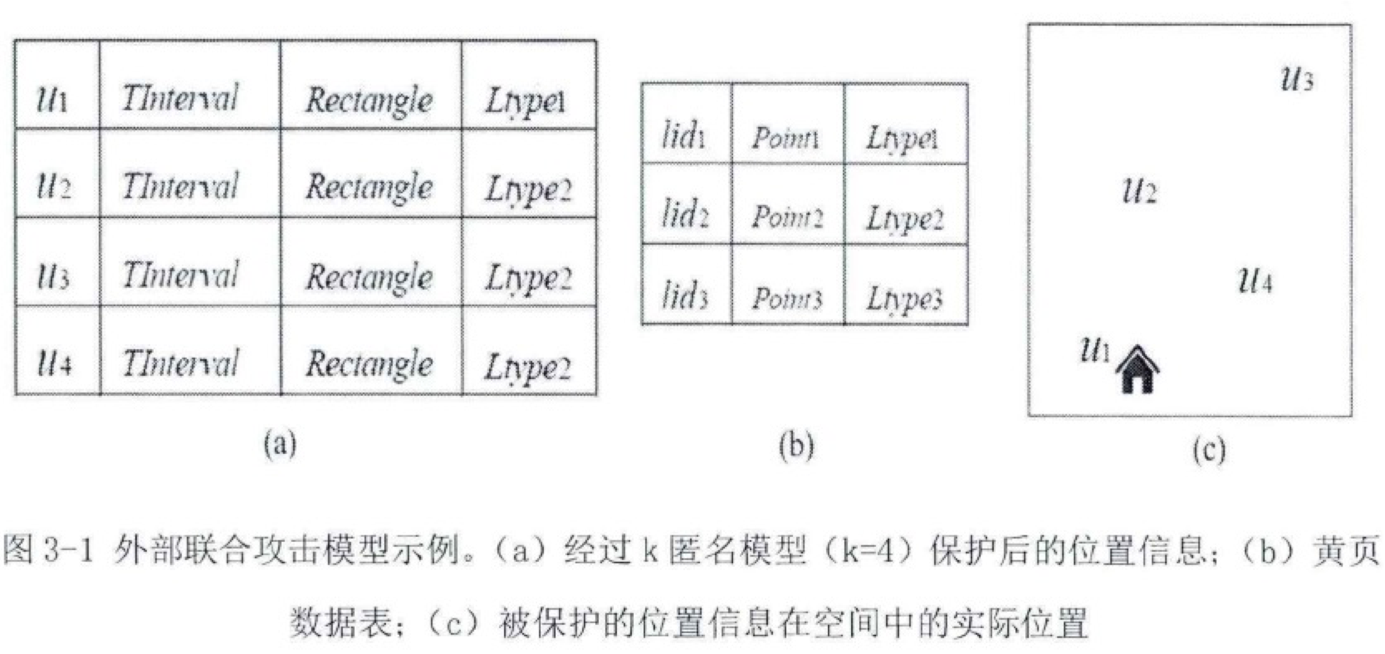
## 位置隐私攻击模型

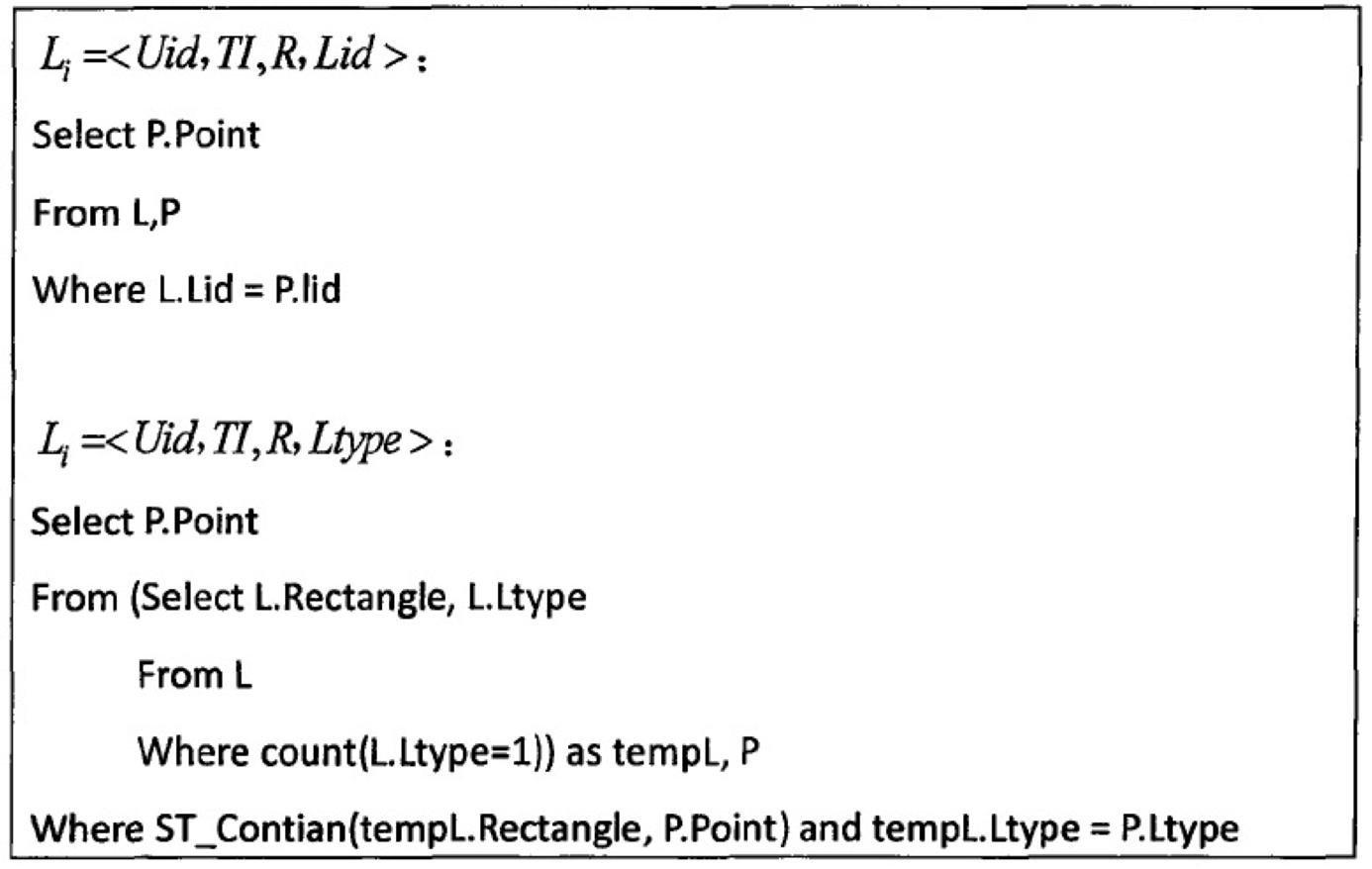
在LBSNS中，一条位置信息可以表示为以下形式：<uid,t,<x,y>,lid>

其中，uid是用户的标识号，t是时间戳，<x,y>是经纬度坐标，lid是该位置点(兴趣点)的标识号，通过lid可以获取到该位置信息所对应地点的名字、地址和类型等其它信息。由于典型的位置信息包含了用户的标识和其时空信息，因而很容易将用户的隐私泄露出去。

传统的LBS攻击模型是指对通过发现位置隐私保护模型的内在缺陷，从而挖掘出用户身份标识，使得相关保护模型失效的攻击方法的一种统称。但与LBS攻击模型不同的是，LBSNS攻击模型的目的是获取用户的敏感位置，而非用户的身份标识。本节提出了两种针对LBSNS的敏感位置攻击模型，分别是外部联合攻击模（External Joint Attack, EJA）和位置重叠攻击（Location Overlapping Attack, LOA）。

### 外部联合攻击模型

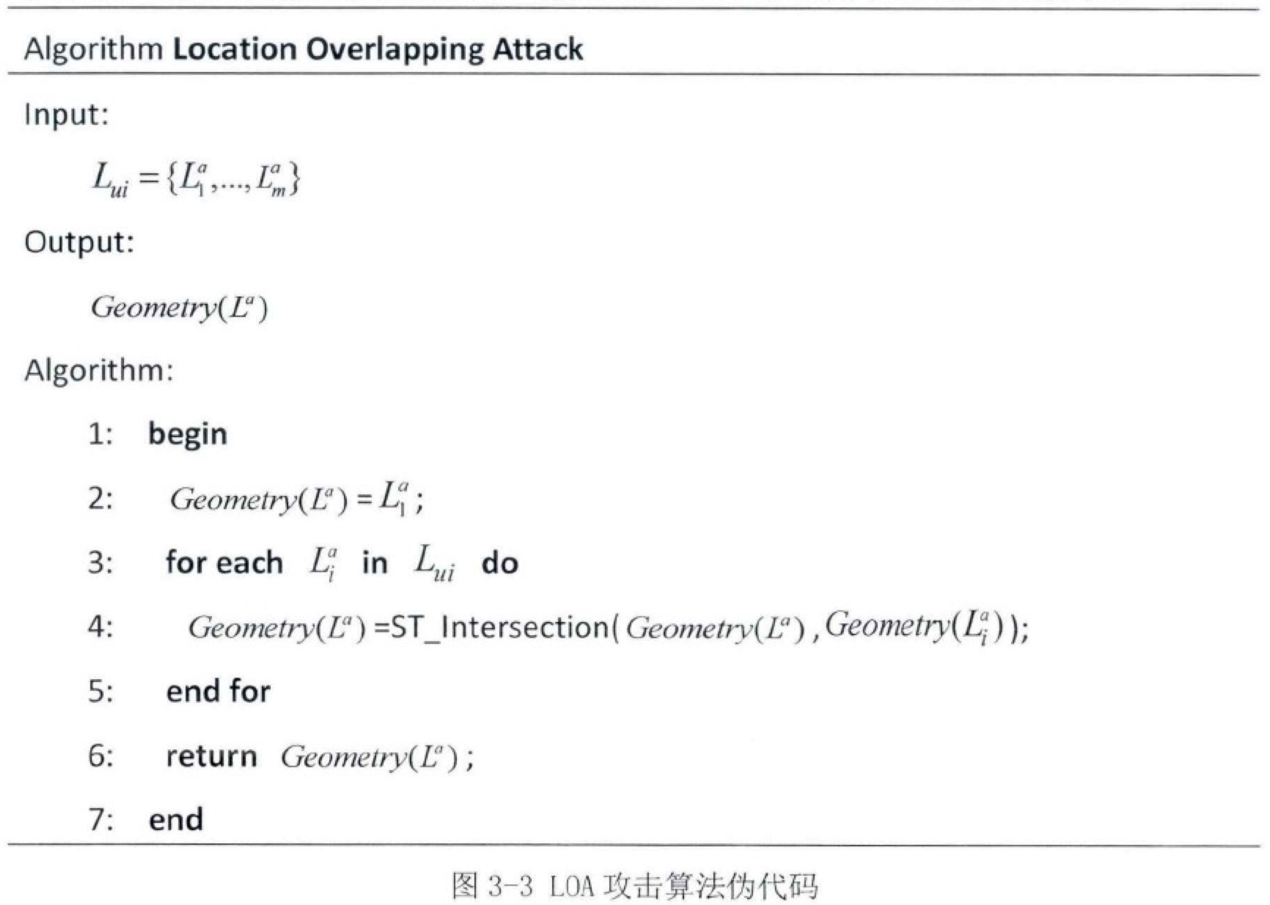
外部联合攻击模型是指通过将被保护后的位置信息记录中的兴趣点标识或者兴趣点类型和相关的黄页数据表进行连接操作，以此获取用户敏感位置的攻击方式。以下图3-1为例，图3-1 (a)是经过k匿名模型(k=4)保护后的位置信息记录，表中每列依次表示用户标识、时间间隔、模糊区域和位置类型。其中，位置类型是指原始位置信息中相应兴趣点的类型,例如“饭店”和“家”等。从直观角度上看，图3-1 (a)己经满足了位置隐私保护的要求，攻击者无法通过该记录直接发现用户的敏感位置。但是，如果攻击者掌握了黄页数据表，如图3-1(b)所示，那么一旦当模糊区域中的位置类型并不丰富，甚至某个类型仅存在一处时，那么用户的敏感位置就有可能被发现。图3-l(c)显示了在模糊区域中，仅包含一处位置类型是Ltype1的地点，则在这种情况下，位置隐私保护模型已经失效。

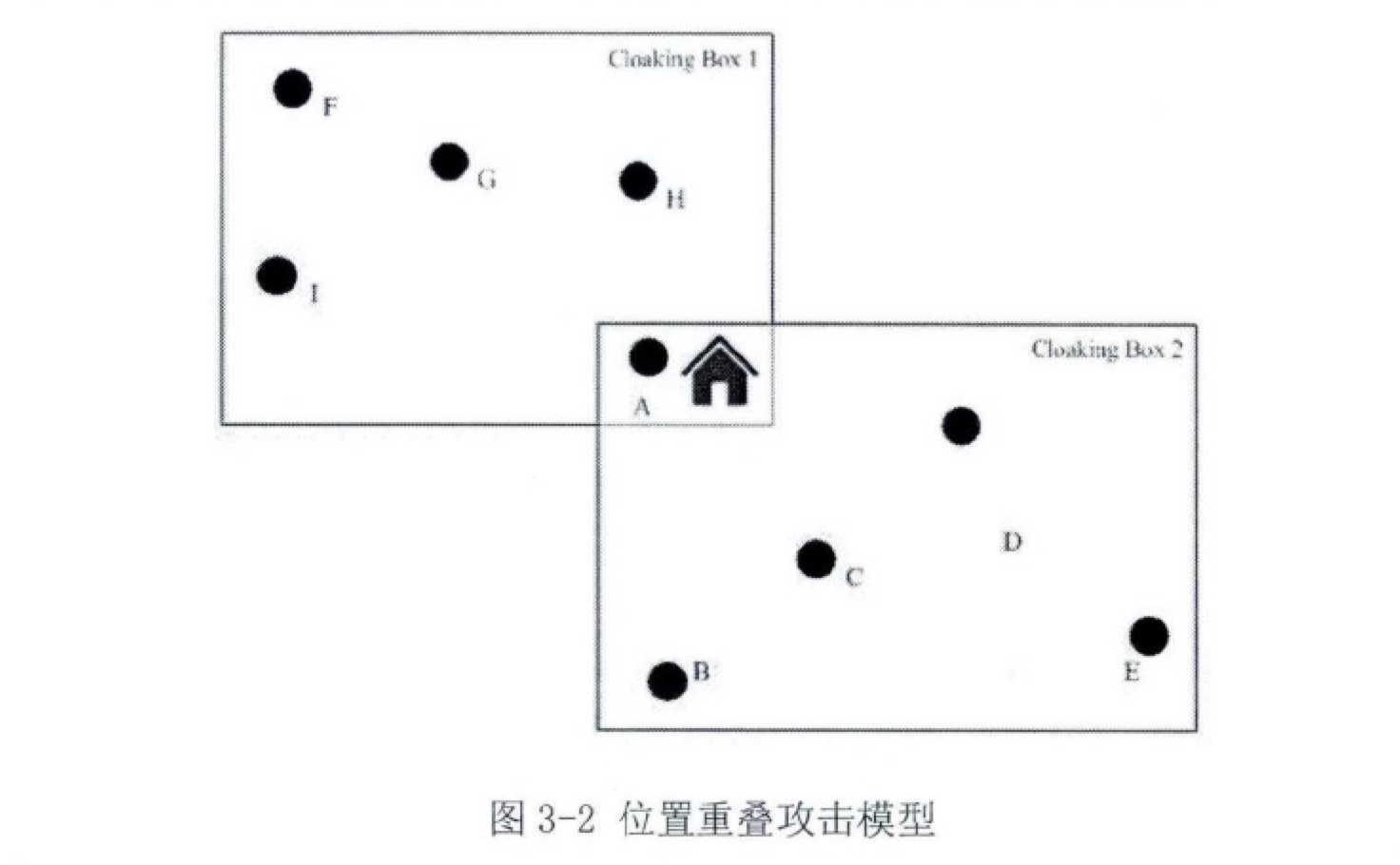
假设Li=<Uid,TI,R,Lid>或Li=<Uid,TI,R,Ltype>表示一条被K匿名位置模型保护后的LBSNS位置信息，L={L0,…,Ln}表示被保护的位置信息的集合，P={P0,…,Pm}表示黄页兴趣点的集合，ST\_Contain(A,B)表示空间对象A包含空间对象B，那么外部联合攻击模型的结果可以用以下时空数据库查询得到：

为了防御外部联合攻击模型，可以将兴趣点标识或者类型从被保护的位置信息中删除，但是这么做的缺点是丢失了位置信息的语义，而使得LBSNS无法发现用户的偏好，降低了推荐服务的质量，并可能使用户的使用体验得不到满足。一种更好的方法是对兴趣点标识或者类型也进行一定程度的模糊，虽然降低了它的信息粒度，但是由此也保留了发现用户偏好的可能。

### 位置重叠攻击模型

传统K匿名位置模型将用户的每次查询都看作是一个孤立的事件，即每次用户进行位置更新，它都会新计算位置信息所对应的模糊区域。换言之，对于两个具有同一位置、不同时间戳的位置信息，它们模糊后的区域可能是不同的。当这些被保护的位置信息被存储时，这种做法就有可能造成用户敏感位置的泄漏。以图3-2为例，用户在不同时间对敏感位置A进行了两次访问，并都与其它4个位置信息进行了模糊处理。如果将两次模糊的结果分开看,那么每一次都能使得用户的位置隐私得到保护。但是当将这两次模糊后得到的模糊区域在空间上进行重叠时，可以发现用户的敏感位置A就可以被限定在一个极小的范围内，由此就有可能泄露了用户的敏感位置隐私。随着用户对同一位置的访问次数的增多，使用位置重叠攻击模型得到用户该敏感位置的可能性也在不停增大。而在极端情况下，用户的敏感位置甚至只需要通过一次空间重叠就能够被发现。很显然,传统的K匿名位置模型无法很好地对LBSNS的位置历史进行保护。

假设Lui={L1a,…, Lma }表示用户ui在敏感位置a所有被保护的位置信息的集合，Geometry(Li)表示位置信息 Li的空间对象，它可以是一个点（Point）或者一个多边形区域（Polygon），ST\_Intersection(A,B)表示空间对象A与空间对象B在空间上的交集，那么位置重叠模型的攻击算法伪代码可以表示如下图3-3所示：



## 本章小结

本章中，我们首先定义了移动定位社交服务的概念，描述了该服务在不同领域的应用现状，总结了现在移动设备的定位技术，并介绍了该服务的基本体系结构。接着，我们对位置隐私保护技术按照三类不同的技术手段进行了详细综述，其中对空间匿名保护技术的两种体系结构进行了详细的介绍。最后，我们总结了移动定位社交服务中针对位置隐私保护的两种攻击模型。

# 基于移动社交定位服务的位置隐私保护模型

位置信息在中具有非常重要的作用，除了支持实时服务外，通过离线分析用户的位置历史，不仅能够标识用户的兴趣点，预测用户的移动行为，而且也可以发现新的社群，为用户推荐新的社交关系。作为位置服务的一个子类，LBSNS 对位置信息隐私保护的需求与传统的 LBS不同，主要体现在以下三方面：

1. 从保护目的上来看，传统 LBS 位置隐私保护的目的是为了防止攻击者获取用户的唯一标识，使得攻击者即使知道了位置信息也无法知道是哪个用户在使用服务。而 LBSNS 位置隐私保护并不是将用户的标识与地理位置信息完全剥离，因为一旦保护过度就无法为用户提供精确的个性化服务。LBSNS位置隐私保护的目的是为了防止用户的敏感隐私地点被攻击者掌握。简而言之，LBS 位置隐私保护是对用户标识匿名，而LBSNS位置隐私保护是对敏感隐私地点匿名。

2. 从使用角度上来看，位置信息不仅作为 LBSNS 实时查询服务的基础，而且通过分析记录的位置历史能够在只是知识发现方面为 LBSNS 的个性化服务提供支持。而传统的 LBS 隐私保护主要从实时查询角度出发，将每次对位置信息的匿名割裂成一个孤立的事件，如果使用传统的LBS 位置匿名方式来存储位置历史记录，这可能导致用户的敏感位置信息被泄露。

3. 从位置信息表现形式上来看，传统 LBS 位置信息仅包括用户标识、经纬度坐标和时间戳三个属性，而在 LBSNS 中记录的位置信息还应该包括该位置对应的兴趣点标识属性。从该标识属性可以知道兴趣点对应的名称、类型和标签，用于知识发现。因此，在 LBSNS 位置隐私保护中，不仅需要在时间和空间属性上匿名，还需要对信息记录中的兴趣点标识进行匿名。

基于上述需求，本文提出了一种增强的 K-匿名位置隐私保护模型。该模型从位置信息的时间和空间两个角度出发，寻找与目标地理信息最接近的其他 K-1条地理位置信息，通过泛化来降低信息粒度，同时对位置信息中的兴趣点标识属性进行处理，从而使它们在时间属性、空间属性和兴趣点标识属性上都彼此无法区分，最终实现用户位置隐私保护的目的。同时，该模型针对位置历史记录的匿名位置保护，提出了位置历史记录匿名算法

## 增强的 K-匿名位置隐私保护模型

对于LBSNS位置隐私的保护，特别是离线的位置历史的保护，它应该满足以下要求：

1、对位置信息的时间、空间以及语义标记都需要进行保护。在时间和空间维度上进行保护，可以使恶意攻击者无法轻易发现用户敏感位置的确切所在。而对语义标记也进行保护，则可以防范攻击者通过EJA模型进行攻击，从而获取敏感位置。

2、对于同一用户在相同敏感位置不同时间所更新位置信息，在空间和语义标记维度上都应该尽量保持一致，由此才能确保敏感位置不能通过LOA模型获取到。换言之，无论是对静态的已经存储的位置历史，还是对动态的即将更新的位置信息，都需要提出相应的解决方法。

为了满足以上要求，本文提出了一个增强的 K-匿名位置隐私保护模型。该模型对时空K-匿名模型进行了扩展，它以K-匿名模型为基础，但是它不严格要求每个位置信息都必须和其它至少K-1一个位置信息相区分，只是尽量保证匿名的条件。这主要是因为在实际应用中，严格的K-匿名条件经常难以达到。而另一方面，为了防范LOA模型的攻击，保证相同用户的具有同一位置坐标的位置信息对象在空间和语义标记维度上相一致，K匿名的条件也可能在建模过程中失效。除了对位置信息采用了灵活的匿名条件和模糊保护方式外，模型在匿名过程中，将被保护的位置信息与其它干扰位置信息在时间和空间维度上进行了相似度比较，从而得到更小的时空匿名框，提高了匿名服务的质量。

### 相关定义

定义 **K-匿名模型** 给定一个包含若干数据对象的数据表，表示数据表中定义的一个属性。如果存在一个属性集合，使得数据表中的任意一个记录，都能在属性维度上至少找到K-1条其它记录与它无法相互区分，那么则称数据表在属性维度上满足K匿名。

定义 **泛化** 为了实现 K-匿名，需要对记录表中的原始属性的值进行修改，使得它们彼此无法区分。最常用的方法是降低原始属性值的信息粒度，以时间戳属性为例，为了满足 K-匿名，需要将这 K 条记录的时间戳属性值修改为一个覆盖这 K 条记录时间戳的时间间隔对象，我们将这种方法称为泛化。

### 增强的 K-匿名位置隐私保护模型

定义 **增强的 K-匿名位置隐私保护模型** 给定一个记录位置信息的集合，对于其中任意一条位置信息，它都能通过泛化操作成 的形式，其中是用户的唯一标识，是一个时间间隔对象，表示一个覆盖地理信息位置坐标的空间矩形区域，是泛化后兴趣点标识的集合。假设集合可以被划分为若干个互不相交的子集，并且对于其中的任意一个子集，它都能满足以下3个条件：



1.集合中元素的个数，并且。



2.若集合中元素个数，那么集合中任意两个元素的用户标识都不相同。



3.若集合中元素个数，那么集合中任意一个元素都无法在时间属性、空间属性和兴趣点标识属性上和其它元素区分。



那么则称是满足增强的时空K-匿名位置隐私保护模型的集合，是集合中的一个匿名子集。



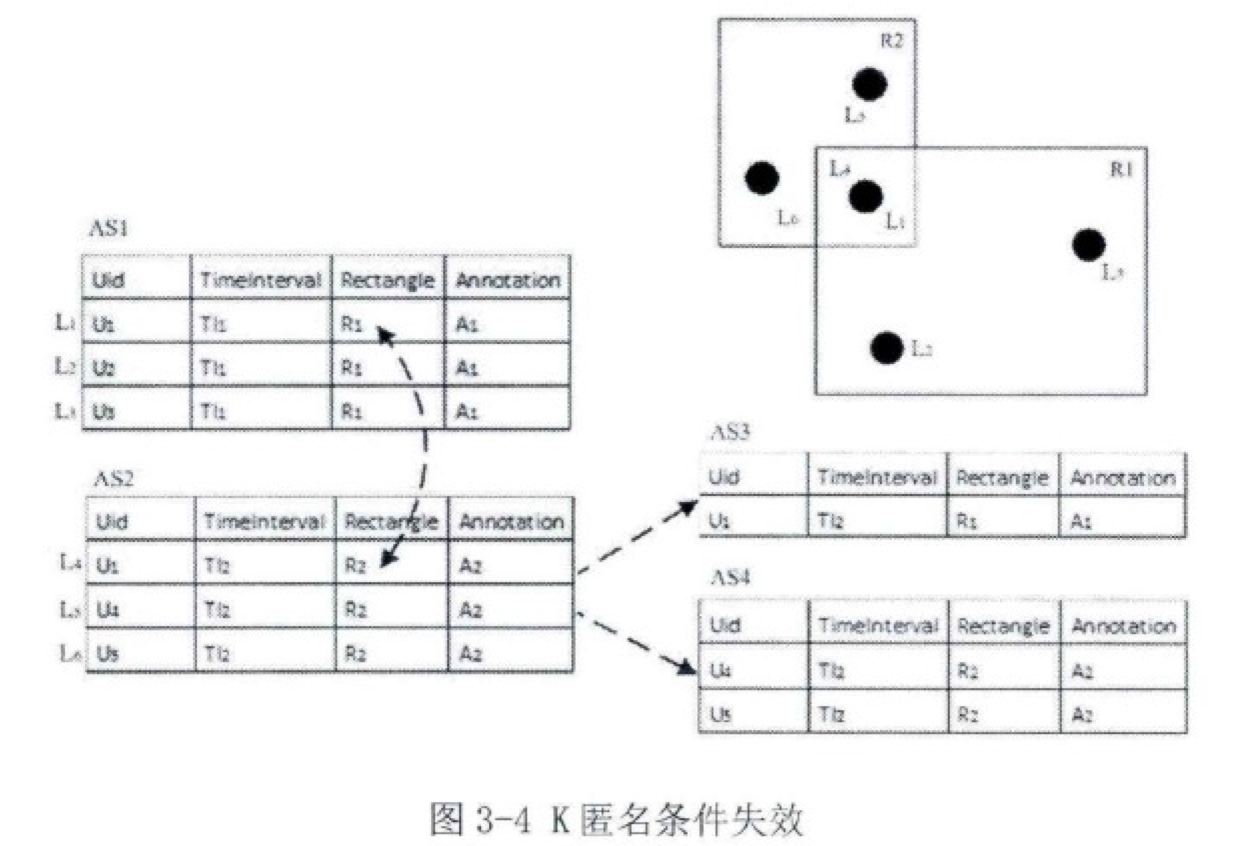
本文提出的匿名位置隐私保护模型是对传统K匿名模型的扩展，以它为基础，但对 K 的要求没有传统的 K-匿名模型那么严格，只需尽量满足K匿名的条件。因为需要保证同一用户位于同一敏感位置时在进行泛化操作后，空间属性能够保持一致，这样就可能导致在泛化过程中无法满足 K 匿名的要求。以下图1为例，、和是时间段内的三条位置信息，它们构成的匿名子集。在时间段内、和构成另一个匿名子集。两个子集都各自满足匿名条件，然而和是同一个用户在相同敏感位置不同时刻进行泛化操作后的位置信息，将它们在空间上进行位置重叠就有可能暴露用户的敏感位置。这时需要对匿名子集重新泛化，如果在时间段内集合中不存在其他位置信息能和、和构成匿名子集，那么匿名子集就被切割成两个匿名子集，从而的匿名条件失效。



图3-1 K -匿名条件失效

在匿名过程中，增强K匿名模型除了采取灵活的匿名条件外，还将需要匿名的位置信息和其他干扰信息在时间和空间上进行相似度比较，计算出更精确的匿名区域，从而提高匿名服务的质量。

## 位置历史匿名保护算法

增强的 K-匿名位置隐私保护模型为了解决 LBSNS 中位置历史记录在离线状态下的匿名保护问题，提出了一个位置历史匿名保护算法(Location History Anonymizing Protection Algorithm, LHAPA)。该算法在对目标位置信息进行匿名保护的时候需要考虑时间、空间和兴趣点标识三个方面，下面对算法涉及到的相关概念做简要介绍。

### 时空相似度

位置历史匿名保护算法在对目标位置信息进行匿名保护、寻找相似的 K-1条位置数据时，需要考虑时间和空间两个维度。目标位置信息和其他 K-1条干扰位置信息在时间、空间上越相似，它们构成的匿名集合保护效果越好，而且将来通过知识发现提供的个性化服务质量越高。这就需要在时间和空间维度上对位置信息之间的相似度进行度量。

位置信息在时间维度上的相似度可以由公式4-1计算得到：

其中，和表示两个位置信息的时间戳，是增强 K-匿名位置隐私保护模型的时间阈值(分钟)。

公式4-2则描述了位置信息在空间维度上的相似度。其中和表示两个位置信息的地理坐标位置，是增强 K-匿名位置隐私保护模型的空间阈值。

结合公式4-1和4-2，则可以通过时间和空间上的相似度联合计算得到两个位置信息之间的相似度。如果LBSNS应用对时间或者空间有特殊要求，可以在计算相似度的时为这两个维度引入权值，公式4-3所描述：

默认情况下，在计算两个位置信息之间相似度时，时间维度和空间维度的权值都为0.5。

### 兴趣点泛化

位置信息中包含兴趣点标识，从该标识中LBSNS 应用可以了解该兴趣点的名称、类型和标签，并从中发现用户的偏好，从而为用户提供个性化服务。然而如果不对兴趣点标识进行泛化处理，攻击者有可能通过将位置信息与其已知的黄页数据连接，从而知道用户的敏感位置信息。在对位置信息进行匿名时，也需要对兴趣点标识进行泛化处理。

对兴趣点的泛化有两种方式。当目标位置信息无法找到另一个能够与其构成匿名子集的其他位置信息时，将泛化后的匿名区域与数据库中的区域相比较，若两者在位置上包含或相交，则将该区域的类型作为兴趣点的语义标记。如果能找到其他位置信息与目标位置信息构成匿名子集，那么则将该子集中所有兴趣点标识对应的类型作为语义标记。

### 位置历史匿名保护算法

本文提出的位置历史匿名保护算法的主要思想是：对于位置历史信息中的任意一条位置信息，在增强的 K-匿名位置隐私保护模型规定的时间阈值和空间阈值范围内，找到一个由其它位置信息构成的集合，并且对于其中每一条位置信息保证。接着将集合中所有位置信息与在时间和空间维度上进行相似度比较，找到其中与相似度最高的K-1个位置信息，然后对这些位置信息在时间、空间和兴趣点标识上进行泛化，形成匿名子集AS。同时，对于每一个用户，需要一张额外的数据表记录该用户的所有位置信息中包含的兴趣点标识，及其对应的匿名位置区域和兴趣标识泛化语义，该表的格式为，这张表的作用是判断该用户是否在该位置上己经存在匿名的位置区域。对于新的匿名子集中的每一条数据，如果存在原有的与其用户标识和原始的兴趣点标识都相同，那么将的匿名空间区域与新的匿名空间区域进行对比，将区域面积较小的作为该位置对应的匿名空间区域，并且修改相应的兴趣点标记语义。如果不存在，则更新数据表。

下面用图4-1来描述位置历史匿名保护算法的执行过程。、和三条未匿名保护的位置信息，根据算法在时间、空间和兴趣点标识上泛化形成匿名子集。同时在数据表中记录匿名子集用户的用户标识、原始兴趣点标识、匿名泛化区域和泛化后的兴趣点标识语义。下一时刻，对另一条目标数据进行匿名保护，它与和一起泛化后构成匿名子集。对于中的每一条数据，发现对应的用户在相同的位置已经存在有匿名区域，这时需要比较这两个匿名区域，若原始的匿名区域小于新的匿名区域，需要对重新泛化，分割成两个匿名子集和。



图4-1 位置历史匿名保护算法示例

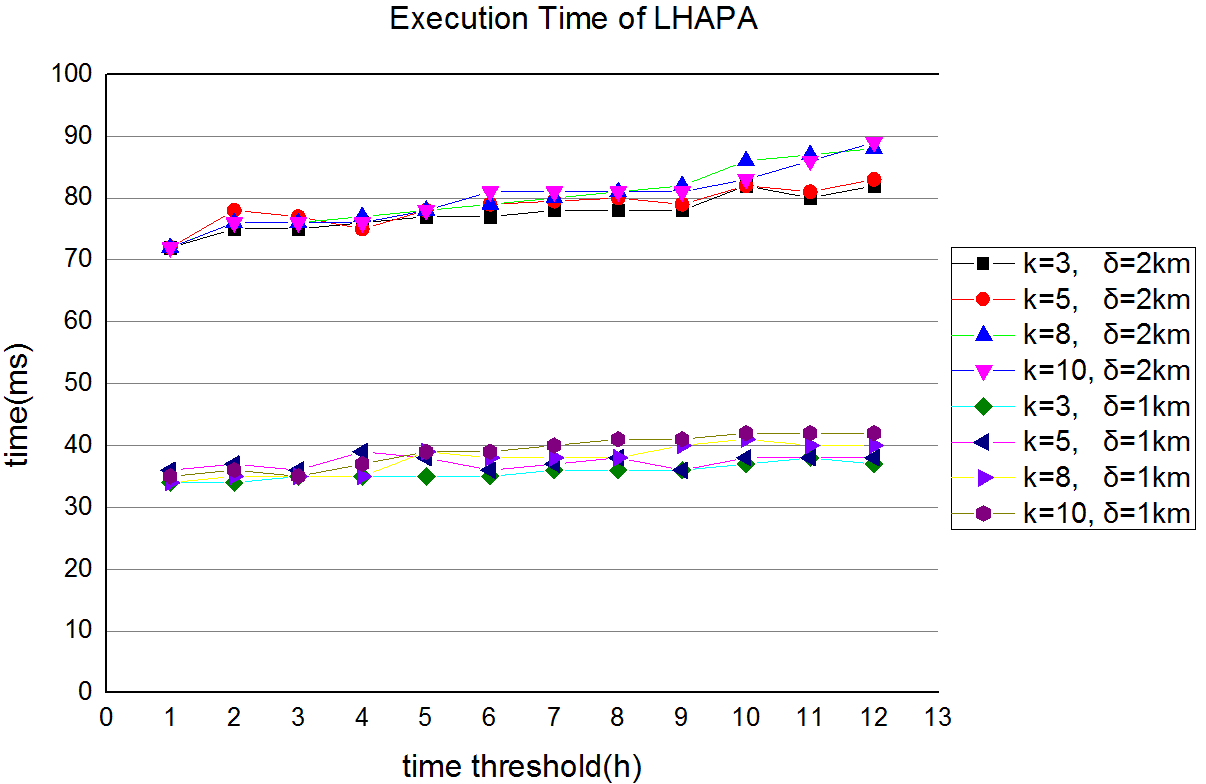
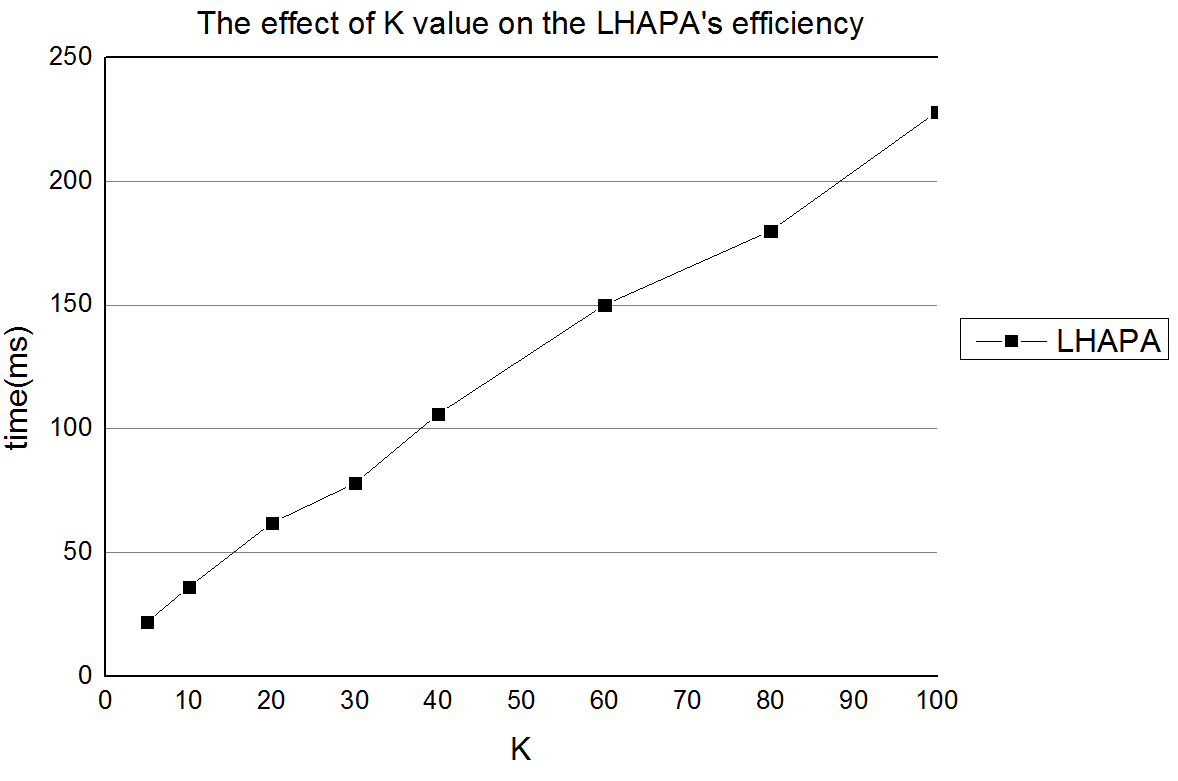
|  |
| --- |
| 位置历史匿名保护算法 |
| 输入： |
| ：位置历史数据集合，：目标位置信息 |
| 输出： |
| ：的匿名子集 |
| 算法： |
| 1：begin |
| 2： 在时间阈值和空间阈值范围内，寻找候选位置信息集合，保证中用户标识与不同 |
| 3： if then |
| 4： foreach in do |
| 5： 计算和相似性 |
| 6： 选取其中最相似的 K-1个位置信息泛化后构成匿名子集 |
| 7： foreach in do |
| 8： if 对应的用户在该位置上已经存在匿名区域 |
| 9： 比较新生成的匿名区域和数据表 T 中原始的匿名区域，若前者小于后者，重新泛化 |
| 10： else |
| 11： 更新数据表 |
| 12： return |
| 13：else if then |
| 14： 泛化中的位置信息构成匿名子集 |
| 15： foreach in do |
| 16： if 对应的用户在该位置上已经存在匿名区域 |
| 17： 比较新生成的匿名区域和数据表 T 中原始的匿名区域，若前者小于后者，重新泛化 |
| 18： else |
| 19： 更新数据表 |
| 20： return |
| 21：else |
| 22： 使用空间和时间阈值单独泛化 |
| 23： 更新数据表 |
| 24： return |

## 实验与分析

为了评估LHAPA算法的有效性，本文通过实验来验证LHAPA算法的构建效率和时空匿名效果。实验的数据主要来自于Gowalla的签到数据集，由于其不含有兴趣点标识，所以本文通过与美国加利福尼亚的兴趣点数据集相连接筛选出最终的测试数据集，总共有667821条数据，包括15039个用户，63种兴趣点类型。实验设置为 DELL Optiplex 780主机，硬件配置为：Core 2 Duo 3.00GHz CPU，4.00GB RAM。算法采用 Python 实现，使用 MySQL 作为数据库。

### LHAPA的整体性能

常见的非对称加密算法有RSA算法、基于离散对数的加密方式和基于椭圆曲线的加密方式，其中RSA算法是目前使用的最广泛的一种非对加密算法。RSA算法都是在整数环中进行计算的，其核心是模运算。其中RSA的加密流程可以用如下公式所表示：

为了评估LHAPA的整体性能，实验一将空间阈值设为1km和2km，时间阈值设为1h~12h，K的值分别设为3、5、8、10，计算采用LHAPA算法对位置历史中的每条位置数据进行匿名保护需要的平均时间，实验一结果如下图5-1所示。

. 图5-1 LHAPA的整体性能 图5-2 K值对LHAPA效率的影响

从图5-1可知，时间阈值对LHAPA的影响很小，其主要作用的是空间阈值。 虽然LHAPA的执行时间随着空间阈值的增加而线性增长，但该算法主要是为了保护离线的位置历史数据，对实时性要求不高，因此LHAPA的整体性能还是很高的。

实验二验证K值对LHAPA效率的影响，本实验将时间阈值设为6h，而不限定空间阈值，K值设定为5、10、20、30、40、60、80、100，同样计算采用LHAPA算法对位置历史中的每条位置数据进行匿名保护需要的平均时间，实验二结果如上图5-2所示。

从图5-2实验结果可以看到，LHAPA的执行时间随着K值的变大而增加，这主要因为LHAPA的效率主要由查询的空间范围决定。在时间阈值不变的情况下，如果K值越大，那么算法所需查询的空间范围就更广，所以查询时空数据库所需要的时间也就越长。

综合实验一和二可知，时间阈值对模型构建效率影响小、空间阈值影响大。实际中，可选择较大的时间阈值、较小的空间阈值，从而获得较高的模型构建效率。

### K值对LHAPA效率的影响

本文模型在构建匿名子集时需考虑空间和时间，而传统模型仅从空间上选取最相似的位置信息构成匿名子集。实验三主要对两种模型的匿名效果进行比较。

首先时间阈值设为1h，空间阈值设为1km和2km，K的值分别设为3、5、8、10，两种模型生成的匿名子集的空间区域面积分别为MS和LS，它们的比值MS/LS如下表4-1所示。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 空间阈值 | 空间权值 | K = 3 | K = 5 | K = 8 | K = 10 |
| 1km | WS = 0.5 | 223% | 235% | 171% | 142% |
| WS = 0.7 | 223% | 107% | 0% | 0% |
| WS = 0.9 | 104% | 0% | 0% | 0% |
| 2km | WS = 0.5 | 352% | 272% | 209% | 193% |
| WS = 0.7 | 352% | 195% | 112% | 193% |
| WS = 0.9 | 194% | 218% | 0% | 0% |

表4-1 两种模型匿名区域大小的比较

从上表结果可知，在默认空间权值下，MS普遍是LS的两倍。随着K的增大，两者间的差距逐渐缩小；而空间阈值越大，两者间的差距也越大。另一方面，随着空间权值的增大，两者差距变小。

接着用MT来表示本文模型匿名子集的时间间隔，LT表示传统模型匿名子集的时间间隔。两者的比值LT/MT如表4-2所示。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 空间阈值 | 时间权值 | K = 3 | K = 5 | K = 8 | K = 10 |
| 1km | Wt = 0.5 | 461% | 225% | 232% | 108% |
| Wt = 0.3 | 461% | 115% | 0% | 0% |
| Wt = 0.1 | 226% | 0% | 0% | 0% |
| 2km | Wt = 0.5 | 802% | 1402% | 207% | 212% |
| Wt = 0.3 | 802% | 365% | 112% | 112% |
| Wt = 0.1 | 131% | 116% | 0% | 0% |

表4-2 两种模型匿名子集时间间隔大小的比较

从上表结果可知，两种模型生成的匿名子集时间间隔差异较大。虽然随着K值的增大，两者间的差异有所减小，但基本上前者是后者的两倍。两一方面，随着空间阈值的增大，该比例也在逐渐增大。

综合两表可以发现，本文模型因需同时考虑时间和空间维度，在空间维度上的匿名效果略不及传统模型，但在时间维度上匿名效果远远优于传统模型。因此，在实践中使用本文模型时，可以依据实际需求选择合适的空间和时间阈值以及两者的权值，以获得最佳的匿名效果。

## **本章小结**

针对LBSNS 的位置隐私保护问题，本文提出一种增强的 K-匿名位置隐私保护模型，在传统的 K-匿名模型上着重考虑如何保护离线情况下的位置历史信息，提出了在保护位置历史信息时构建增强 K-匿名隐私保护模型采用的构建算法位置历史匿名算法，同时采用实验验证了该算法在构建 K 匿名集合有较高的效率和较好的匿名效果。该模型不仅弥补了传统隐私保护模型在位置历史保护上的不足，而且接下来将使用匿名数据为用户提供个性化服务。

# 位置隐私保护在LBSNS个人兴趣区域发现中的应用

## 引言

随着移动定位技术的发展以及智能便携设备的普及，通过对移动用户的位置信息进行挖掘和分析，发现用户的兴趣区域（Region of Interest，ROI），并以此为用户提供智能化和个性化的服务，已经成为了许多移动应用吸引用户、增加用户粘性的重要手段之一。对于LBSNS而言，无论是社交关系的发现还是内容推送都可以受益于对兴趣区域的发现。一方面，位置因素在建立新的移动社交关系时扮演着重要的角色，彼此生活中兴趣区域的交集越大，分享共同话题的可能性就越大，对于彼此的认同感可能更强。另一方面，LBSNS在进行内容推送时，若所推送的内容与用户的兴趣区域之间有很高的相关性，则更有可能引起用户的关注，从而起到有的放矢的作用。

在进行兴趣区域抽取的过程中，目前广泛采用的位置数据来源主要分为两种：签到活动对应的位置信息以及用户日常的轨迹信息。对于前者，它通常会包含该位置所属的兴趣点信息，一般是指用户访问的黄页对象，包含了空间属性和一系列非空间属性。例如一家餐馆，它的空间属性是它的地理位置信息，非空间属性则可能是它的菜系、价格和服务质量等。对于后者，它是一种移动模型，是在一定时间和空间范围内，移动对象位置变化过程的一个记录，可以表示为一系列位置实体和时间实体的序列。相对于兴趣点，对轨迹信息的挖掘和分析可以发现更多的用户信息，包括用户重要的活动场所、生活习惯、行为动作以及偏好等。

但是，对于LBSNS而言，无论是采用签到位置信息，还是采用轨迹信息去发现用户的兴趣区域，两者都存在一定的缺陷。一方面，轨迹信息的获取需要长时间记录用户的位置数据。对于便携移动设备，轨迹信息获取的过程对它有限资源的消耗会非常大，由此可能会影响其它工作的正常进行。而另一方面，签到位置信息虽然能够更加准确地定位用户的兴趣区域所在，但是也很容易造成用户的位置隐私泄露，从而可能产生不可预知的潜在危险。

针对上述问题，本章以采用增强K-匿名模型保护的签到活动的位置信息为基础，定义了一种兴趣区域类型：个人兴趣区域，并分别提出了个人兴趣区域发现方法。

对于个人兴趣区域，本章将其分为了私密区域和偏好区域。私密区域指那些用户隐私敏感的区域，如居住区域；而偏好区域则是指用户喜欢去的公共场所，如购物中心。本章分析了两种个人兴趣区域之间的异同，并根据行为规律、活动范围和社交关系三种影响用户签到活动的主要因素，总结了签到规律、签到密度分布和区域信息熵三种重要的签到特征，以此为基础提出了对这些区域进行抽取和归类的方法。实验结果显示了本章提出的方法能够发现一部分用户的私密区域和大部分偏好区域。同时，实验结果也显示了随着增强K-匿名模型保护力度的增大，即K值的增大，可发现用户的私密区域也随之减少，从而从侧面反映了增强K-匿名模型的位置隐私保护效果。

兴趣点和兴趣区域都是地理信息系统中的术语。前者泛指可以抽象为地理位置坐标点的位置实体，如学校、医院、银行和餐馆等；后者则是以空间区域为范围的位置对象，如大型公园、中央商圈和居住区等。在空间地理数据库中，兴趣点或兴趣区域对象一般都会包含名字、类别和经纬度坐标点集合这三个属性。因为从用户对兴趣点和兴趣区域的访问频率以及滞留时间等因素，能够挖掘出用户潜在的一些个人偏好，所以围绕如何发现兴趣点和兴趣区域也一直是位置相关服务领域中的研究热点。

通常来说，兴趣点和兴趣区域的发现有以下几个步骤：

1、采集用户时空数据。

2、将一段时间内稳定在一个经纬度坐标范围内的数据作为有效的位置对象，同时提取其位置语义。

3、相同的位置对象进行统计归类，并以此作为数据集。

4、依照一定的算法识别用户访问位置对象的偏好。

5、对结果进行准确性的判断。

本文研究的LBSNS个人兴趣区域的发现方法同前人的研宄工作的区别主要有以下几点：（1）传统的兴趣点和兴趣区域的发现方法缺乏对兴趣区域与签到活动之间的内在联系的分析，通常仅在空间维度上采用聚类等方法发现此类区域。而本文提出的兴趣区域发现方法则通过分析兴趣区域的被签到次数、用户计数和信息熵之间的典型关系，并结合聚类、排序和过滤的方法，提出相关解决方法；（2）本文研究的基础数据集是基于经过增强K-匿名模型保护后的签到位置数据，这些数据在时间和空间维度上存在了大量的不确定性，使得对兴趣区域的发现由基于点的形式转变为了基于区域的形式，因此也更具挑战。

## 个人兴趣区域的发现方法

对LBSNS的个人用户而言，他更关心与其日常生活息息相关的那些区域附近所发生的事。例如，当LBSNS向用户推荐购物优惠时，如果被推荐的对象处于用户居住区域附近，那么用户去访问的概率肯定会大幅增加。因此，抽取出个人兴趣区域对LBSNS为用户进行精准定位，以此提供更好的个性化服务起到了至关重要的作用。

在这一小节，本文对个人兴趣区域的一些特点进行了系统的描述，并以此为基础，定义了两种特别的个人兴趣区域：私密区域和偏好区域。其中，私密区域是指对用户而言最需要进行隐私保护的区域，这类区域通常不是广泛兴趣区域，具有较高的私密性，例如用户的居住区域。偏好区域是指用户经常会去那些公共区域，这些区域可以反映出用户的一些偏好，在用户生活中同样扮演重要角色，例如用户习惯购物消费的区域。本节随后也提出了对应的个人兴趣区域发现方法。

### 个人兴趣区域的特点

LBSNS用户进行签到活动的区域并非完全是随机性质的，通常都会受到一些主观和客观因素的影响。例如，如果用户经常在各种饭店进行签到，那么有可能该用户的兴趣是美食；如果用户经常在同一中央商务区签到，那么有可能他就在那里工作。从已有的研究工作已经证明了绝大多数人的生活都是有规律，并且活动的范围也都是有限的。因此，如果以此为假设，LBSNS用户的签到活动也应该是有迹可循的。从时空角度出发，本文列举了一些影响用户在私密区域和偏好区域进行签到活动的影响因素以及对应的个人兴趣区域的特点，总结为以下几个方面：

1、生活规律

对于绝大多数的LBSNS用户来说，他们的签到活动符合以下特征：在一些较少的区域进行了多次签到，在大部分区域偶尔有签到。这些较少的区域就可能就包含了用户的私密区域和偏好区域。

对于私密区域和偏好区域，它们之间既有相同点也有不同点。两者的相同点是，它们都是用户签到概率最高的区域之一，并且签到时间的分布都比较规律。

两者的不同点在于，私密区域除了用户自己签到外，其他用户签到的可能性较低，而偏好区域则可能被许多人签到。因此，从信息熵的角度，用户私密区域的信息熵可能较低，而偏好区域的信息熵则可能较高。

2、活动范围

已有研究工作己经证明了人们日常出行的距离符合幂指数分布的特点，即地点距离用户的私密区域越近，它被LBSNS用户签到的概率越高。换言之，如果在一片用户签到区域周围存在许多其它用户的签到区域，那么该片区域很有可能是用户的个人兴趣区域。

3、社交关系

社交关系对于用户的签到活动也具有非常大的影响力。日常生活中，同事之间一起出去吃工作餐或者朋友之间的串门都是频繁发生的活动。因此，如果一个区域被若干个具有社交关系的用户频繁签到，那么这些区域就可能是个人兴趣区域。

### 个人兴趣区域发现方法

本节根据上一小节总结的个人兴趣区域的特点，提出了一种个人兴趣区域的发现方法，该方法的思想是首先将潜在的用户个人兴趣区域抽取出来，然后通过排序和过滤的方法，将最有可能为是私密区域和偏好区域的两个区域抽取出来。下面首先对该方法中使用的相关的排序和过滤的方法进行描述。

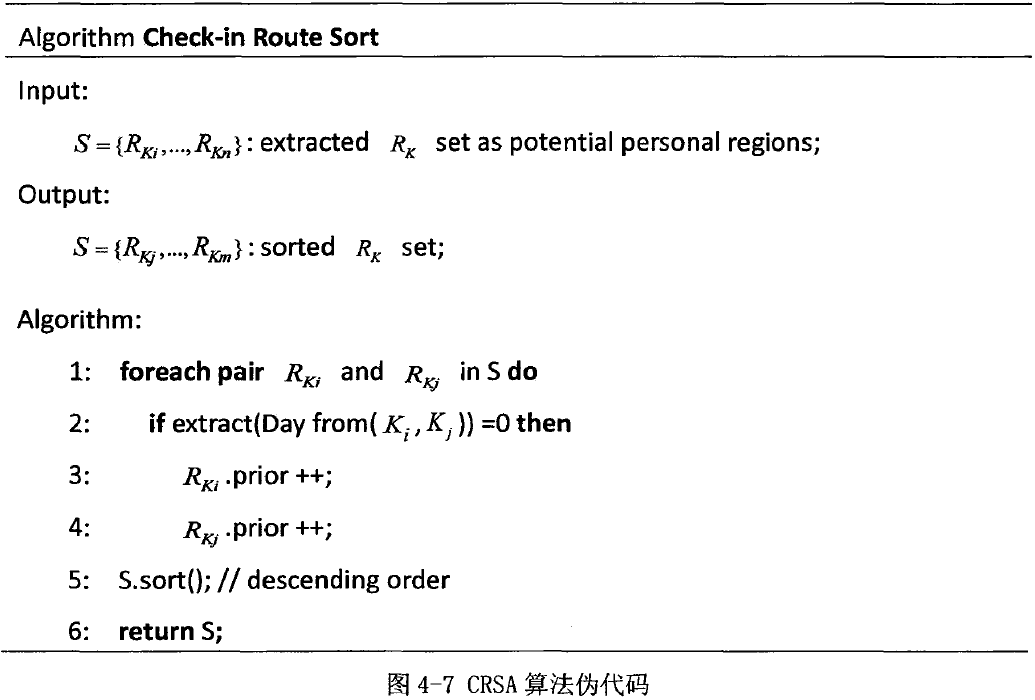
#### 签到规律排序算法

签到规律排序是根据上一小节中描述的用户日常生活规律对用户签到行为影响的特点所总结出的，对潜在兴趣区域是否为私密区域和偏好区域的可能性大小进行排序的方法，它的排序依据是一个区域在同一天与其它区域共同被签到的次数。下面本文对签到规律排序所涉及相关的概念进行描述和定义。

根据增强K-匿名模型对象的定义，对于原始签到点为同一对象的增强K-匿名模型对象集合，它们在空间维度上的矩形区域是相同的。为了方便描述，本章将在后续小节中使用RK表示在空间维度上相同的增强K-匿名模型对象集合K的空间对象，其具体定义如下：

定义 **区域对象RK** 如果存在一个增强K-匿名模型对象集合K={K0,...,Kn}，对于其中任意一个增强K-匿名模型对象Ki，它在空间维度上的矩形区域和集合K中的其它增强K-匿名模型对象都相同，那么使用表示RK集合K在空间维度上的矩形区域，并且矩形区域Rk被签到的计数Count(RK)=|K|。

定义 **规律签到关系** 对于任意两个增强K-匿名模型对象Ki和Kj，如果它们的时间间隔对象位于同一天范围内，那么则称该两个对象之间存在规律签到关系。

签到规律排序算法（Check-In Route Sort Algorithm，CRSA）的思想是依据增强K-匿名模型对象之间存在规律签到关系的次数对潜在的个人私密区域或者偏好区域进行排序。算法首先对抽取出的增强K-匿名模型对象集合中的规律签到关系进行统计，如果两个增强K-匿名模型对象Ki和Kj之间存在规律签到关系，则对它们对应的区域对象RKi和RKj的优先级Rkiprior和RKjprior分别加1。当整个算法结束时，优先级最大的区域对象的排序位置最前。下图4-7描述了算法的伪代码：

#### 密度分布排序算法

密度分布排序算法（Density Distribution Sort Algorithm,DDSA）的理论基础是人们日常生活中的活动范围分布。在上一小节已经指出，人们的活动范围都较有限，私密区域和偏好区域周围是人们活动概率较高的区域。因此统计潜在兴趣区域周围的签到对象情况，有助于帮助确定用户的个人兴趣区域。

DDSA算法将潜在的兴趣区域对象之间的周围空间进行分割，将分割后空间中所包含的增强K-匿名模型对象的数量作为该空间的密度，密度值越高的空间所对应的兴趣区域对象排序顺序越高。该方法将潜在兴趣区域对象的空间矩形的中心点作为其周围空间分割的基础点，并采用Voronoi图的方法对空间进行分割。

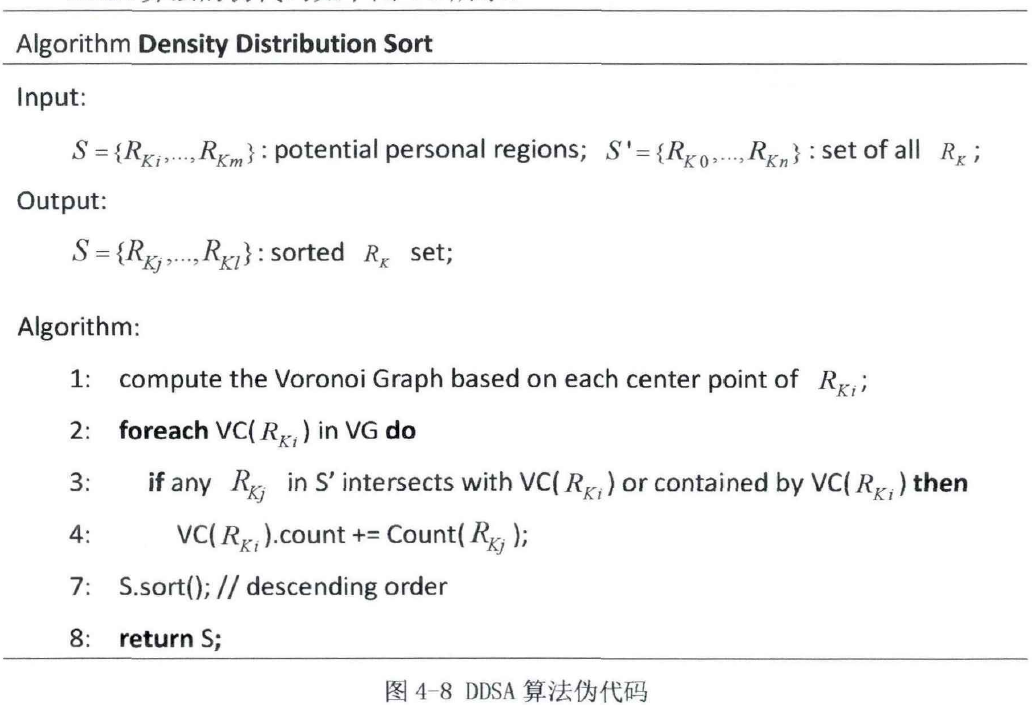
DDSA算法的伪代码如下图所示：

图4-9显示了密度分布排序方法中的空间分割情况。图中深红色矩形为抽取出的潜在兴趣区域，浅红色矩形为一般签到区域。可以看到，整个空间被划分为了三个Voronoi单元（VC）。对于每一个VC，DDSA算法将与它的最大内切圆在空间上存在包含或者相交关系的增强K-匿名模型对象计数作为该Voronoi单元对应的潜在兴趣区域的密度。

#### 信息熵过滤

对于个人兴趣区域的发现，信息熵可以发挥作用。首先，在通常情况下，用户的私密区域和偏好区域的信息熵的大小显然是不同的。通常来说，在偏好区域活动的用户数较多，故而它的信息熵也会较大。相反，在用户的私密区域可能只有用户的家人和朋友才会签到，所以信息熵会较低。

其次，如果从社交关系对签到活动的影响来看，用户的私密区域和偏好区域都会存在与用户好友进行签到的情况，但是这两种情况下的信息熵也会有所不同。由于偏好区域通常是一些较公共的区域，因此用户的好友在该类区域进行签到的次数也会较多，所对应的信息熵也会较大。而在私密区域和用户一起签到的好友，如果两者仅是朋友关系的话，其签到的次数可能较少，因而信息熵的值也会较小。

这里提出两种类型的信息熵，分别称为公共信息熵和社交信息熵。**公共信息熵和在广泛兴趣区域发现方法中的信息熵是同一个概念**，在计算对应的信息熵值大小的时候使用的是整个用户空间的数据集合作为信息空间。而社交信息熵则使用与用户存在社交关系的用户数据集合作为计算信息熵的信息空间。当公共信息熵和社交信息熵都较处于high簇时，潜在兴趣区域被归为偏好区域，而当公共信息熵和社交信息熵都处于low簇时，则被标记为私密区域。

综上，个人兴趣区域发现方法通过比较两种区域的公共信息熵和社交信息熵的方法对区域的类型进行判断和标记。

#### 个人兴趣区域发现方法

个人兴趣区域的发现方法首先根据特定用户的签到次数，釆用聚类算法将签到次数值较大簇中的区域对象RK抽取出来，并分别采用签到规律排序和密度排序的方法对这些RK对象进行排序。其次，对于这些RK对象在空间维度上所涉及的网格单元分别计算公共信息熵和社交信息熵，并根据两种信息熵的分布情况，对相关的网格单元进行归类。最后，将私密区域和偏好区域排序最靠前的两个RK对象所涉及的网格单元抽取出来，作为所发现的两种个人兴趣区域。

详细流程见下图4-10所示：

该方法的具体步骤描述如下：

1、对于用户i，根据他所有的RK对象的签到次数大小，使用K均值聚类算法，将其划分为3个簇，分别为low、medium和high簇。

2、将划分为high族的RK对象集合提取出来作为潜在的兴趣区域集合。如果high簇中的集合大小小于等于1，那么跳至步骤5。

3、对于集合中任意两个RK对象，使用CRSA算法进行排序。

4、同时对潜在的兴趣区域集合中的RK对象采用DDSA算法进行排序。

5、对集合中RK进行公共信息熵和社交信息熵的计算，根据两者的分布情况分别对网格单元进行类型标记。

6、将两种类型排序最靠前的区域对象抽取出来，作为发现的两种个人兴趣区域。

## 实验与分析

## 本章小结

本章研究了采用增强K-匿名模型保护后的位置信息在LBSNS的个人兴趣区域发现方面的应用。本章定义了个人兴趣区域，并提出其发现方法。个人兴趣区域可细化为两类：私密区域和偏好区域。私密区域是用户最需要隐私保护的区域，例如用户的居住区域；而偏好区域则是用户喜欢去的公共区域，例如购物中心。实验结果显示，可抽取私密区域的用户数占用户总数的比例并不高，而且随着增强K-匿名模型K值的增大，该比例更会逐渐下降，这也从侧面反映了增强K-匿名模型对用户位置隐私的保护效果与K值关系密切。和私密区域相比，可抽取偏好区域的用户数占用户总数的比例较高，并且受到增强K-匿名模型K值的影响较小。两种个人兴趣区域的抽取结果显示，被抽取的私密区域和偏好区域能够准确地反映出用户日常生活中的一些重要区域，因而为个人化的服务提供了很好的支持。

本章的研究结果表明了增强K-匿名模型在LBSNS的个人兴趣区域发现上作用明显，既能够对用户的签到信息起到较好的保护作用，也能够发现LBSNS的个人兴趣区域。

# 总结与展望

## 工作总结

## 未来展望

# 致 谢