RSA CONFERENCE
CHINA 2012
RSA信息安全大会2012

THE GREAT CIPHER

MIGHTIER THAN THE SWORD

伟大的密码胜于利剑

基于移动互联网络的位置服务隐私保护

演讲人姓名：王丽娜

演讲人公司：武汉大学

RSA CONFERENCE CHINA 2012

RSA信息安全大会2012

专题会议主题：

专题会议分类：

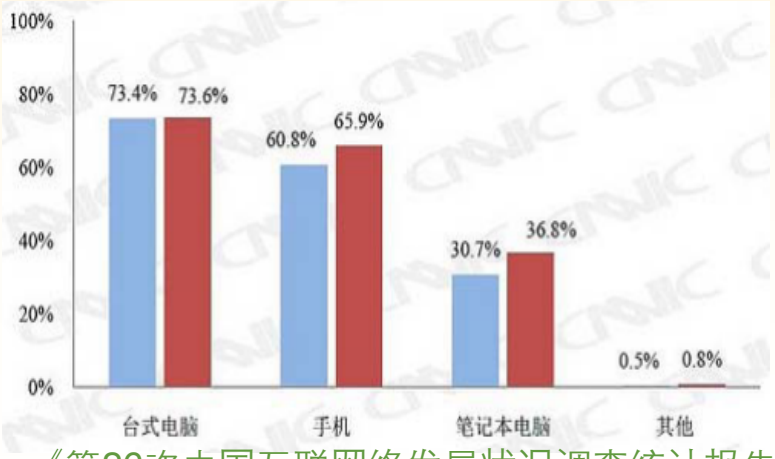
大纲

1. 移动网络时代的位置服务
2. 位置服务的特征及内涵
3. 位置服务中存在的安全问题
4. 位置服务的用户隐私保护

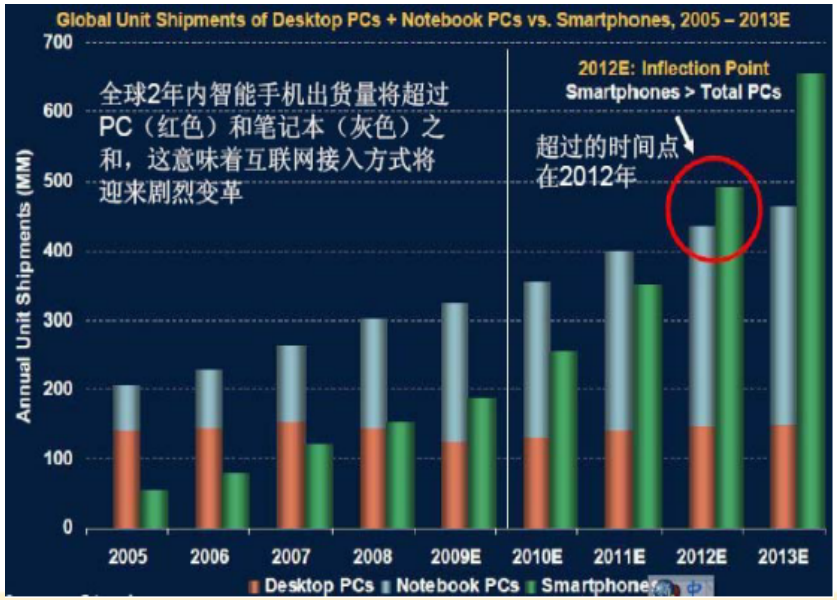
一、移动网络时代的位置服务

1.1 移动互联网时代的来临

- 2011年全球移动互联网大会：截止到2010年底，中国手机网民规模达到3.03亿，占网民总数的66.2%，较2009年底增加了6930万人。2011年第一季度，中国移动互联网市场规模达64.4亿元人民币，同比增长43.4%，环比增长23%。

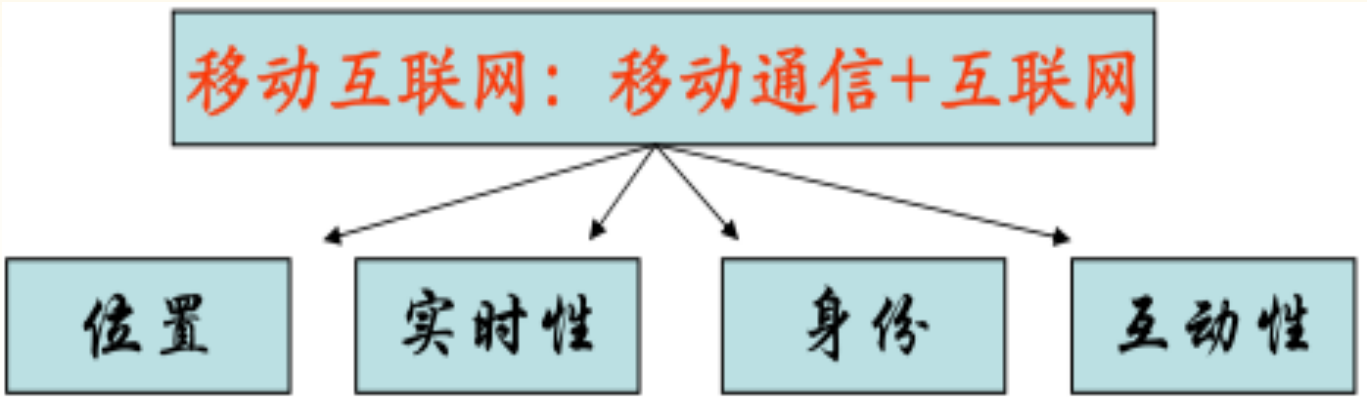


《第26次中国互联网络发展状况调查统计报告》2011♪

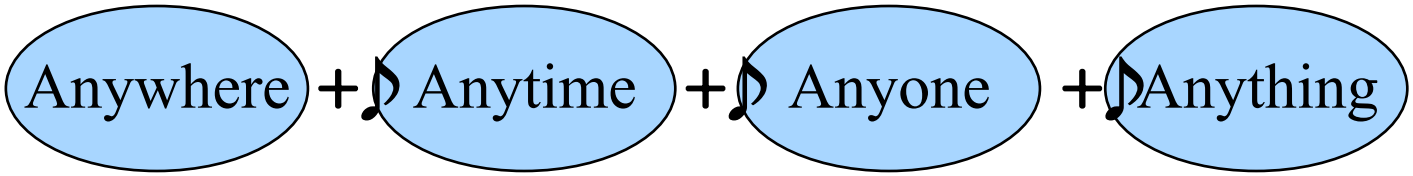


1.2 移动性：位置知识的天然反映

移动性：



- 以上的移动性是与传统互联网区别的关键
- 位置和移动性的天然结合，相互促进

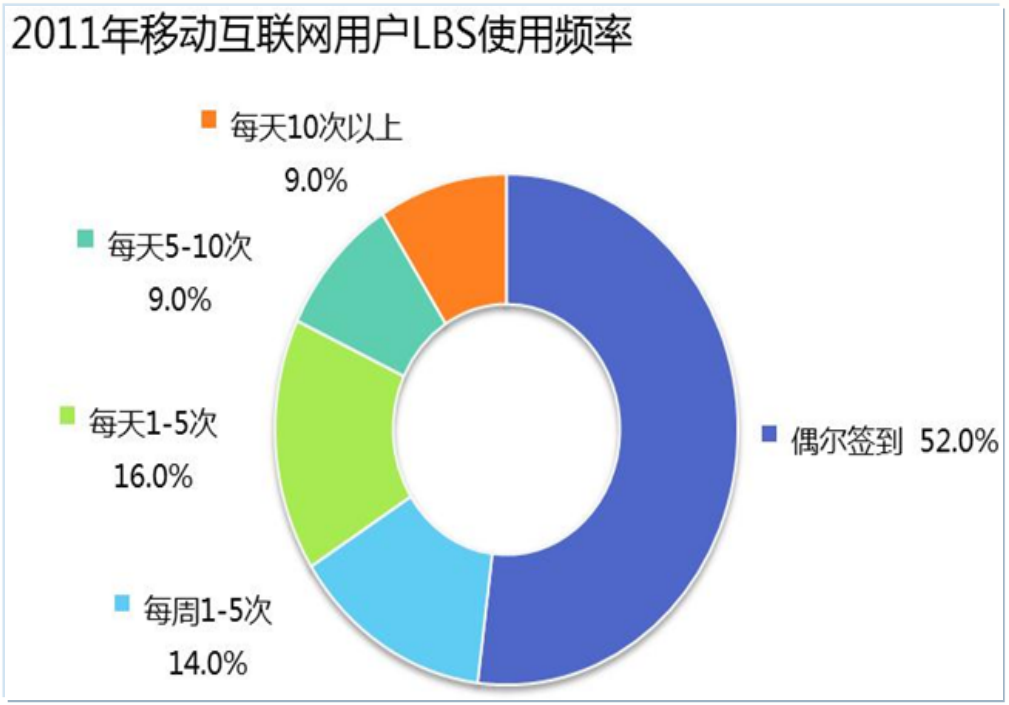


位置服务的目标：4A

1.3 位置服务成为移动互联网的标配

- 2011年8月，第10届中国互联网大会：位置服务的前景被一致看好。位置服务将会成为新互联网时代的标准配置。

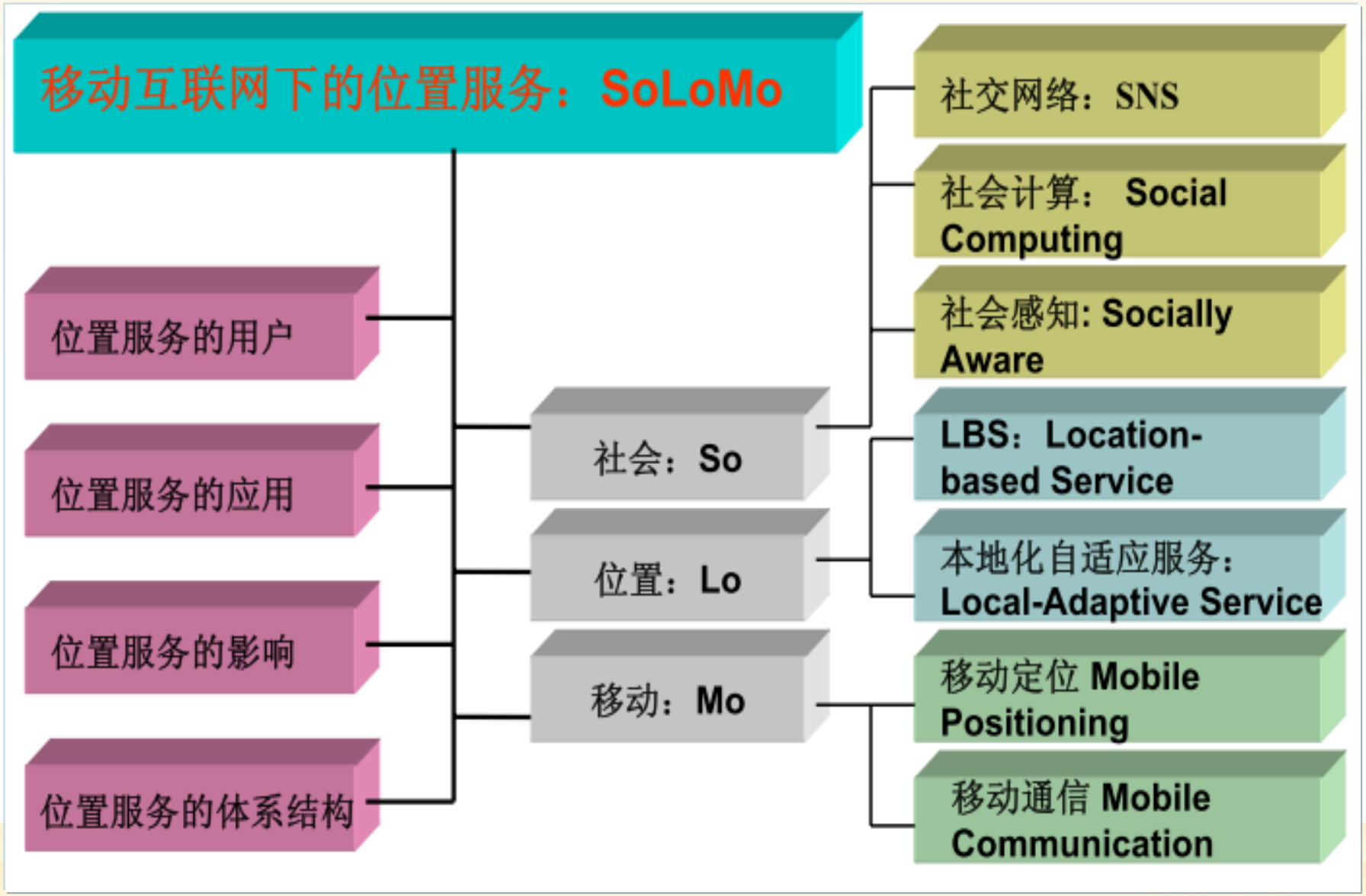
  - 位置服务与SNS等移动互联网应用相结合，形成了一种普遍认同的新发展模式：O2O（online to offline）。



易观国际 研究数据，2011

1.4 移动互联网下的位置服务

RSA CONFERENCE CHINA 2012

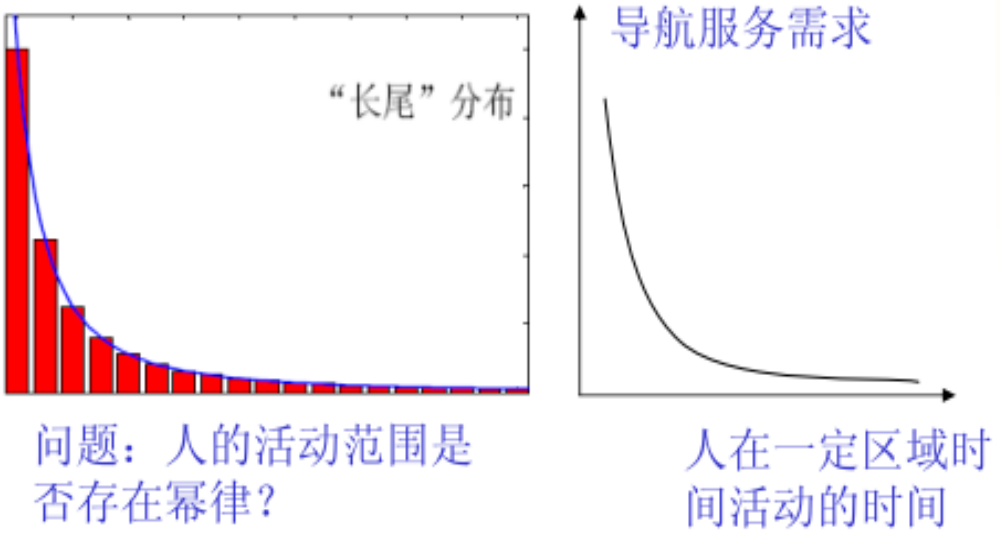


二、位置服务的特征及内涵

2.1基于位置服务(LBS)的含义

- LBS包括两层含义：首先是确定移动设备或用户所在的地理位置；其次是提供与位置相关的各类信息服务。

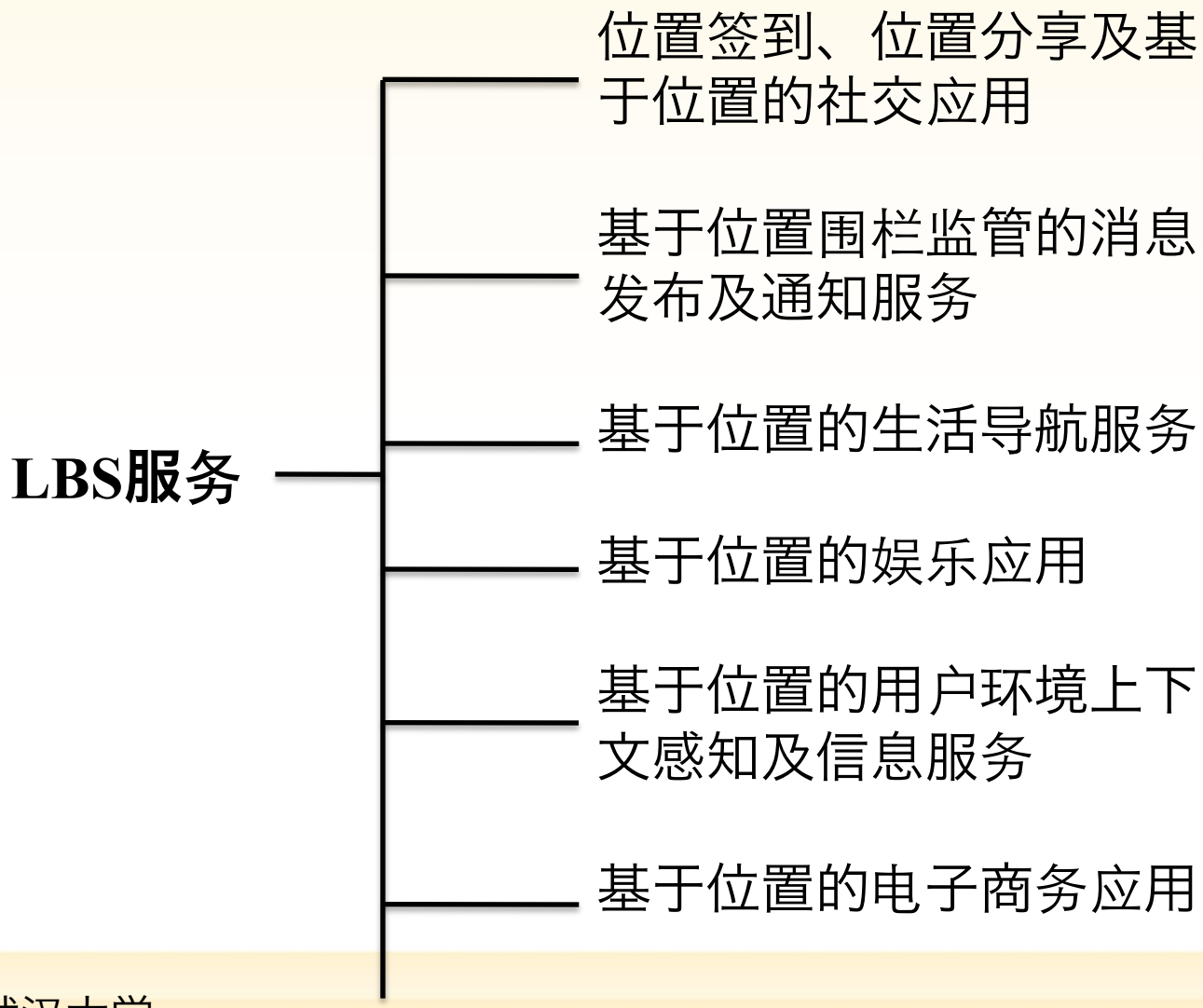
- 传统“导航”服务的先天缺陷：
  - 缺乏服务持久性
  - 用户黏性



社交→基于位置的社交
电子商务→基于位置的电子商务
搜索引擎→基于位置的搜索引擎
游戏→基于位置的游戏
广告→基于位置的广告 …

在一定程度上“位置”已经不是“位置服务”的内容，而是构成服务的输入性关键性因素。

2.2位置服务的应用分类



2.3典型应用

Foursquare

Foursquare是一个基于地理位置、对用户进行定位，融合了Twitter、LBS、趣味性和商家点评等概念的**新型Mobile SNS服务。**

SK Telecom

用户在超市里移动的时候，推车上的屏幕会针对用户所在的位置（误差小于1米）来推送商品信息和优惠券。

虚拟超市

使用手机扫描所选商品，并通过手机在网上结算，超市就会将顾客所购产品按时送到其家中。

2.4 位置服务模式逐步成熟

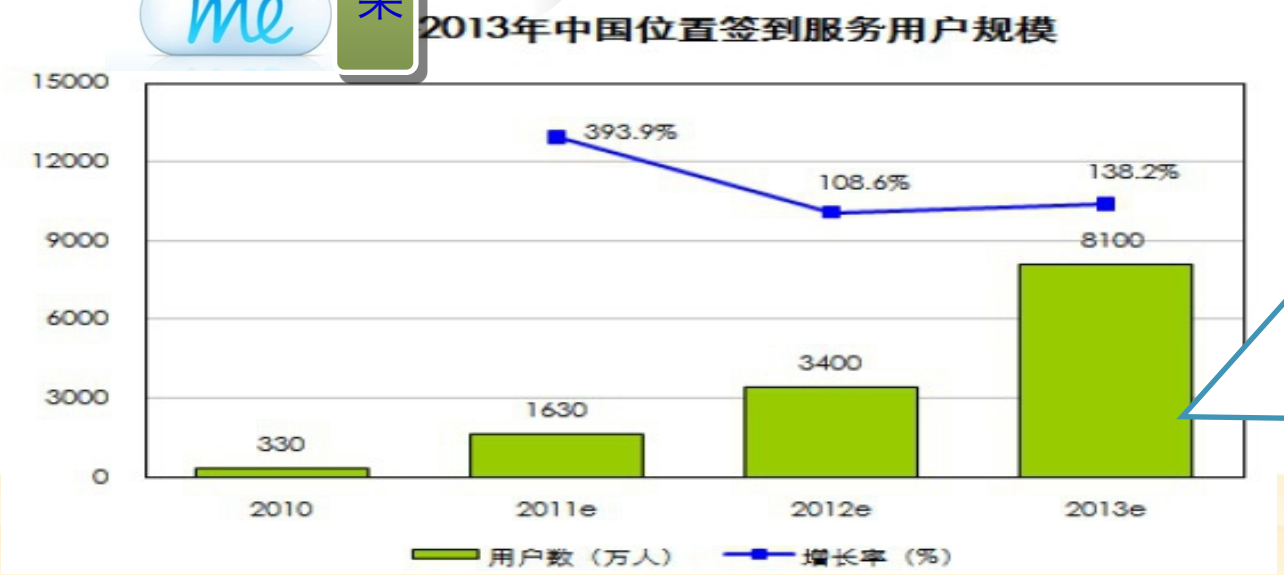
谷歌

安卓

苹果

- 智能终端的日益普及
- 无线互联网快速发展
- 位置服务市场日渐成熟

位置服务与移动社交、移动支付相结合产生新兴模式应用更加广泛功能更加灵活



预计2013年用户规模将达到8100万亿，市场增长率将达到138%。

2.4 位置服务模式逐步成熟

谷歌

安卓

苹果

- 智能终端的日益普及
- 无线互联网快速发展
- 位置服务市场日渐成熟

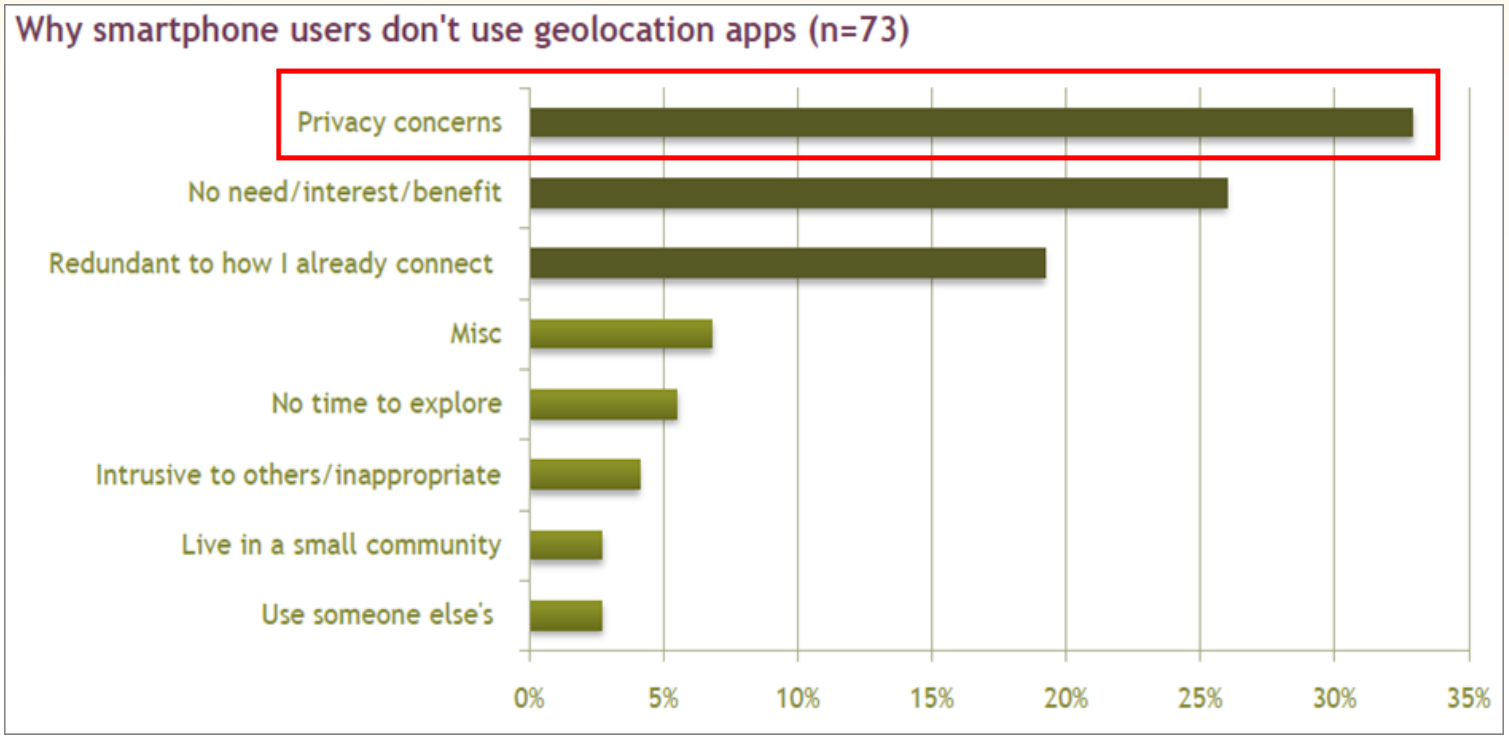
位置服务与移动社交、移动支付相结合产生新兴模式应用更加广泛功能更加灵活

位置服务必将给地球空间信息全新应用模式及新型业务拓展带来历史机遇和巨大的发展空间。

达到138%。

三、位置服务中存在的安全问题

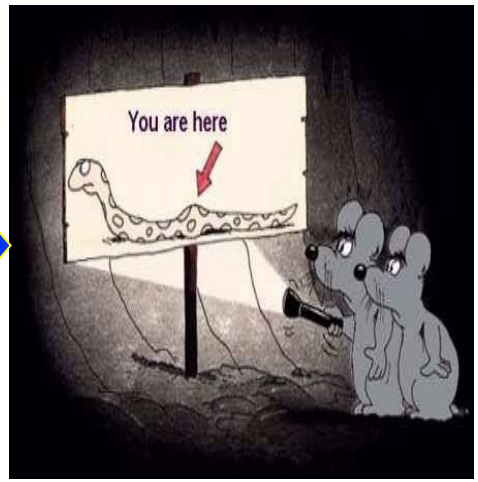
3.1 LBS用户最担心的问题？



该调查结果显示，影响LBS广泛普及的一个主要原因是人们对隐私问题的担心。其中最大的担心是害怕未经允许向不明组织或其他人透露自的己位置信息。

3.2 LBS隐私泄露风险

追踪...!!!!



“新的技术可以随时随地的定位你的位置. 虽然他们承诺安全和便捷, 但依然存在隐私和安全威胁”

Cover story, IEEE Spectrum, July 2003

3.2 LBS隐私泄露风险（案例）

FOX NEWS .com — U.S. & WORLD

Updated: 3:28:06 9:42pm ET

SEARCH GO

E-MAIL STORY | PRINTER FRIENDLY | FOXFAN CENTRAL

Man Accused of Stalking Ex-Girlfriend With GPS

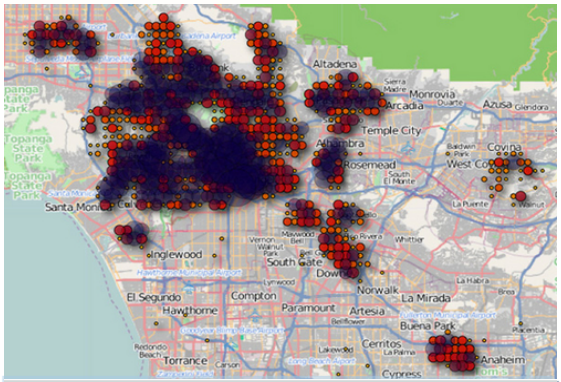
Saturday, September 04, 2004
Associated Press

GLENDALE, Calif. — Police arrested a man they said tracked his ex-girlfriend's whereabouts by attaching a global positioning system (search) to her car.

Ara Gabrielyan, 32, was arrested Aug. 29 on one count of stalking (search) and three counts of making criminal threats. He was being held on $500,000 bail and was to be arraigned Wednesday.

"This is what I would consider stalking of the 21st century," police Lt. Jon Perkins said.

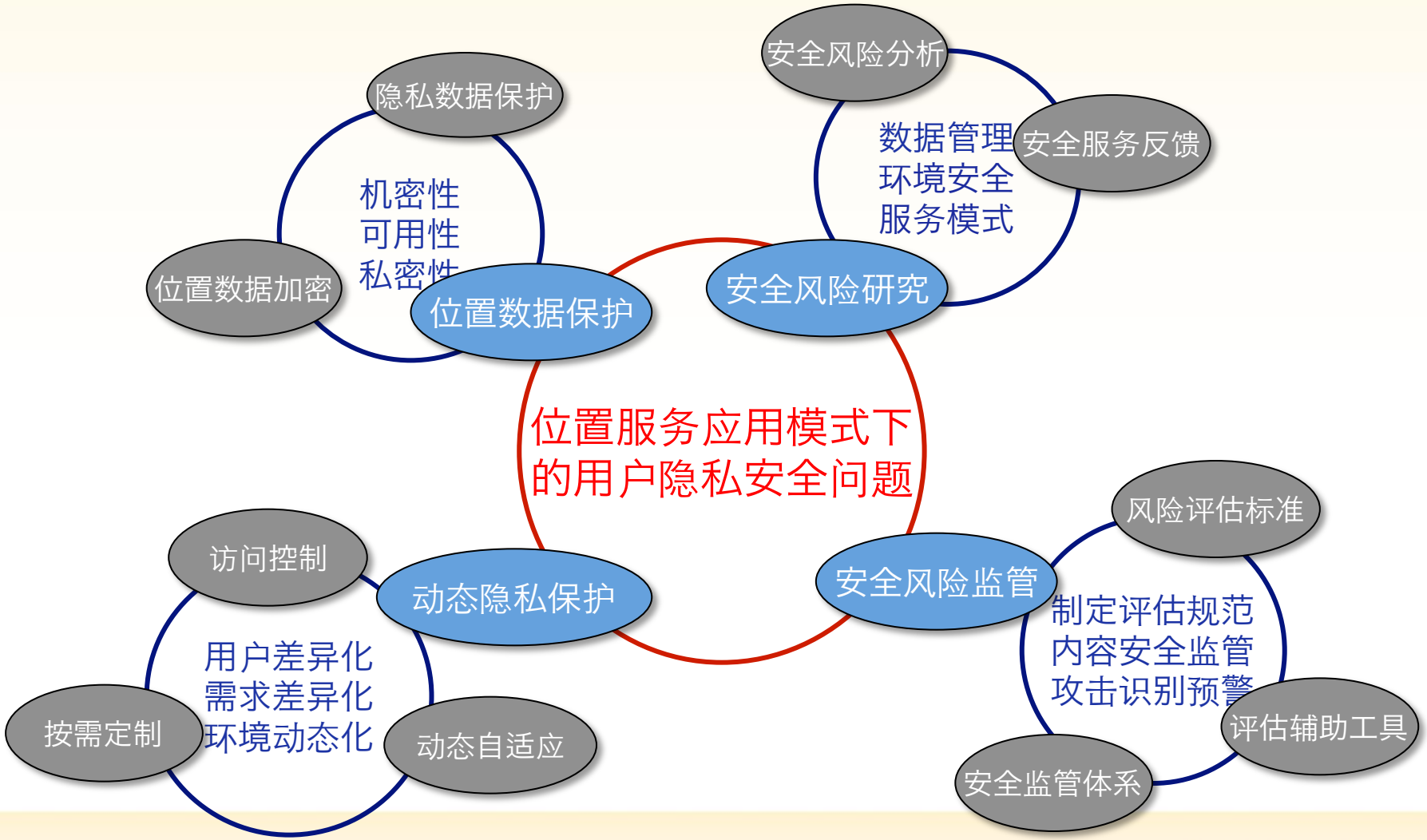
美国福克斯新闻报道：一男子通过GPS定位对其女友进行跟踪



多家媒体报道苹果iPhone 4以及iPad记录用户行踪

国外媒体统计：78%的小偷使用facebook、twitter来定位目标；15%的小偷因社会化媒体知道屋内没人

3.3 LBS隐私安全问题



3.4 LBS隐私安全威胁

隐私：用户过去或现在的位置、活动的轨迹、查询的内容

位置服务市场庞杂，缺乏全面的隐私安全保障

移动互联网络环境下的隐私安全问题是位置服务推广应用的核心问题之一

四、位置服务的用户隐私保护

4.1 LBS隐私分类

- 隐私风险通常可以分为四类
  - 关联风险：
    - 用户身份和位置之间的对应关系导致位置信息敏感。

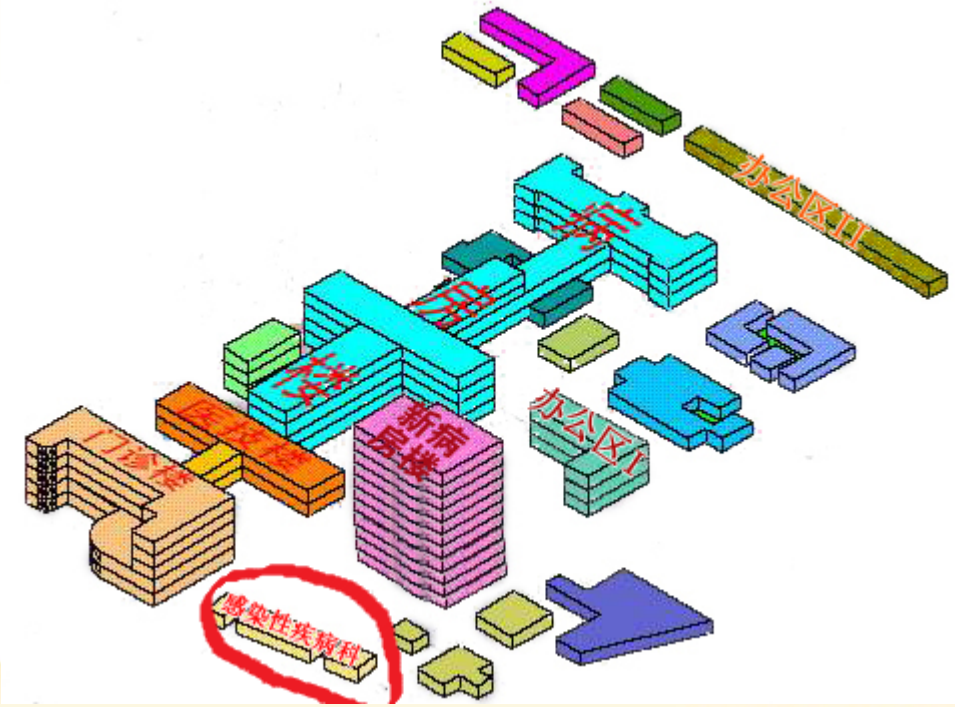
如：攻击者获取用户的位置在第六医院，则可以推测其可能患病或者去看病人。

4.1 LBS隐私分类

RSA CONFERENCE CHINA 2012

- 隐私风险通常可以分为四类
  - 质量风险：
    - 信息的精度越高（时间精度、空间精度、查询精度），用户等信息所蕴含的用户隐私越多；

如：用户在医院内的使用位置的精度较高，则更会暴露所处的科室（如：感染病科）。



4.1 LBS隐私分类

- 隐私风险通常可以分为四类
  - 发现风险：
    - 用户的位置信息未经授权而被他人获得；

如：用户在连续使用位置服务的过程中，其多个位置信息被攻击者所获取，则攻击者可以根据所获取的位置点推测出大概的行动轨迹，从而造成更大的隐私危害。

4.1 LBS隐私分类

- 隐私风险通常可以分为四类
  - 数量风险：
    - 多个个体的位置信息泄露会导致整个部署方案的泄露。

如：攻击者获取某物流公司多数员工的行动轨迹，则可以推测出该公司的整体商业部署，从而造成商业秘密的泄漏。

4.2 LBS隐私保护目的

- 位置隐私保护
  - 避免用户与某一精确位置匹配
- 查询隐私保护
  - 避免用户与某一敏感查询匹配

隐私保护不是指要保护用户的个人信息不被他人使用，而是指用户对个人信息进行有效控制的权利。

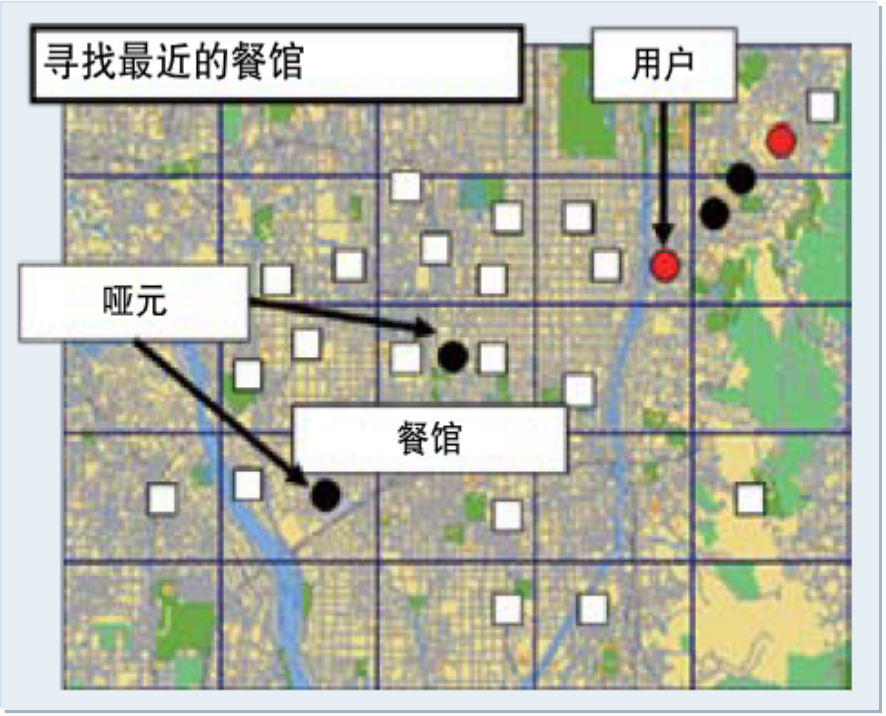
4.3 LBS隐私保护方法

- 假位置
- 时空匿名
- 空间加密

4.3 LBS隐私保护方法

➢假位置

通过制造假位置，达到以假乱真的效果。移动对象数据库中的查询处理器无需作任何修改。

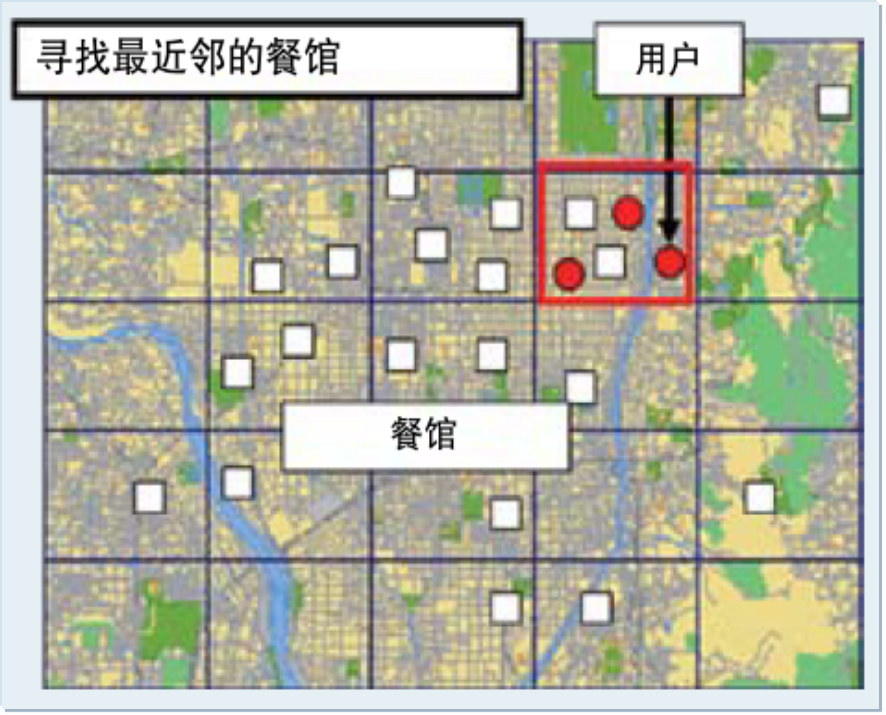


如：用户寻找最近的餐馆。白色方块是餐馆位置，红色点是用户的真实位置。当该用户提出查询时，为其生成两个假位置，即哑元（如图中的黑色点）。真假位置一同发送给服务提供商。从攻击者的角度，同时看到三个位置，无法区分哪个是真实的哪个是虚假的。

4.3 LBS隐私保护方法

➢时空匿名

将一个用户的位置通过扩展变成时空区域，达到匿名的效果。设计基于区域位置的查询处理技术；查询结果是一个包含真实结果的超集

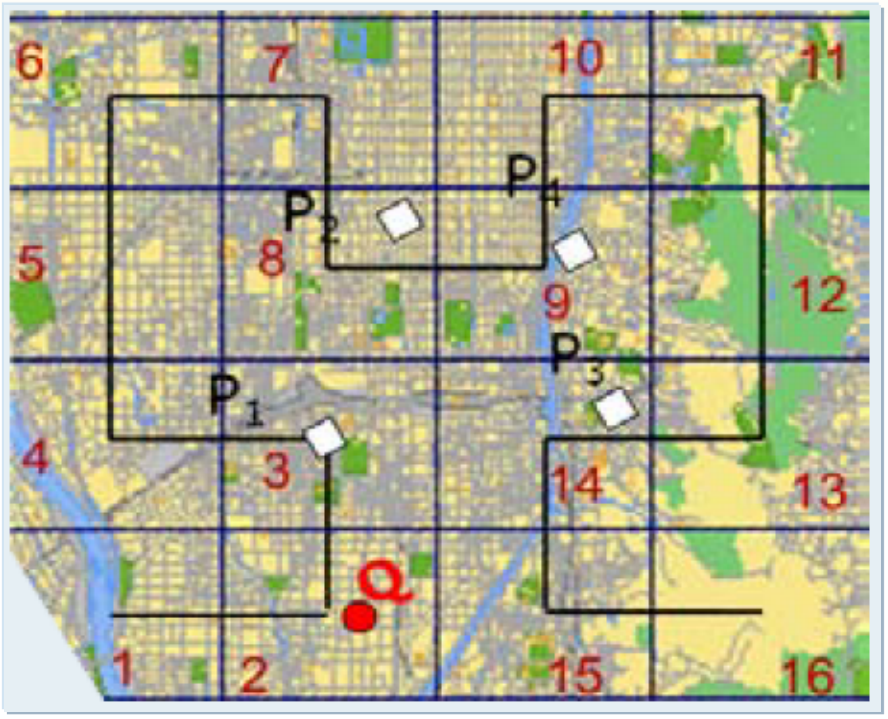


如：延续上图寻找餐馆的例子，当用户提出查询时，用一个空间区域表示用户位置，如图中的红色框。从服务提供商角度只能看到这个区域，无法确定用户是在整个区域内的哪个具体位置上。

4.3 LBS隐私保护方法

➢空间加密

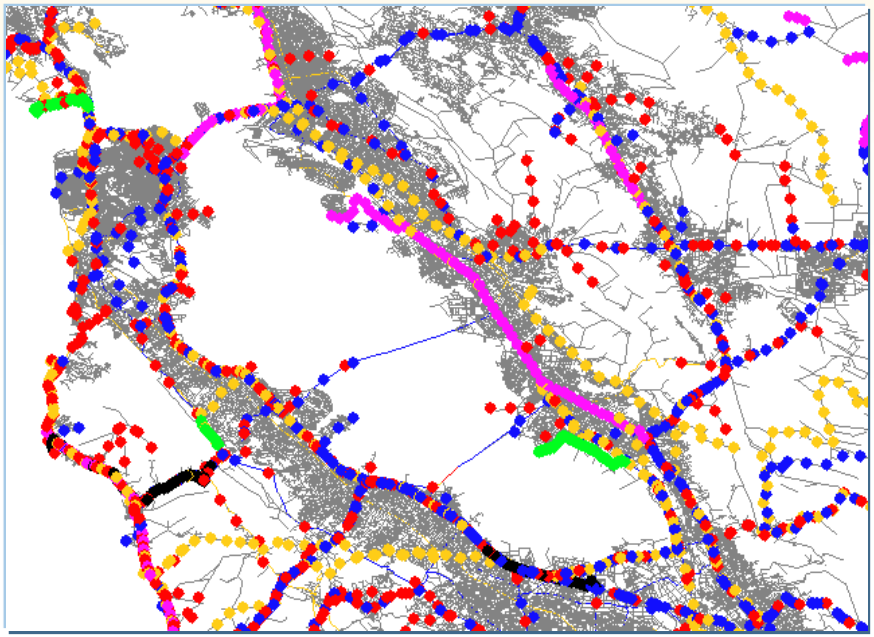
通过对位置加密从而达到匿名的效果。查询方法与使用的加密协议有关。



继续前面的例子，首先将整个空间旋转一个角度，在旋转后的空间中建立希尔伯特（Hilbert）曲线。每一个被查询点P（即图中的白色方块）对应的希尔伯特值如该点所在的方格数字所示。当某用户提出查询Q时，计算出加密空间中Q的希尔伯特值。在此例子中，该值等于2。寻找与2最近的希尔伯特值所对应的P，即P1。将P1返回给用户。

4.4 LBS隐私保护中面临的挑战(1)

- 随着移动设备的使用快速增长，存储时空数据的价值是显而易见的。商业公司、政府以及科研院所大量的收集和存储时空数据，以获取有价值的信息。



多辆车在某城市的行驶状态

- 然而，由于轨迹数据库包含大量的个人信息，发布这些数据库需要受到隐私规则的限制。

4.4 LBS隐私保护中面临的挑战(2)

- 简单的删除和使用假名替换用户身份的方法，不能有效的保护用户的隐私信息。如下图：

t_{id}	trajectory
t_1	$a_1 \rightarrow b_1 \rightarrow a_2$
t_2	$a_1 \rightarrow b_1 \rightarrow a_2 \rightarrow b_3$
t_3	$a_1 \rightarrow b_2 \rightarrow a_2$
t_4	$a_1 \rightarrow a_2 \rightarrow b_2$
t_5	$a_1 \rightarrow a_3 \rightarrow b_1$
t_6	$a_3 \rightarrow b_1$
t_7	$a_3 \rightarrow b_2$
t_8	$a_3 \rightarrow b_2 \rightarrow b_3$

(a)

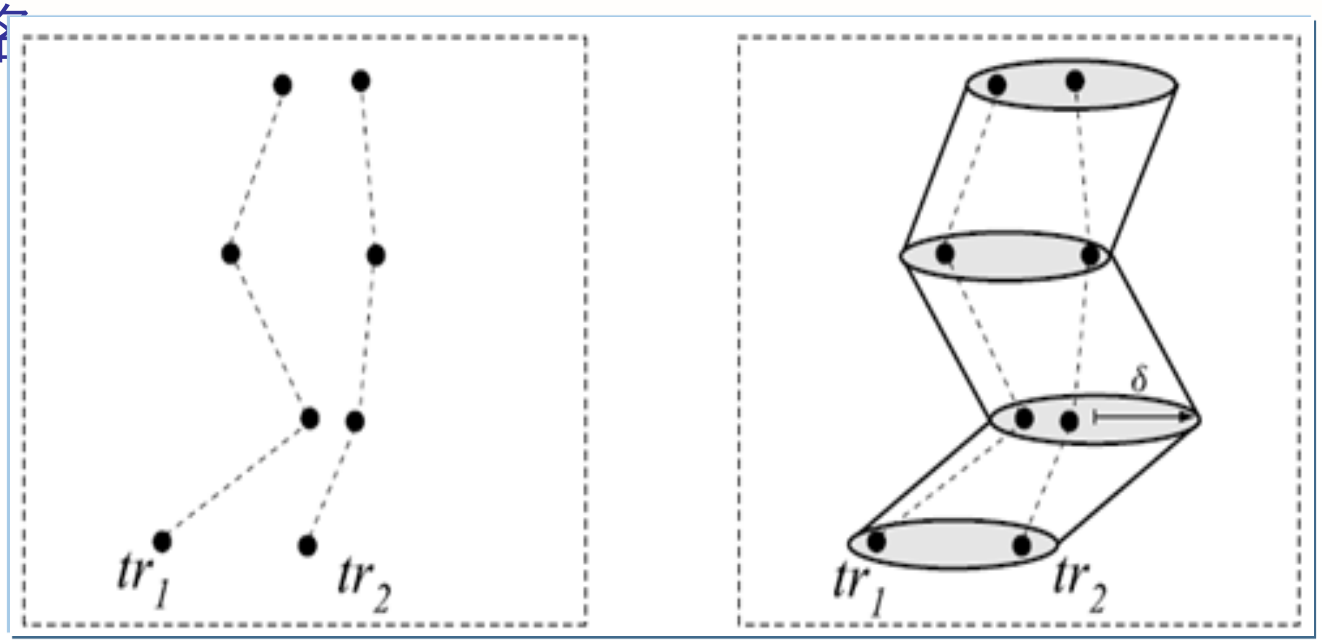
t_{id}	trajectory
t_1^A	$a_1 \rightarrow a_2$
t_2^A	$a_1 \rightarrow a_2$
t_3^A	$a_1 \rightarrow a_2$
t_4^A	$a_1 \rightarrow a_2$
t_5^A	$a_1 \rightarrow a_3$
t_6^A	a_3
t_7^A	a_3
t_8^A	a_3

(b)

图(a)为轨迹数据库D，（b）为对手A所掌握的知识——轨迹片段。则A在只掌握片段轨迹a_1,a_3的前提下，可以唯一的确定t_5的下一个位置点为b_1。

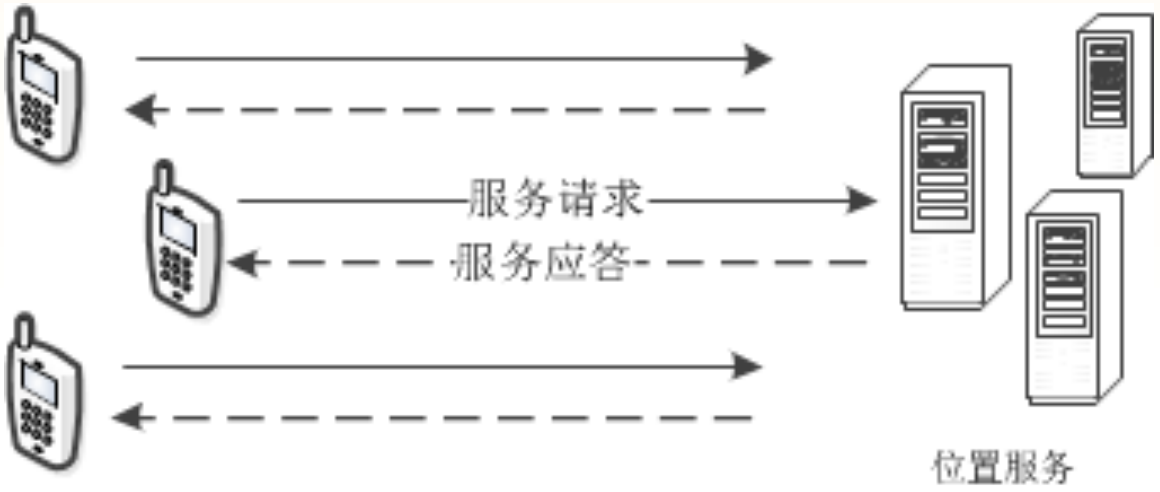
4.4 LBS隐私保护中面临的挑战(3)

- 目前，匿名是应用最为广泛的隐私保护技术。其基本的思想为：匿名数据集中每一条轨迹，使得任一条轨迹不能和其他K-1条轨迹区分开。下图为K=2时的轨迹匿名



4.5 LBS隐私保护系统架构

（1）独立式架构

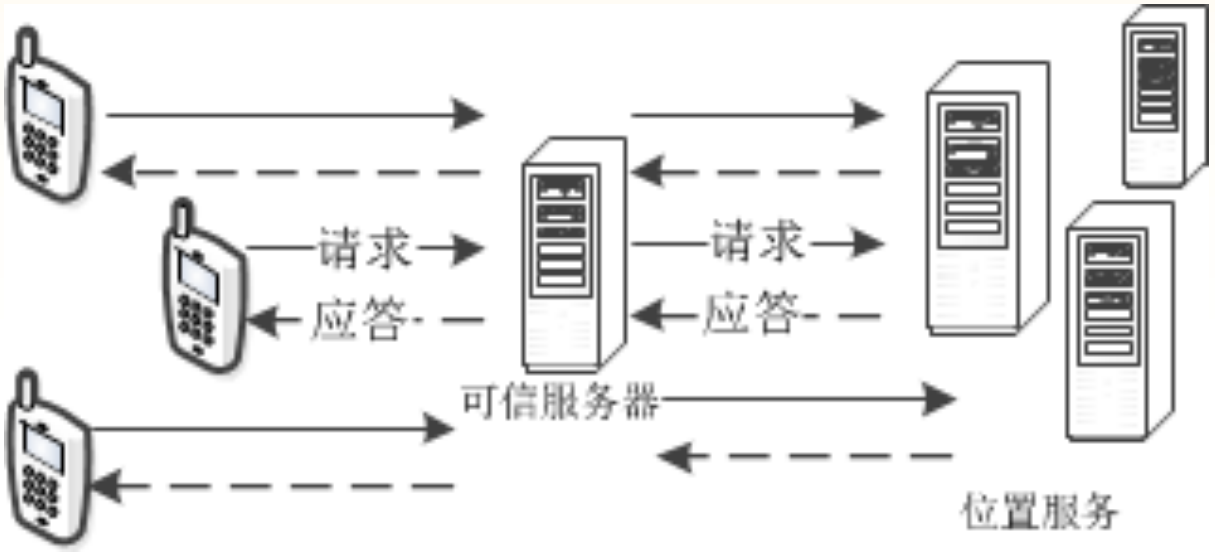


优点：结构简单，易于配置

缺点：增加客户端负担，缺乏全局信息

4.5 LBS隐私保护系统架构

（2）中心服务器架构

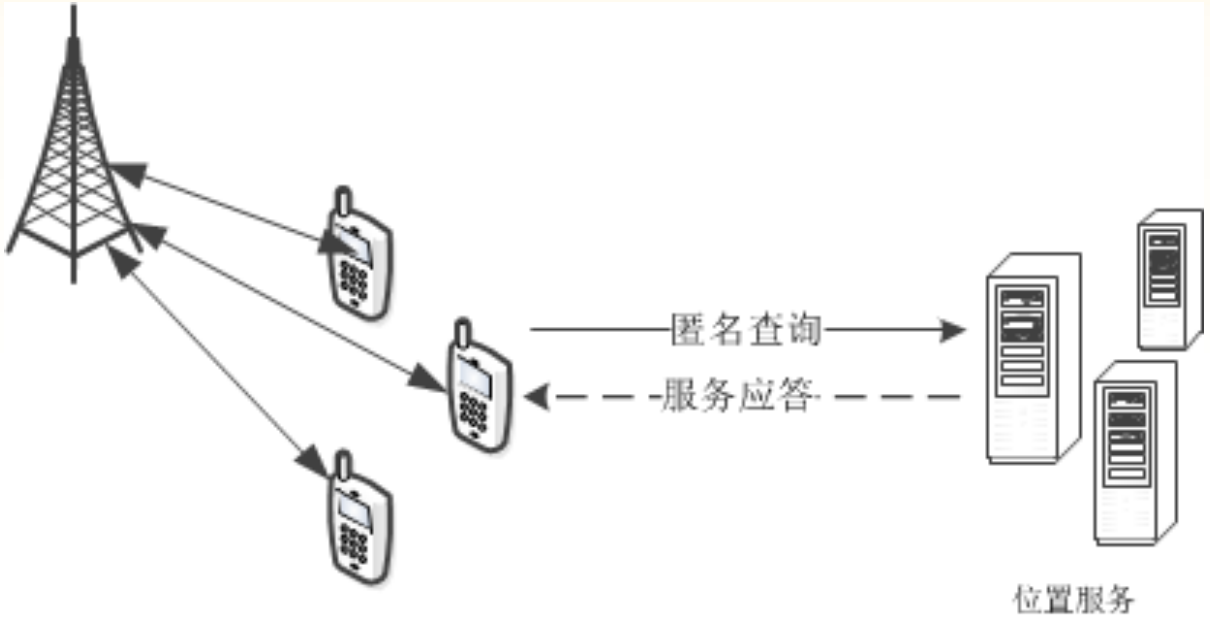


优点：拥有全局信息，隐私性好

缺点：容易成为系统瓶颈，成为攻击点

4.5 LBS隐私保护系统架构

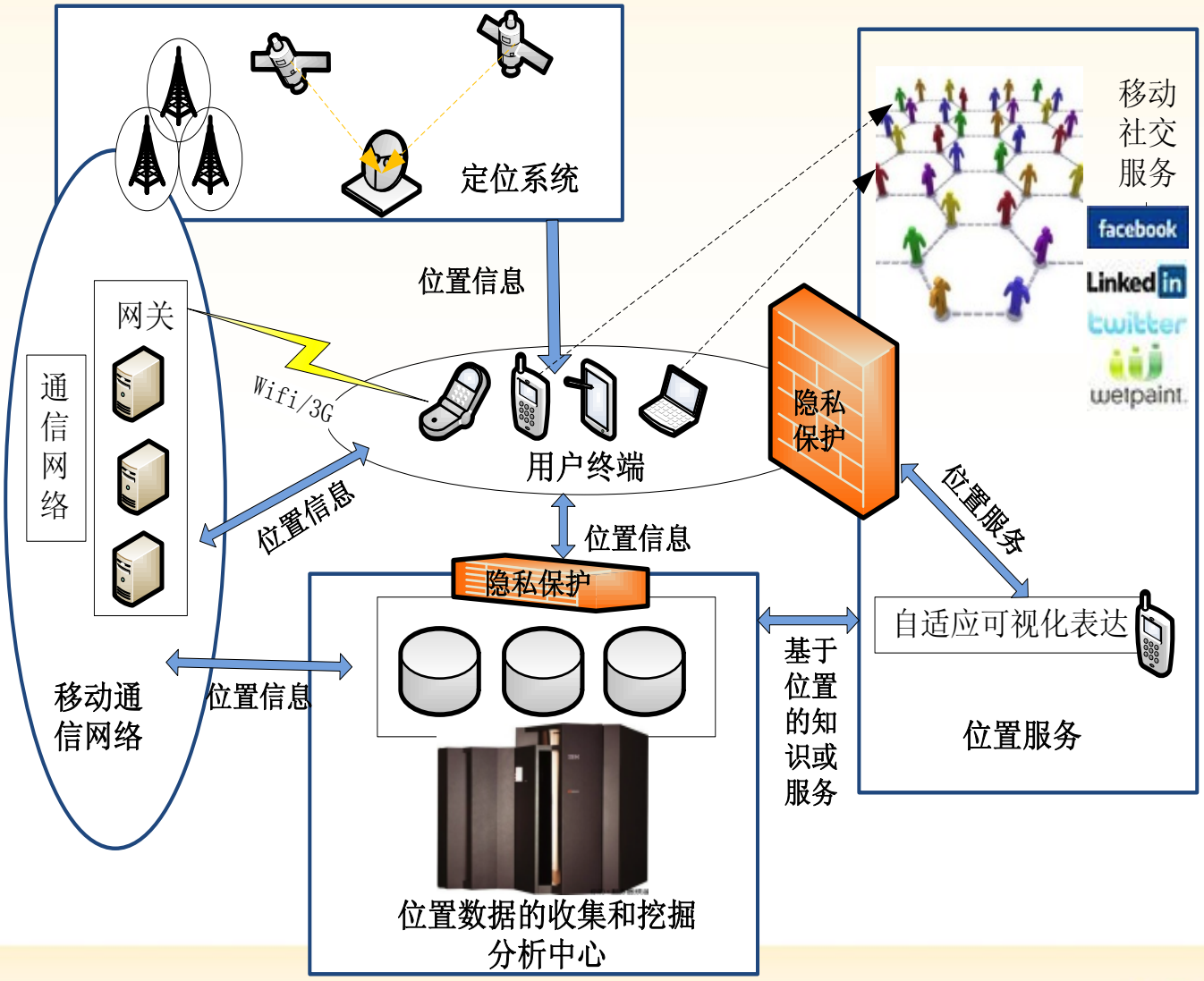
（3）合作式架构



优点：具有全局信息，消除系统瓶颈

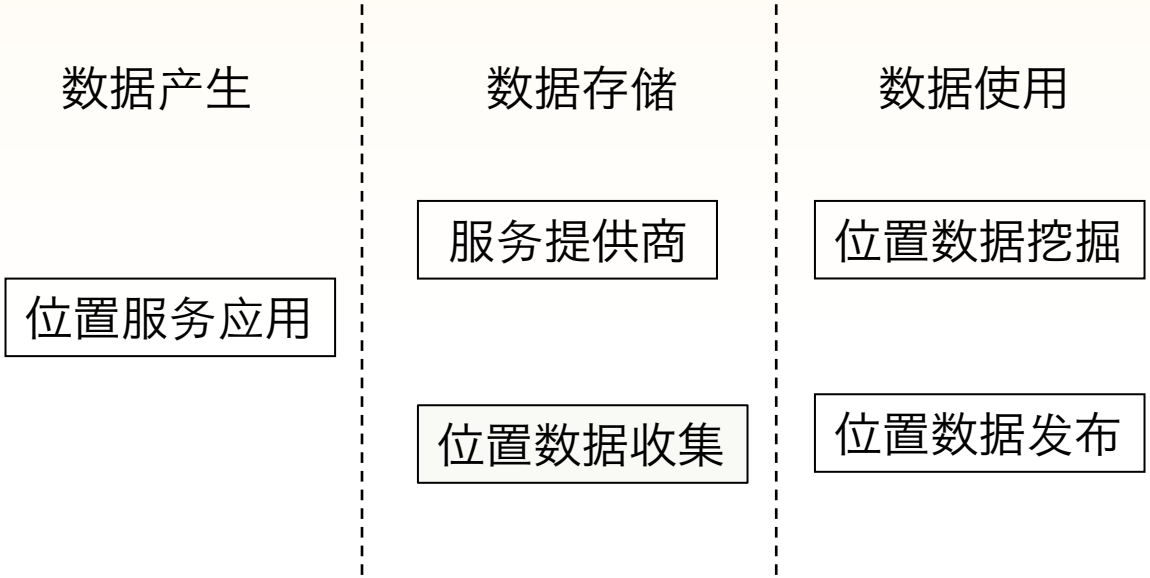
缺点：通讯代价高

4.6 LBS隐私保护总体框架



4.7 轨迹隐私保护

- 从上图中可以发现轨迹数据的生命周期为：



- 隐私保护处理越处于生命周期的前端则其泄露隐私的概率也越低。接下来将详细介绍一种位置轨迹收集的隐私保护方法。

4.8 轨迹收集隐私保护系统

- **环　境：**

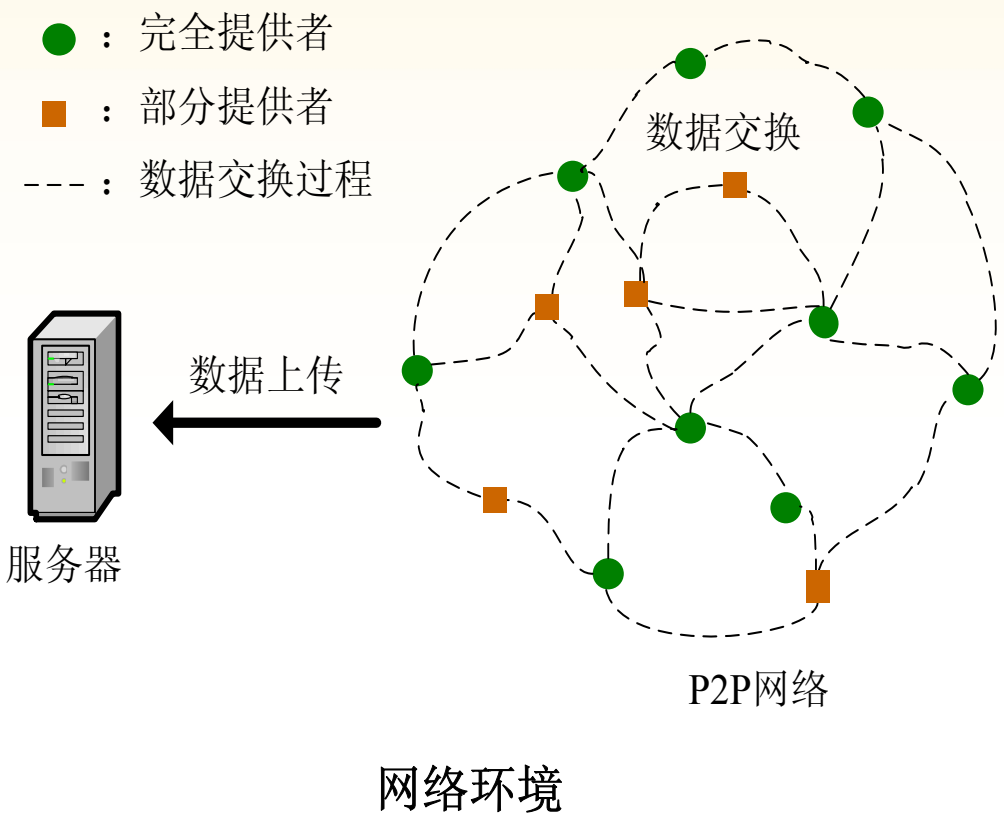
　无线P2P网络(蓝牙、Zigbee等)

- **完全提供者：**

　隐私需求较低，提供真实的位置数据

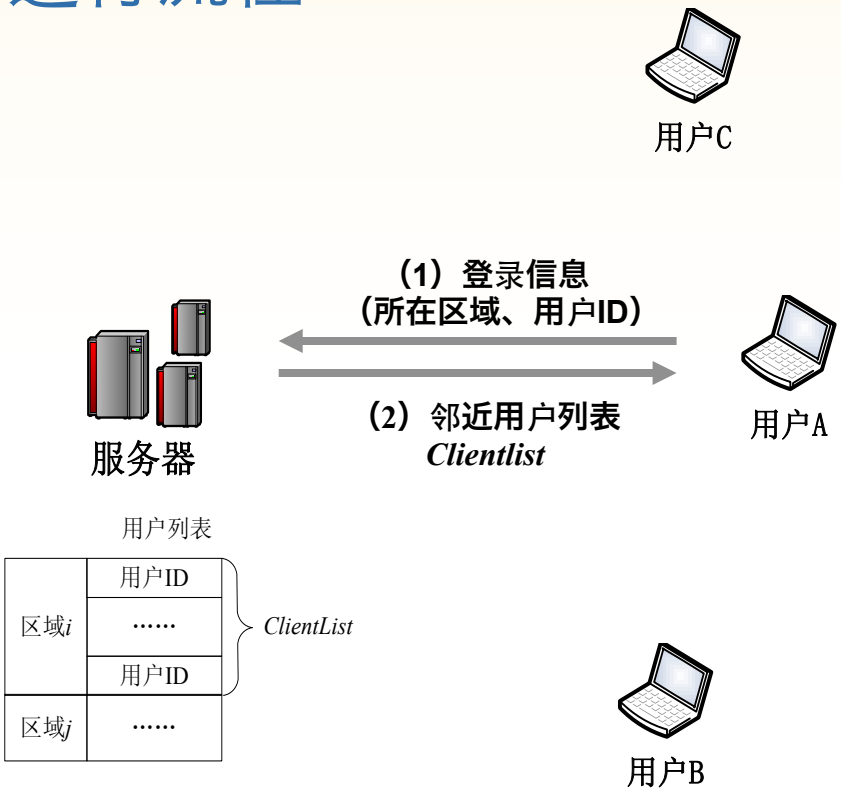
- **部分提供者：**

　隐私需求较高，不提供位置数据，仅参与数据交换帮助其它用户隐匿。

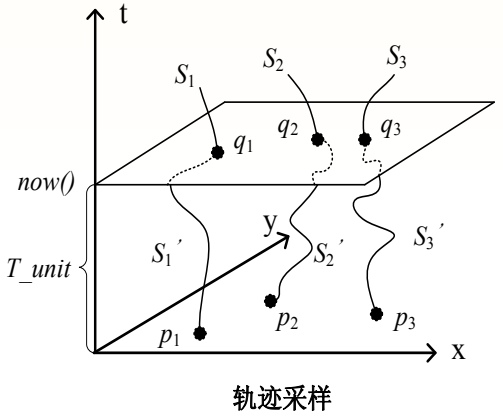


位置数据收集K-匿名模型

运行流程：



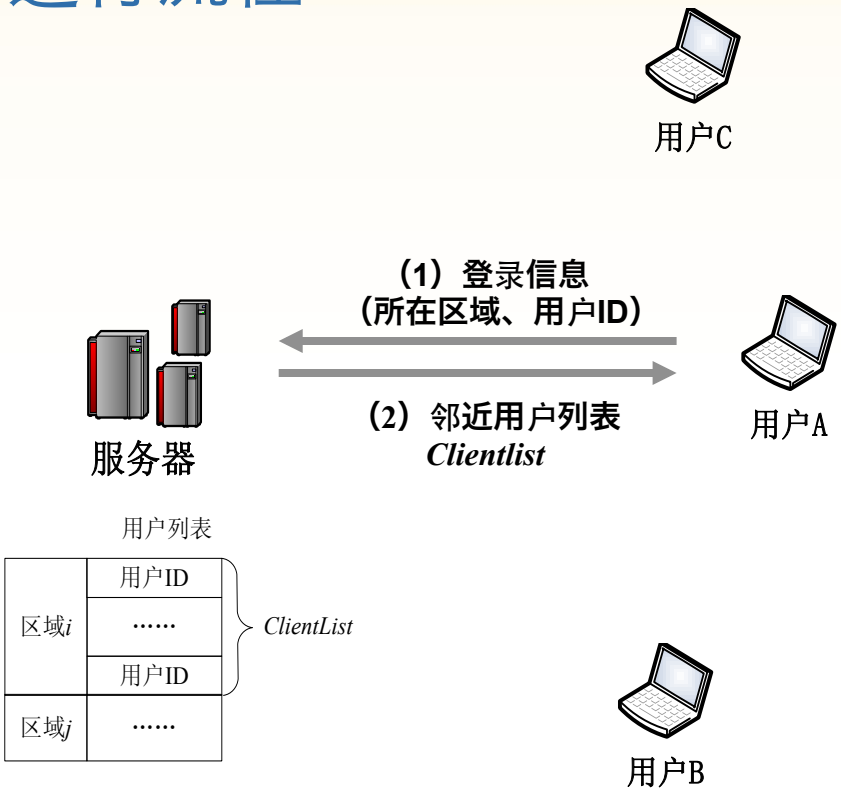
轨迹采样



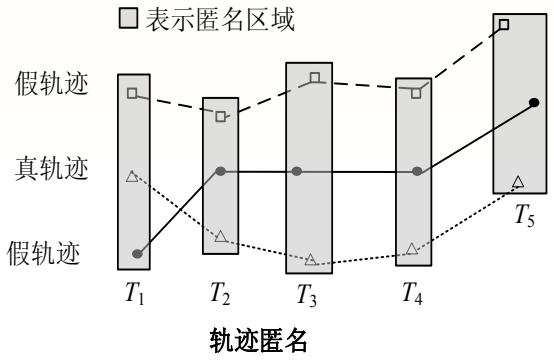
轨迹采样

位置数据收集K-匿名模型

运行流程：

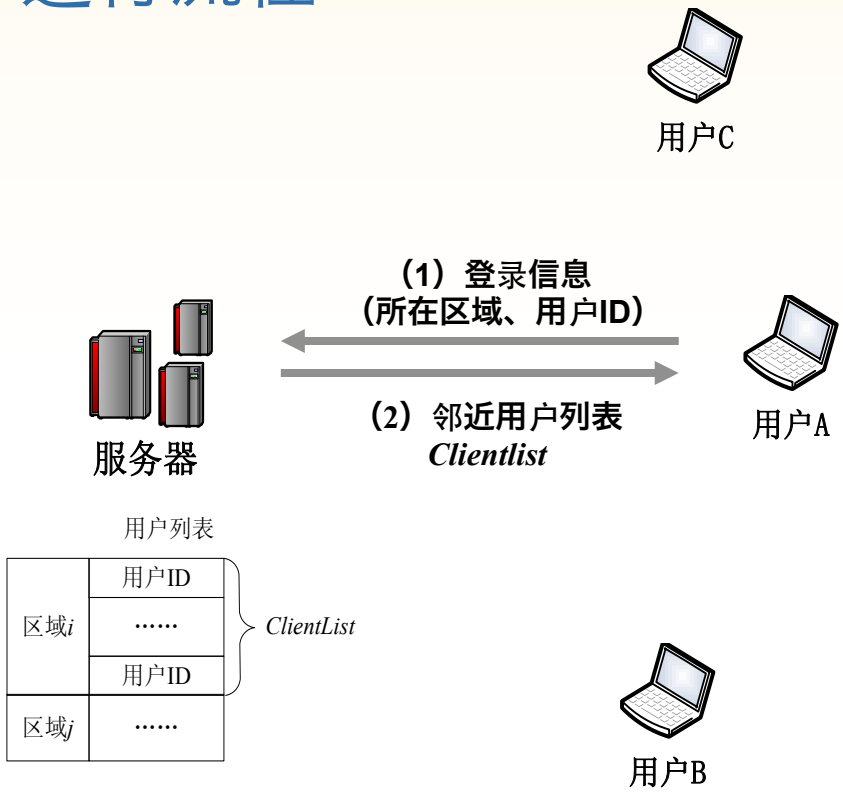


轨迹匿名



位置数据收集K-匿名模型

运行流程：



轨迹匿名

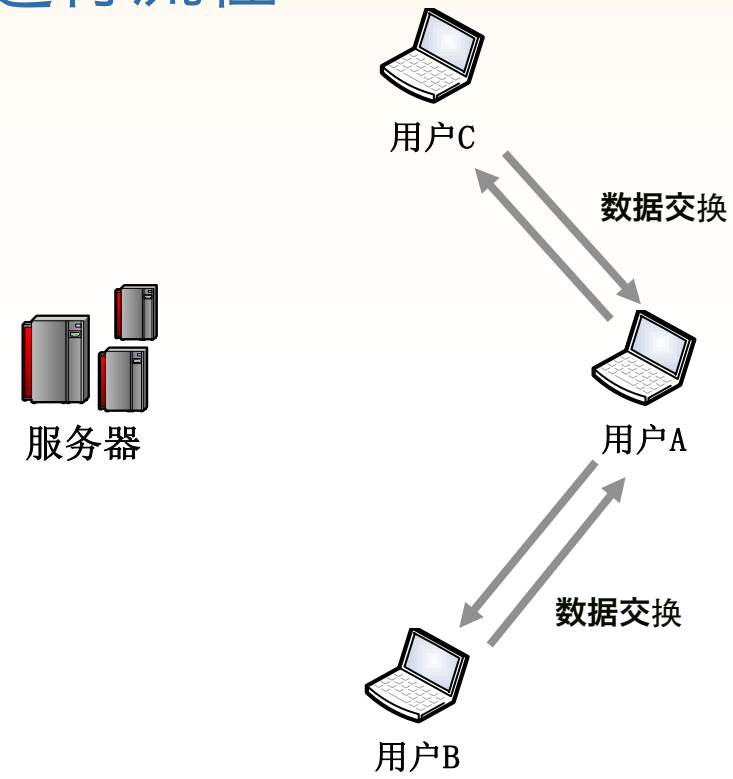
轨迹片段 副本数

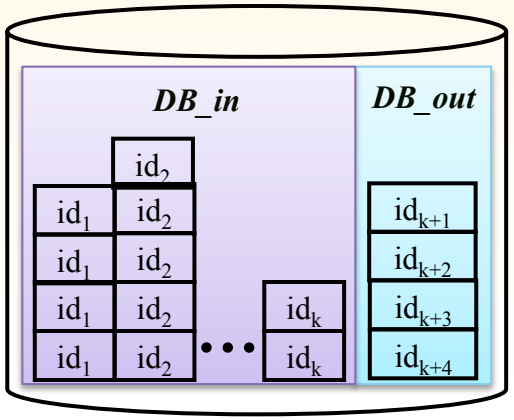
(ID_1, S_1^1)	2	(ID_2, S_2^1)	3	……		(ID_k, S_k^1)	3
(ID_1, S_1^2)	2	(ID_2, S_2^2)	3	……		(ID_k, S_k^2)	3
(ID_1, S_1^3)	2	(ID_2, S_2^3)	3	……		(ID_k, S_k^3)	3
(ID_1, S_1^4)	2	(ID_2, S_2^4)	3			(ID_k, S_k^4)	3
(ID_1, S_1^5)	2	(ID_2, S_2^5)	3			(ID_k, S_k^5)	3
……	…	……	…			……	…

完全提供者 轨迹数据库

位置数据收集K-匿名模型

运行流程：



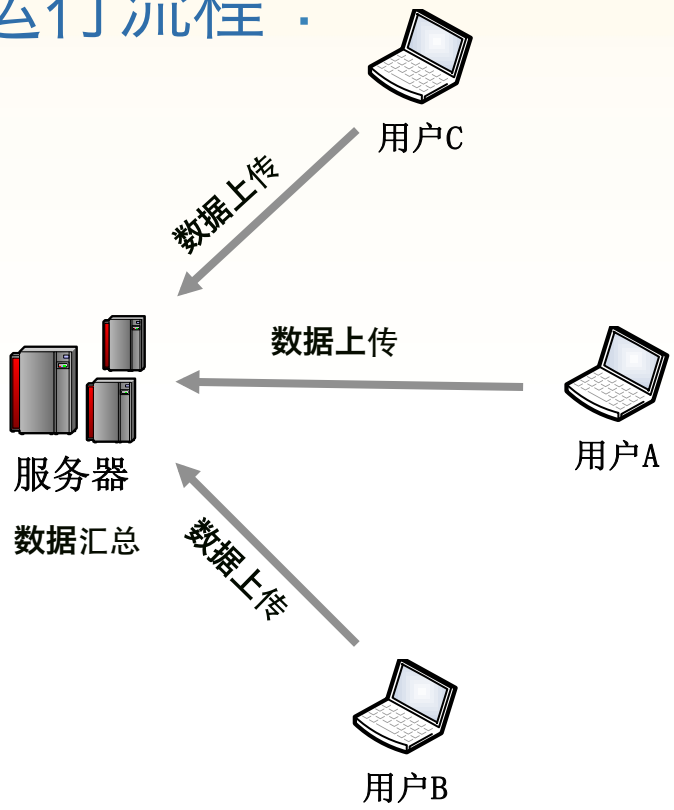


数据交换：

- 随机选择*K*个用户交换*K*K*条数据
- 优先交换时间较早、副本较多的片段
- 等概率选择*DB_in*和*DB_out*中的数据

位置数据收集K-匿名模型

运行流程：



数据上传：

●数据产生时间过早或*DB_out*达到容量　　阈值则开始数据上传过程

●仅上传*DB_out*中的数据

数据汇总：

完整轨迹
$ID_1:S_1^1, S_1^2, S_1^3, S_1^4$ ...
$ID_5:S_5^1, S_5^2, S_5^3, S_5^4$ ...
$ID_{12}:S_{12}^1, S_{12}^2, S_{12}^3, S_{12}^4$ ...
.....

服务器端数据库

时间	轨迹片段	份数
t_1	(ID_1, S_1^1)	2
	$(ID_3:S_3^1)$	3
	$(ID_5:S_5^1)$	2
t_2	$(ID_5:S_5^2)$	2
t_3	$(ID_1:S_1^3)$	2
	$(ID_3:S_3^3)$	3
.....		

临时列表

位置数据收集K-匿名模型

RSA CONFERENCE CHINA 2012

数据来源：

武汉大学自主研发的一套基于位置的网络信息自适应推送系统，该系统所提供的轨迹数据库包含了约1.4万条位置点记录。

(a)真轨迹片段

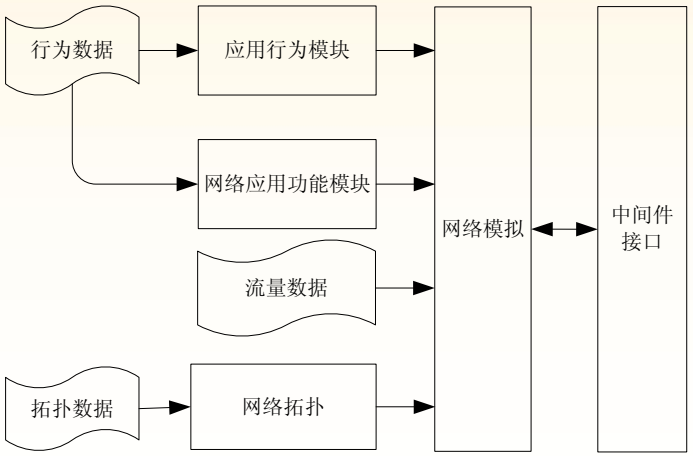
(b) 假轨迹片段

位置数据收集K-匿名模型

RSA CONFERENCE CHINA 2012

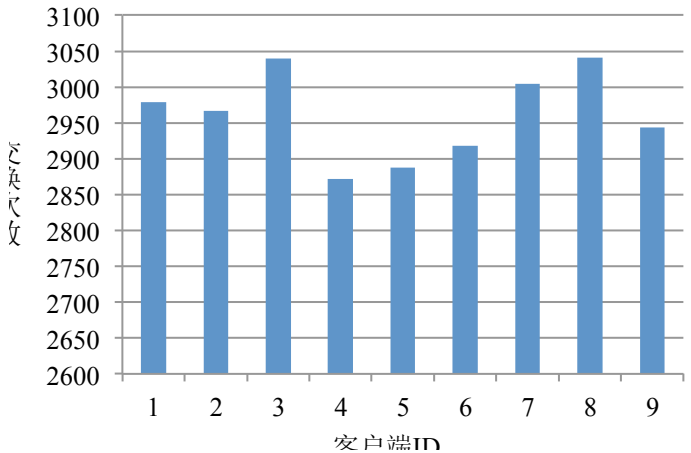
仿真平台：

WhuCNetSim网络仿真器

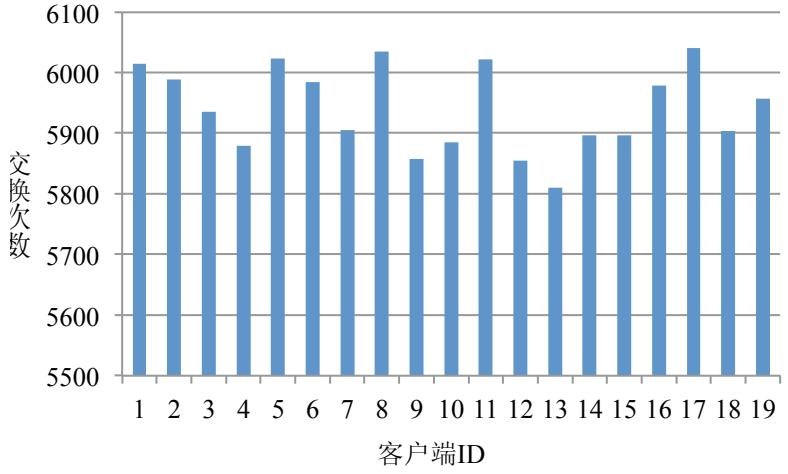


仿真实验：

1.交换次数测试



K=10♪



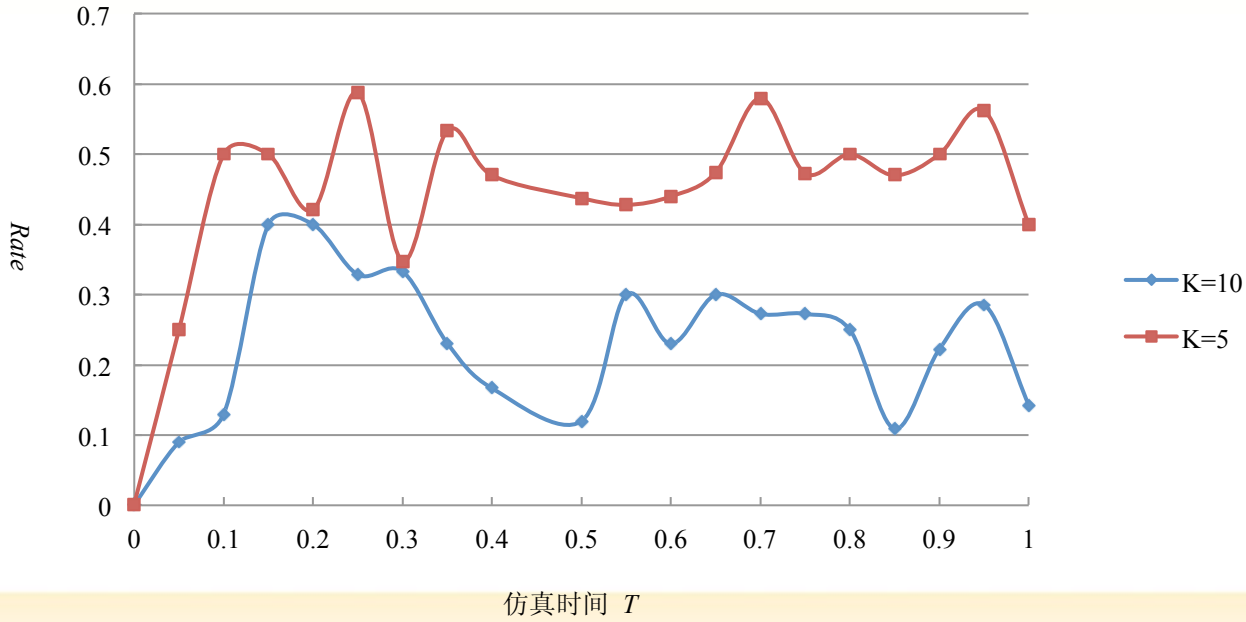
K=20♪

位置数据收集K-匿名模型

仿真实验：

2.匿名状态测试

变量*Rate*，它表示*DB_in*与*DB_out*中由用户自己产生的轨迹片段在整个数据库中所占的比例。*Rate*取值越低，则说明本地数据库中来自于其他用户的数据越多，用户的匿名程度越高。



结束语

- 移动互联网环境下的位置服务是当前学术界和产业界共同关注的热点；
- 在一定程度上，“位置”已经不是“位置服务”的唯一服务内容，而更多的是生成服务的输入性关键性因素。位置服务的内涵需要继续思考和挖掘；
- 隐私风险将是未来移动互联网发展的关键安全因素。基于位置的隐私保护技术还有需要进步一的研究以满足用户的需要；
- 移动互联网环境下的社会行为监管，对于社会的稳定着重要的作用。

谢谢

RSA CONFERENCE
CHINA 2012
RSA信息安全大会2012