软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2015,26(9):2373−2395 [doi: 10.13328/j.cnki.jos.004857]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

位置服务隐私保护研究综述*

张学军[1,2,3], 桂小林[1,3], 伍忠东[2]

[1](西安交通大学 电子与信息工程学院,陕西 西安 710049)

[2](兰州交通大学 电子与信息工程学院,甘肃 兰州 730070)

[3](陕西省计算机网络重点实验室(西安交通大学),陕西 西安 710049)

通讯作者: 桂小林, E-mail: xlgui@mail.xjtu.edu.cn

摘 要: 由于位置感知移动电子设备的繁荣,位置服务(LBS)几乎在所有的社会和商业领域广泛流行.虽然 LBS 给个人和社会带来了巨大利益,但也给用户的隐私造成了严重威胁.因为用户享受 LBS 的同时需要向不可信的 LBS 提供商泄露其位置和查询属性,而附加在这些信息上的上下文揭露了用户的兴趣爱好、生活习惯、健康状况等.如何保护用户的隐私免受恶意提供商的侵犯,对 LBS 生态系统的健康发展至关重要,因而引起了研究者的广泛关注.对 LBS 隐私保护的研究现状与进展进行综述.首先介绍 LBS 隐私的概念和威胁模型;然后,从系统结构、度量指标、保护技术等方面对现有的研究工作进行细致的分类归纳和阐述,重点阐述当前 LBS 隐私保护研究的主流技术:基于扭曲法的隐私保护技术;通过对各类技术性能和优缺点的分析比较,指出了 LBS 隐私保护研究存在的问题及可能的解决方法;最后,对未来研究方向进行了展望.

关键词: 基于位置的服务/位置服务;位置隐私;查询隐私;隐私度量

中图法分类号: TP309

中文引用格式: 张学军,桂小林,伍忠东.位置服务隐私保护研究综述.软件学报,2015,26(9):2373−2395. http://www.jos.org.cn/1000-9825/4857.htm

英文引用格式: Zhang XJ, Gui XL, Wu ZD. Privacy preservation for location-based services: A survey. Ruan Jian Xue Bao/ Journal of Software, 2015,26(9):2373−2395 (in Chinese). http://www.jos.org.cn/1000-9825/4857.htm

Privacy Preservation for Location-Based Services: A Survey

ZHANG Xue-Jun[1,2,3], GUI Xiao-Lin[1,3], WU Zhong-Dong[2]

[1](School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

[2](School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

[3](Shaanxi Province Key Laboratory of Computer Network (Xi'an Jiaotong University), Xi'an 710049, China)

Abstract: Location-based service (LBS) has recently become popular in almost all social and business fields due to the boom of location-aware mobile electronic devices. LBS, albeit providing enormous benefits to individuals and society, poses a serious threat to users' privacy as they are enticed to disclose their locations and query attributes to untrusted LBS providers via their LBS queries. Moreover, the contextual information attached to these locations and service attributes can reveal users' personal interests, life styles, health conditions, etc. How to preserve users' privacy against potentially malicious LBS providers is of vital importance to the well-being of LBS ecosystem, and as such, it attracts great attentions from many researchers. This paper provides a review of the state-of-the-art of privacy preserving for LBS. First, the concept and threat model of LBS privacy are presented. Then, the existing schemes for preserving users' LBS privacy are described in detail from the aspects of architecture, metric and technology. Next, a pointed discussion is placed on

* 基金项目: 国家科技重大专项 (2012ZX03002001); 国家自然科学基金(61472316, 61172090, 61163009, 61163010); 高等学校博士学科点专项科研基金(20120201110013); 陕西省科技攻关项目(2012K06-30, 2014JQ8322); 中央高校基础科学研究基金(XJJ2014049, XKJC2014008); 陕西科技创新项目(2013SZS16-Z01/P01/K01); 兰州交通大学青年科学基金研究项目(2014026)

收稿时间: 2014-05-22; 修改时间: 2014-06-20, 2015-03-30; 定稿时间: 2015-05-14

the latest mainstream technology, with emphasis on the distortion-based technology. Further, following a comprehensive comparison and analysis of the performance and defects of various technologies, the problems and possible solutions for LBS privacy preserving are pointed out. Finally, some future research directions are provided.

Key words: location-based service; location privacy; query privacy; privacy quantification

随着无线通信技术和移动定位技术的发展,越来越多的移动设备具备 GPS 精确定位功能,使得位置服务(LBS)日益风行,是为移动用户提供的最有前途的服务之一.LBS 是指基于移动设备的地理位置和其他信息,为移动用户提供的信息和娱乐服务[1],其典型应用包括地图类应用(如 Google Maps)、兴趣点检索(如 AroundMe)、优惠券或折扣提供(如 GroupOn)、GPS 导航(如 TomTom)和位置感知社会网络(如 Foursquare)等.早期的 LBS 系统主要用于军事和涉及国家重要利益的民用领域.目前,LBS 已被广泛应用在军事、政府产业、商业、医疗、紧急救援、民生等领域[2].美国著名的市场研究公司 ABI Research 发布预测,LBS 全球总收入将由 2009 年的 26 亿美元上升到 2014 年的 140 多亿美元.然而,LBS 在给个人和社会带来巨大好处的同时,也引发了严重的隐私关注.因为用户获取 LBS 时需要报告他们的位置信息,而位置数据既直接包含了用户的隐私信息,又隐含了用户通常想保护的其他个人敏感信息,如家庭住址、生活习惯、健康状况和社会关系等.因此,把这些私人信息泄露给不可信的第三方(如 LBS 提供商),会打开滥用个人数据的大门,对用户各方面的隐私带来严重威胁.例如,从匿名 GPS 数据中能推断出个人的家庭地址、工作单位和社会关系[3,4],预测出用户过去、现在和将来的位置[5],推断出个人的行踪[6];甚至可以利用室内位置信息推断出个人的工作角色、年龄、爱好(如是否抽烟)等[7].因此,对用户的 LBS 隐私进行保护是一个至关重要的问题.

移动互联网、社会网络、大数据等新兴技术的发展和广泛使用,使得 LBS 隐私保护问题越来越严重,成为工业界和学术界广泛关注的热点问题.2003 年,Beresford 提出位置隐私的概念[8],开启了对 LBS 隐私保护研究的先河.此后,LBS 隐私保护日渐成为信息技术领域的研究热点,许多著名的国际会议及期刊陆续发表了大量 LBS 隐私保护的相关研究成果.本文综述 LBS 隐私保护研究的最新进展.发表的文献中,已有一些关于 LBS 隐私保护的综述文献:Ghinita[9]从私有查询和轨迹匿名两个方面对位置隐私进行了综述,但没有涉及隐私度量和查询隐私;Krumm[10]着重评述了匿名、模糊化隐私保护技术和一些利用位置数据几何性质的隐私侵犯算法,但没有涉及系统结构和查询隐私;霍峥等人[11]从传统关系数据隐私保护向时空方向拓展的角度对数据发布中的轨迹隐私保护和 LBS 中的轨迹隐私保护关键技术进行了分析和比较,但没有涉及查询隐私;Shin 等人[12]总结了 LBS 通用隐私威胁模型,分析比较了各种 LBS 隐私保护技术及度量指标,但对系统结构和保护技术的综述不够全面.而且,这几篇论文发表较早,无法收录之后的 LBS 隐私保护研究新进展.本文从 LBS 隐私的概念入手,分析 LBS 的隐私威胁模型,总结 LBS 隐私保护的体系结构和度量指标.依据不同 LBS 隐私保护技术在隐私保护度和服务质量之间的权衡,分类阐述基于政策法、扭曲法和加密法的 LBS 隐私保护技术研究进展,特别是新近的研究进展,全面比较和分析不同技术的优缺点及适用场景,并探讨未来的研究方向.

本文第 1 节介绍 LBS 隐私保护关键问题.第 2 节介绍 LBS 隐私保护系统结构.第 3 节阐述 LBS 隐私保护度量指标.第 4 节详细阐述和分析各种 LBS 隐私保护技术.第 5 节是各种主要 LBS 隐私保护技术的性能评估和比较.第 6 节总结全文并探讨未来的研究方向.

1 LBS 隐私保护关键问题

1.1 理解LBS隐私

LBS 应用的一个典型例子是关于最近兴趣点(POIs)的搜索引擎,如 Google Maps.用户使用具有 GPS 定位功能的移动设备向 LBS 服务器(也称 LBS 提供商)发送包含他当前位置和感兴趣服务属性(如医院等)的 LBS 请求,服务器会返回少量和用户指定服务属性相匹配的兴趣点,而且这些兴趣点在地理上接近于用户的当前位置.一般而言,一个 LBS 请求可看成是 LBS 服务器空间数据库上的一个查询,如:

SELECT TOP(k) FROM *POI* WHERE type=U_{poi} ORDER BY DISTANCE (*POI.location*,*userLoc*) ASC,

其中,*POI* 是兴趣点空间数据库;U_{poi} 是服务属性值,即查询内容;*userLoc* 是用户的当前位置;*k* 是预定义参数.通常,*userLoc* 在排序函数中被指定为常量,且要与 U_{poi} 一起发送给 LBS 服务器.为了叙述方便,将 LBS 查询定义为一个 4 元组$\langle u,t,loc,U_{poi}\rangle$,*u* 是用户标识,*loc* 是用户在 *t* 时刻提交查询的位置,U_{poi} 是服务属性.

用户在方便地访问各种 LBS 时,会不可避免地在网络中遗留下大量的数字踪迹和服务属性,附加在这些数字踪迹或服务属性上的上下文能揭露用户的个人习惯、兴趣爱好、人际关系、身体状况等私人信息.因此,将这些个人信息暴露给不可信的第三方势必会造成严重的隐私关注.LBS 隐私关注存在两个方面:查询隐私和位置隐私.查询隐私和 LBS 查询中 U_{poi} 的泄露相关,位置隐私与 LBS 查询中 *loc* 的泄露和滥用有关.

下面综合各种文献,给出与 LBS 隐私相关的一些描述性定义.

定义 1(隐私). 隐私是指个人、组织或机构等实体不愿意被外部知晓的信息,如个人兴趣爱好、政治信仰、公司的财务状况等等.

定义 2(个人隐私). 个人隐私一般是指数据拥有者不愿意披露的私人敏感信息,如兴趣爱好、健康状况、收入水平、宗教信仰和政治倾向等.由于人们对隐私的限定标准不同,所以对隐私的定义也就有所差异.一般来说,任何可以确定是个人的,但个人不愿意披露的信息都可以认为是个人隐私.

定义 3(位置隐私). 位置隐私是一种特殊的个人隐私,是指直接与 LBS 查询中的 *loc* 相关的个人敏感信息(如访问的位置是敏感的)以及由 *loc* 推理出的其他个人敏感信息(如兴趣爱好、健康状况、宗教信仰等).

个人的位置包括其现在或过去访问的位置[8,13]:实时位置可使攻击者(本文定义 LBS 服务器为攻击者)找到用户在哪,过去的位置则可帮助攻击者发现用户是谁、住哪、想做什么等.所以,位置隐私涉及:① 用户是否被精确定位;② 用户的个人敏感信息是否被从他访问的位置中推断出来.因此,位置隐私保护既要保护用户过去和现在位置本身的敏感信息不被泄露,又要防止攻击者通过用户位置推理出其他个人敏感信息.

定义 4(查询隐私). 查询隐私是一种特殊的个人隐私,是指与 LBS 查询中的 U_{poi} 相关的个人敏感信息(如查询内容是敏感的)或者从 U_{poi} 中推理出的其他个人敏感信息(如兴趣爱好,健康状况等).

服务属性及其与用户之间的关联往往被认为是敏感的,因为服务属性本身说明了用户的兴趣类别,会直接泄露用户的个人偏好或需求[14].所以,查询隐私涉及:① 用户是否被攻击者识别或去匿名;② 用户的私人敏感信息是否被从他请求的服务属性中推断出来.因此,查询隐私保护既要保证服务属性本身的敏感信息不被泄露,又要防止攻击者通过服务属性和用户的关联推断出其他个人敏感信息.

虽然位置隐私和查询隐私有区别,但它们也密切相关:一方面,如果用户被定位或跟踪(位置隐私泄露),那么他也相对容易被去匿名(查询隐私泄露),如,采用空间受限识别攻击[15]可以很容易俘获用户的查询隐私;另一方面,如果用户被识别,那么他的位置隐私也相对容易被泄露,因为在这种情况下,攻击者有更多可用的信息(如对用户的历史跟踪记录)来成功获得与位置相关的推理性攻击(如移动追踪攻击).

定义 5(LBS 隐私). LBS 隐私是一种特殊的个人隐私,是指由请求 LBS 所产生的查询隐私和位置隐私.

1.2 LBS隐私威胁

LBS 系统通常由移动终端、定位系统、通信网络和 LBS 服务器 4 部分组成,如图 1 所示:移动终端(如智能手机)向 LBS 服务器发送包含用户位置的 LBS 查询;定位系统(如 GPS)实时获取移动终端发送 LBS 查询时的位置;通信网络(如 3G 网络)传输 LBS 查询和从服务器返回的查询结果;LBS 服务器响应用户的查询,并返回定制结果.用户的隐私可能会在 3 个地方泄露:首先是移动终端,如果用户的移动设备被捕获或劫持,那么就会变成恶意的,可能会主动泄露用户的私有信息(包括但不限于位置信息),如何保护用户移动终端的安全本身也是一个非常活跃的研究话题,这里不做深入探讨,有兴趣的读者可参阅文献[16];其次是用户的 LBS 查询和返回结果在通过无线网络传输时,有可能被窃听或遭受中间人攻击,这可以通过传统的加密和散列机制解决;最后是 LBS 服务器,因为一个恶意的攻击者可能就是 LBS 服务器的拥有者或维护者,也可能是俘获并掌控 LBS 服务器的恶意攻击者.这两种情况下,恶意攻击者都能够访问存储在 LBS 服务器上的所有信息,如 IP 地址、用户每次提交的位置和服务属性等.虽然在说明通用 LBS 隐私威胁模型时没有对 LBS 的使用设定额外的要求(如假设用户必须登录后才能使用 LBS),但是攻击者仍能够利用一些侧通道(如每个查询的 IP 地址)和复杂的目标追踪算法把连续

的匿名LBS查询和用户关联起来.为了简化隐私保护问题,通用LBS隐私威胁模型都假定LBS服务器是恶意或不可信的,其他部分则是安全可信的.本文讨论的所有隐私保护技术都基于该通用隐私威胁模型.

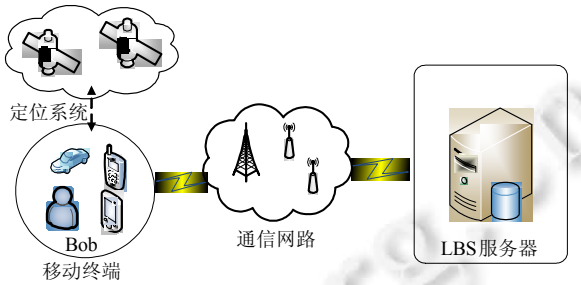

Fig.1 A common LBS architecture

图 1 LBS 通用系统架构

用户使用各种LBS应用时,LBS服务器会收集包含在查询中的用户位置或服务属性.受利益或好奇心驱使,攻击者会结合相关的背景知识,并利用各种攻击方法(如空间受限攻击和基于观察的识别攻击[15]、关联推理攻击[17]、位置相依性攻击[18−20]等),直接或间接的重构出用户希望保护的LBS隐私.经典的LBS隐私保护技术针对这种隐私威胁,确保用户在访问各种LBS应用时不会泄露个人隐私且能获得有用的服务,如使用隐私政策的访问控制策略[21,22]、针对快照查询的位置匿名化方法[15,23]、针对连续查询的轨迹匿名化方法[24−26]等.但是这些经典的隐私保护技术在面对大数据、社会网络等新兴技术时会遇到许多新的隐私威胁,例如:考虑攻击者从社会网络中获取的用户资料等背景知识,基于位置 k 匿名的隐私保护机制就会泄露用户的查询隐私[27];利用获取的位置大数据之间的关联等背景知识,可对精心匿名的位置数据进行成功的反匿名[28].这是因为攻击者可以利用这些新兴技术持续地收集用户的历史信息,并能从更多渠道获取用户的位置或与位置相关的非位置等各种类型的数据,从而使其具备掌握更多背景知识的能力.这样,攻击者将获取的各方面数据、位置和非位置数据之间的关系以及可用的背景知识等相结合,就能以很高的确定性推测出用户的个人敏感信息,从而使得经典LBS隐私保护技术不能提供充分的隐私安全保障.因此,如何防止攻击者利用获取到的各方面的数据并结合可用的背景知识来全面推测用户的LBS隐私,是应对新形势下LBS隐私威胁亟待解决的关键问题.

1.3 LBS隐私保护技术分类及性能评估

1.3.1 LBS隐私保护场景

下面来看一个例子,Alice使用智能手机请求“查询离我最近的肿瘤医院”.该查询由LBS服务器做出响应,并返回结果.由于LBS服务器不可信,Alice的敏感信息有可能被泄露或滥用.为了保护隐私,Alice通常不会与LBS服务器直接交互,而事先通过隐私保护技术对查询进行模糊化处理后再提交.该场景实际包含了LBS隐私保护的两个方面:位置隐私保护和查询隐私保护.例如,Alice不希望任何人知道她目前所在的位置(医院),也不希望任何人知道自己提出了哪方面的查询(与肿瘤相关的医院查询).前者是位置隐私保护的范畴,后者是查询隐私保护的范畴.这种场景下的LBS隐私保护技术是根据用户的个性化隐私需求进行信息论意义上的全面保护.因此,隐私保护技术需要解决以下几个关键问题:

(1) 如何准确地度量用户隐私的披露风险;

(2) 如何选择有效的隐私保护机制全面保护用户的隐私;

(3) 如何权衡用户的隐私水平、服务质量和资源开销.

当前,该场景下的LBS隐私保护技术主要针对用户不同程度的隐私需求,权衡隐私保护效果和服务可用性.

1.3.2 LBS隐私保护技术分类

保护用户的隐私是成功部署LBS应用的基本要求[9].目前,已提出了许多隐私保护策略来增强用户的隐私,但没有任何一种单一策略能够提供完全的解决方案.本文将LBS隐私保护技术分为3类.

(1) 基于政策法的LBS隐私保护技术.

这一技术是指通过制定一些常用的隐私管理规则和可信任的隐私协定来约束服务提供商能公平、安全的使用用户 LBS 查询中的位置信息或服务属性,如 IETF 的 GeoPriv[21]和 W3C 的 P3P[22].由于隐私政策系统本身并不能够执行隐私保护,往往依赖经济、社会和监管压力等,因而不能实现对用户隐私的有效保护,也不会在隐私和服务质量损失之间进行一定的权衡.

(2) 基于扭曲法的 LBS 隐私保护技术.

这一技术是指在 LBS 查询暴露给 LBS 服务器之前,事先对查询中的时空信息或服务属性进行适当地修改或扭曲,使LBS服务器无法获得精确的位置信息或服务属性.由于发送给LBS服务器的是经过扭曲的信息,会导致服务质量的损失,所以必须在用户希望保护的隐私水平和他们必须接受的服务质量损失之间进行必要的权衡.但是这类技术的实现往往依赖于攻击者的先验知识,易于遭受具有数据分布特征等背景知识的攻击.差分隐私[29]对背景知识不够敏感,可应用在 LBS 隐私保护中以抵御具有任意背景知识的攻击者[30,31].基于扭曲法技术的关键是:如何在考虑攻击者背景知识和推理能力的情况下设计出最佳的隐私保护方法,从而在满足用户最大容忍服务质量的前提下,尽可能降低用户隐私的披露风险[32].

(3) 基于加密法的 LBS 隐私保护技术.

这一技术指的是通过使用加密技术使用户的 LBS 查询对 LBS 服务器完全不可见,从而达到隐私保护的目的.这类技术在确保服务质量的情况下,不会泄露任何用户的位置信息,实现了更严格的隐私保护,如基于隐私信息检索的 LBS 保护技术[33].但是,基于加密法的技术没有考虑隐私度量问题,不能实现对位置隐私的全面保护,因此也无法在隐私和服务质量损失之间做出权衡.虽然最近提出的全同态加密技术[34]可以在不解密用户查询的情况下返回正确的查询结果,但效率仍是很大的问题.最新的研究说明:由于高效的数据访问方式会暴露数据之间的顺序,所以能提供完全隐私的高效加密方法是不存在的[35].

3 类技术各有优缺点:基于政策法的技术实现简单,服务质量高,但隐私保护效果差;基于扭曲法的技术效率较高,在服务质量和隐私保护上取得了较好的平衡,但位置信息或服务属性存在一定的不准确性,易遭受具有完全背景知识的攻击;基于加密法的技术能够完全保证数据的准确性和安全性,可以提供更严格的隐私保护,但需要额外的硬件和复杂的算法支持,计算和通信开销很大.

1.3.3 LBS 隐私保护技术性能评估

LBS 隐私保护技术需要在保护用户隐私的同时兼顾服务质量与开销.本文从以下 3 个方面度量 LBS 隐私保护技术的性能:

(1) 隐私保护度:反映 LBS 隐私保护技术披露隐私多寡的程度,一般用披露风险来描述,即:攻击者根据观察到的LBS查询以及其他可用背景知识可能披露LBS隐私的概率.披露风险依赖于攻击者掌握的背景知识,攻击者掌握的背景知识越多,隐私披露风险越大.本文使用位置隐私度量指标和查询隐私度量指标来量化披露风险;

(2) 服务质量:反映用户的 LBS 查询经隐私保护技术处理后获得服务结果的好坏,一般由查询响应时间和查询准确性来度量.在相同的隐私保护强度下,用户获得的服务质量越高,说明隐私保护技术越好;

(3) 开销:反映使用LBS隐私保护技术所带来的代价,包括预处理和运行时发生的存储、计算和传输代价.存储代价主要发生在预处理时;现有技术下,预处理代价通常较小,在选择隐私保护技术时被忽略;运行时的计算和传输代价一般使用隐私保护技术实现算法的时间复杂度和通信协议的通信复杂度来度量.LBS 隐私保护技术要在满足隐私保护度和服务质量的前提下尽量减少开销.

2 LBS 隐私保护系统结构

LBS 隐私保护技术以在线或离线的方式和不同的体系结构来实现:离线方式中,所有 LBS 查询中的时空信息对 LBS 隐私保护技术都是可用的;在线方式中,对 LBS 查询信息的修改是在用户不同时间访问新位置时实时执行的[23].LBS 隐私保护技术一般通过 3 种系统结构实现:集中式、分布式和混合式.

- 集中式结构由移动终端、可信匿名服务器、LBS 服务器组成,如图 2 所示.

用户使用移动终端向 LBS 服务器发送 LBS 查询,并获得最终的查询结果.可信匿名服务器包含匿名处理模块和查询结果精炼模块:匿名处理模块把移动终端发送过来的精确位置模糊化,并转发给 LBS 服务器;查询结果精炼模块接收 LBS 服务器返回的结果集对其进行精炼,并将精炼后的最终结果返回给移动终端.集中式结构具有用户的全局信息,隐私保护效果好,移动终端和匿名服务器之间的通信开销较小.但缺点是:① 匿名服务器可能成为系统的性能瓶颈和唯一攻击点;② 匿名服务器拥有所有用户的位置信息或服务属性等完全知识,一旦匿名服务器被攻破,可能会带来严重的隐私威胁;③ 现实中,部署具有大量用户的可信匿名服务器非常困难.

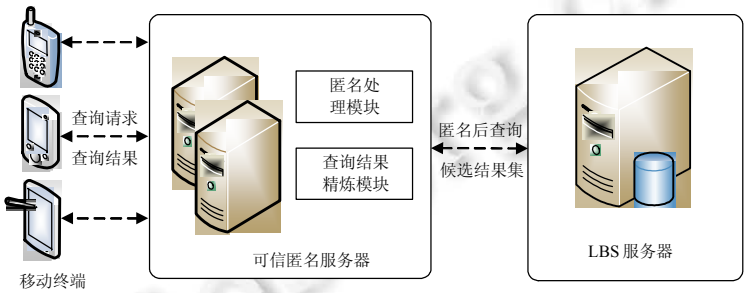

Fig.2 Centralized anonymization server-based architecture

图 2 集中式结构

- 分布式结构由移动终端和 LBS 服务器组成,如图 3 所示.

移动终端之间通过 P2P 协议,利用单跳和多跳通信形成一个匿名组,查询用户模糊化其位置为包含组内所有用户的空间区域,并将其转发给服务器,LBS 服务器返回包含正确结果的候选集,用户之间通过彼此协作完成隐私保护.分布式结构的优点:① 消除了系统的性能瓶颈;② 具有用户的全局信息,隐私保护效果好.缺点:① 增加了移动终端的通信和计算开销,在实际应用中无法有效保证参与隐私保护的其他用户是可信的;② 当服务请求用户附近没有足够的对等用户时,匿名过程很难完成.

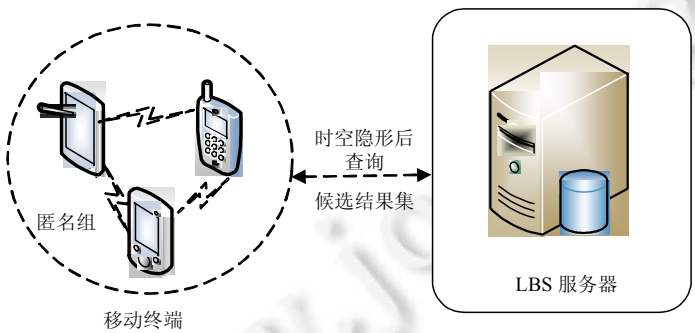

Fig.3 Distributed mobile device-based architecture

图 3 分布式结构

- 混合式结构由移动终端、可信匿名服务器、LBS 服务器组成,如图 4 所示.

移动终端通过可信匿名服务器请求服务,也可基于个性化的隐私、响应时间以及服务质量需求使用 P2P 协议完成隐私保护.匿名服务器拥有用户的身份、服务请求、位置等完全知识.混合式结构集成了集中式和分布式结构的优点,能够很好地平衡客户端和匿名服务器之间的负载[36]减少了匿名服务器由大量移动终端位置更新导致的负荷;在用户分布稀疏的情况下,仍能保证服务的可用性;但缺点是系统参数众多,设置和调整非常复杂,严重影响了它的实用性.