doi:10.3969/j.issn.1000-436x.2014.08.015

面向 LBS 的隐私保护模型及方案

杨松涛[1,2]，马春光[1,3]，周长利[1]

（1. 哈尔滨工程大学 计算机科学与技术学院，黑龙江 哈尔滨 150001；
2. 佳木斯大学 信息电子技术学院，黑龙江 佳木斯 154007；
3. 哈尔滨工程大学 国家保密学院，黑龙江 哈尔滨 150001）

摘 要：首先，提出一种基于中心服务器结构的位置隐私保护模型，然后，针对该模型设计了一种基于伪随机置换的位置隐私保护方案，此方案借鉴 *k*-匿名技术、秘密信息检索技术的设计理念和方法，实现了完美匿名和基于位置的盲查询。最后，证明此方案具备不可追踪性和不可关联性等安全属性，并对方案的效率问题进行了分析。

关键词：隐私保护；基于位置的服务；*k*-匿名；盲查询

中图分类号：TP309.2 **文献标识码**：A **文章编号**：1000-436X(2014)08-0116-09

LBS-oriented location privacy protection model and scheme

YANG Song-tao[1,2], MA Chun-guang[1,3], ZHOU Chang-li[1]

(1.College of Computer Science and Technology, Harbin Engineering University, Harbin 150001,China;
2. College of Information and Electronic Technology, Jiamusi University, Jiamusi 154007, China;
3.College of National Secrecy, Harbin Engineering University, Harbin 150001, China)

Abstract: A location privacy protection model was proposed based on the central server structure, designed a location privacy protection scheme based on pseudo-random permutation for the model. Proposed scheme builds on *k*-anonymous and secret information retrieval technology design concept and methods, achieves a blind query and perfect anonymous. This scheme has been proven to have untraceability and unlinkability, and the efficiency is analyzed as well.

Key words: privacy protection; location-based services; *k*-anonymity; blind query

1 引言

基于位置的服务(LBS, location based services)是指融合无线通信、数据库等相关技术的一种基于空间位置的移动信息服务。LBS 的丰富性和多样性大大地提高了人们的生活质量[1]，但是，LBS 带来的隐私安全性问题越来越受人们的关注。位置隐私保护是挑战性问题，近 10 年来该领域的研究非常活跃，取得了丰硕的研究成果，现在大部分的位置隐私研究工作基于中心服务器结构和分布式点对点结构。

Marco Gruteser 借鉴数据库中常用的 *k*-匿名(*k*-anonymity)机制提出了位置 *k*-匿名模型[2]，基于 *k*-匿名的隐私保护方法主要关注匿名查询，其本质是使查询发送者不可区分的，即切断用户身份与查询内容的联系，抵抗攻击者通过查询内容推理出用户的隐私信息。一些研究进一步分析了用户隐私需求并提出了位置多样性[3]、查询多样性[4]和语义多样性[5]等概念。由于用户位置分布的不同，同一匿名集内用户形成的匿名区域可能不同，大量的研究方法容易受到多查询推理攻击。Hilbert Cloak[6]引入 Hilbert 空间曲线理论，首先定义了 reciprocity 特性，同一匿名集的用户对于匿名区域的依赖度相同。Pingley A[7]和 Ni W[8]等人的研究利用 Hilbert 曲线的

收稿日期：2013-03-28；**修回日期**：2013-06-09

基金项目：国家自然科学基金资助项目(61170241)；高等学校博士学科点专项科研基金资助项目(20132304110017)；黑龙江省教育厅科学技术研究基金资助项目(12513049)；哈尔滨市科技创新人才专项基金资助项目(2012RFXXG086)

Foundation Items: The National Natural Science Foundation of China (61170241); Specialized Research Fund for the Doctoral Program of Higher Education (20132304110017); The Education Department of Heilongjiang Province Science and Technology Research Projects (12513049); Harbin Technological Innovation Talent Special Funds (2012RFXXG086)

聚类特性满足了 reciprocity 特性，产生了更小的匿名空间区域，提高了计算效率和通信效率。

Chow[9]等人提出分布式点对点结构下的时空匿名算法。分布式点对点结构下移动用户之间相互信任、相互协作，从而寻找合适的匿名空间。黄毅[10]提出了一种用户协作无匿名区域的隐私保护方法，该方法通过用户之间协作形成匿名组，匿名组内的用户用该组的密度中心代替真实位置发出查询。大量的研究假设参与匿名的各方均是可信的，这与现实情况不符，匿名参与者可能直接或间接泄露精确的位置信息。Hu H B [11]等人的研究仍然采用 *k*-匿名思想，根据信号强度或到达时间差实现多节点构成簇，用近似最小 *k* 聚集簇信息代替坐标来满足 reciprocity 特性。但是，如果用户不能侦测到足够数量的邻居，则算法失效。Hashem T[12]等人的研究利用移动终端形成无线个人 ad-hoc 网络，用户不需要相信匿名器和其他合作匿名者的任何一方来实现匿名。

保护某一时刻的位置信息不能保证用户的轨迹隐私。例如：移动用户在发出连续请求时候，每一请求时刻均发布了一个匿名区域，这些匿名区域被连接起来就会暴露移动对象的轨迹。轨迹 *k*-匿名[13]要求任一条轨迹的所有采样位置和相同的 *k*−1 条轨迹匿名。Cache Cloak[14]借鉴 mix zones[15]和 Path Confusion[16]的优点，采用一阶马尔科夫模型预测用户行进的路径，在路径的交叉点实现了匿名，攻击者无法确定为什么产生新的路径或者什么原因触发了新的查询。

最近的研究将安全多方计算、秘密信息检索和同态加密等技术引入LBS隐私保护领域。Gabriel G[17]等人的研究将位置隐私保护问题转化为最近邻查询问题。Khoshgozaran A[18]提出了更严格的隐私限制，优化了计算量和通信量。Ashouri-Talouki M[19]基于匿名否决网络和同态加密技术提出了组位置隐私协议（group location privacy protocol），该协议在保护用户位置隐私的前提下，大大降低了通信负载。许多存在的研究考虑在最坏的场景下，用户连续共享位置信息带来的隐私问题，但是大部分的研究方案无法真正应用于实践中，同时存在一些有争议的问题尚待解决[20]。

问题：1) 位置服务中所有的服务都是基于位置信息提供的，位置信息本身就是用户隐私的重要部分。位置隐私保护是一把双刃剑，如何在用户享受高质量服务的前提下，保障用户的隐私？

2) LBS 是大规模的即时性服务，在隐私保护机制下，用户终端的计算、存储能力和通信网络负载量成为了瓶颈；如何减轻用户的计算负担和降低网络通信量？

贡献：1) 提出一种位置隐私保护系统模型（LPPM，location privacy protection model）。LPPM 中不需要周期性地位置更新，实现即时的按需查询，减轻了通信网络负载。

2) 设计了一种基于伪随机置换的位置隐私保护方案（LPPRPS，location privacy pseudo-random permutation scheme）。不需要构建匿名框或匿名集，利用随机加密实现隐私数据的安全性，可以抵抗连续多查询攻击和推理攻击。查询过程中，用户与 LBS 服务器没有直接通信，LBS 服务器无法确定哪个用户发出查询、在什么位置发出查询、在什么时间发出查询和查询内容及结果，真正地实现了基于位置的盲查询和完美匿名。

2　位置隐私保护模型

本文提出的位置隐私保护模型采用中心服务器结构，因为中心服务器结构在大量的位置隐私保护方案中得到应用并证明了它的有效性。

定义 1　LBS 查询（LQ），通常是四元组形式表示，形式化地表示每一个 LBS 查询为

$$LQ=(u, loc, t, c)$$

其中，*u* 代表用户身份标识；*loc* 代表用户的当前位置坐标；*t* 代表当前查询时间；*c* 代表用户的查询内容。

定义 2　一个位置隐私保护系统可定义为三元组：（*S*，*A*，*P*），*S* 为参与主体的集合，*A* 为算法的集合，*P* 为协议的集合。

$S=\{U, CS, LBS, SC\}$，其中，$U=\{u_1, u_2, \ldots, u_n\}$为携带移动终端设备用户的集合；*CS* 为中心服务器；*LBS* 为提供基于位置服务的服务器，*SC* 为内置在 *LBS* 中的安全处理器。

$A=\{Cal, En, De\}$，其中，*Cal* 是在 *CS* 内多项式时间内可完成的算法，输入 *LQ*(*u*)，输出加密记录的序号；*En* 是在 *SC* 内多项式时间内可完成的算法，输入原始数据库，输出加密置换后的数据库；*De* 是在 *CS* 内执行的算法，将密文计算为明文。

$P=\{Per, Que\}$，其中，*Per* 是 *SC* 和 *LBS* 之间

的协议，用于加密数据库；*Que* 是 *U*、*CS* 和 *LBS* 之间的三方协议，用于 *U* 查询服务信息。

定义 3 主体集合为 $S=\{S_1, S_2, \cdots, S_m\}$，算法集合为 $A=\{A_1, A_2, \cdots, A_n\}$，协议集合为 $P=\{P_1, P_2, \cdots, P_k\}$，任何一个 $S_i(i=1, \cdots, m)$能够严格执行算法 $A_j(j=1, \cdots, n)$或协议 $P_t(t=1, \cdots, k)$。如果 $S_i(i=1, \cdots, m)$不会泄露任何私有信息给 $S_k(k=1, \cdots, i-1, i+1, \cdots, m)$，则 S_i 是可信的，否则是半可信的。

图 1 描述了 LPPM 模型结构。从可信性角度看，U、CS 和 SC 是可信的，LBS 是半可信的，通信信道是不可信的；从计算和存储能力角度看，U 具有较弱的计算和存储能力，CS 具有较强的计算和存储能力，LBS 具有超强的计算、存储和概率统计分析能力。U 具有定位和发送查询请求功能；CS 具有收集 U 的位置请求信息和代理 U 计算、查询的功能；LBS 具有地理信息库的检索和存储功能。

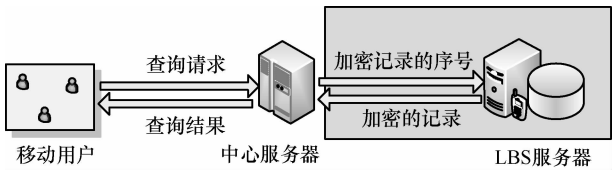


图 1 LPPM 模型结构

在 LPPM 模型中，首先，LBS 初始化数据库，SC 执行 Per 协议，调用 En 算法加密 LBS 数据库；其次，u_i（$i=1, \cdots, n$）执行 Que 协议发送 LQ(u_i)给 CS，CS 调用 Cal 算法计算加密记录的序号并执行 Que 协议查询加密记录；然后，LBS 检索被加密的数据库，并将结果发送给 CS；最后，CS 调用 De 算法计算明文，并将结果发送给 u_i。LPPM 模型具有不可追踪性和不可关联性。

定义 4 不可追踪性是指 LBS 通过公共信息和 LQ(*u*)信息分析出来 *u* 的位置轨迹的概率很低。公共信息是通过公共信息库就可以得到的信息，位置轨迹是指某个移动对象的位置信息按时间排序的序列，位置轨迹可以表示为 $T=\{T_i, (x_1, y_1, t_1), (x_2, y_2, t_2), \cdots, (x_n, y_n, t_n)\}$，其中，$T_i$ 表示该轨迹的标识符，$(x_i, y_i, t_i)(i=1, \cdots, n)$表示移动对象在时刻 t_i 的位置为 (x_i, y_i)，t_i 为采样时间。

定义 5 不可关联性是指 $u_i(i=1, \cdots, n)$与 *C* 之间相互关联的概率很低。用户集合为 $U=\{u_1, u_2, \cdots, u_n\}$，查询内容为 *C*。*k*-匿名度量是针对位置隐私保护最流行的度量标准，量化的匿名度 *k* 可以有效地度量不可关联性，*k* 值越大，不可关联性越好。

为了更好地阐述第 3 节的基于伪随机置换位置隐私保护方案，下面定义一些相关概念：位置 *k*-匿名、位置隐私、秘密信息检索和安全处理器。

定义 6 位置 *k*-匿名(*k*-anonymity)[2]，给定用户集合 $S_1=\{u_1, u_2, \cdots, u_k\}$ 和位置集合 $S_2=\{loc_1, loc_2, \cdots, loc_k\}$，其中 u_i 表示用户身份标识，loc_i 表示用户 u_i 所在的当前位置。从攻击者角度来说，用户 *u* 实现位置 *k*-匿名当且仅当 *u* 的位置 *loc* 与 S_2 中其他元素是无法区分的，且 $u \in S_1$, $loc \in S_2$。

匿名的本意是指在同一匿名集 *AS* 内的主体是不可区分的。相对攻击者而言，主体没有唯一的特征标识。匿名不是保护某一具体的主体标识，而是使攻击者不能建立主体与其行为动作的联系。匿名集的量化需要一个阈值 *k*，*k* 在一般意义下代表匿名集的势|*AS*|，*k* 值越大，匿名效果越好。从用户角度看，*k* 值是静态的，代表用户要求的隐私度。从攻击者角度看，*k* 值是动态的，根据先验知识将匿名集划分为 *n* 等价类。

定义 7 位置隐私(location privacy)[2]。基于位置查询时，用户 *u* 向 LBS 服务器提供了时空数据 $Data=\{u, l, t, c\}$，其中，*u* 表示用户身份标识，*l* 表示用户 *u* 所在的当前位置，*t* 表示当前时间，*c* 表示查询内容。$T_u=\{u_{t1}, u_{t2}, \cdots, u_{tm}\}$ 是 *u* 按照时间序列形成的位置轨迹。位置隐私是指攻击者通过分析 T_u 来定位和跟踪 *u* 得到的用户隐私信息。位置查询隐私（query privacy）是位置隐私的一种，是指攻击者基于 *l* 来分析 *u* 和 *c* 之间关联关系得到的用户隐私信息。

定义 8 秘密信息检索(PIR, private information retrieval)[21]。对于用户 *u* 和数据库 DB，秘密信息检索是指 *u* 从 DB 中秘密检索第 *i* 条记录，服务器无法确定 *i* 的值。换句话说，服务器无法确定 *u* 对第 *i* 条记录感兴趣。

秘密信息检索主要分为 2 类：信息论的秘密信息检索和可计算的秘密信息检索。信息论的秘密信息检索是假设服务器的计算能力是无限的，缺乏实用性；可计算的秘密信息检索是假设服务器的计算能力有限的情况下提出的，安全性基于数学上的困难计算。

定义 9 安全处理器(SC, secure coprocessor)。

SC 为一个计算环境，并集成了 RAM、高速缓存和某些加密算法的硬件加速器，SC 可以发现对其硬件和软件的任何改动并采取相应措施，同时也禁止外界对其 RAM 的任何访问。

3　基于伪随机置换的位置隐私保护方案

依照第 2 节的位置隐私保护模型，提出基于伪随机置换的位置隐私保护方案，S={U, CS, LBS, SC}，A={Cal, En, De}，P={Per, Que}分别主体、算法和协议的集合。LQ(u)=(u, loc, t, c)代表用户 u 的服务请求。针对以往的 PIR 方案中查询计算复杂度过大的问题，通过 CS 和 LBS 协同处理用户查询，主要计算量由 LBS 在初始化阶段完成。初始化部分的核心内容是加密原始数据库，因此本文设计了 En 算法，SC 负责将原始数据库的记录顺序按随机的方式打乱，加密后存入另一个新数据库，这样 LBS 就不可能知道记录的对应关系，SC 虽然被放置在 LBS 端，但是不会泄漏任何信息给服务器。该方案的主要思想：将查询过程分为 2 个阶段，初始化阶段在线下操作，用以 CS 与 LBS 之间交互索引数据库和为原始数据库生成茫然数据；查询阶段中 CS 代理用户计算最近邻信息并得到一个加密记录序号，再通过此序号查询数据库的茫然数据得到结果。

3.1　初始化

令 DB_{ori}、DB_{en} 和 DB_{in} 分别代表原始数据库、加密数据库和索引数据库。

DB_{ori}={Rec[1], Rec[2], …, Rec[n]}, Rec[i]代表 DB_{ori} 中第 i 条记录，1≤i≤n。

DB_{en}={Rec'[1], Rec'[2], …, Rec'[n]}, Rec'[i]代表 DB_{en} 中第 i 条记录，1≤i≤n。

DB_{in}={Rec"[1], Rec"[2], …, Rec"[n]}, Rec"[i]代表 DB_{in} 中第 i 条记录，1≤i≤n。

兴趣点(PoI, point of interest)，代表用户感兴趣的公共设施，PoI=(Id, $Name$, Loc, $Info$)，其中：Id 代表记录序号；$Name$ 代表 PoI 的名称；Loc 代表 PoI 的地理坐标；$Info$ 代表 PoI 的详细服务信息。

1) 生成 DB_{in} 数据库

首先，LBS 将 DB_{ori} 按照 $Name$ 字段排序；其次，提取 Id、$Name$ 和 Loc 字段的记录形成索引数据库 DB_{in}；最后，LBS 将 DB_{in} 发送给 CS。图 2 描述了 DB_{in} 数据结构及生成过程。

Id	Name	Loc	Info
1	学校	(x_1,y_1)	*** ***
2	加油站	(x_2,y_2)	*** ***
3	医院	(x_3,y_3)	*** ***
4	…	…	…
5	宾馆	(x_i,y_i)	*** ***
6	商场	(x_n,y_n)	*** ***

$DB_{\text{ori}} \rightarrow DB_{\text{in}}$

Id	Name	Loc
1	宾馆	(x_i,y_i)
2	加油站	(x_2,y_2)
3	商场	(x_n,y_n)
4	…	…
5	学校	(x_1,y_1)
6	医院	(x_3,y_3)

图 2　生成 DB_{in} 数据库

2) 加密 DB_{ori} 数据库

加密 DB_{ori} 是由 SC 和 LBS 之间交互形成 DB_{en} 的过程。设 $\boldsymbol{DB}_{\text{ori}}$ 和 $\boldsymbol{DB}_{\text{en}}$ 是{0，1}^{n*l} 矩阵，则 $\boldsymbol{DB}_{\text{ori}}$ 和 $\boldsymbol{DB}_{\text{en}}$ 分别为

$$\boldsymbol{DB}_{\text{ori}}=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{bmatrix},$$

$$\boldsymbol{DB}_{\text{en}}=\begin{bmatrix} b'_{11} & b'_{12} & \cdots & b'_{1m} \\ b'_{21} & b'_{22} & \cdots & b'_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ b'_{n1} & b'_{n2} & \cdots & b'_{nm} \end{bmatrix}$$

其中，n 为 $\boldsymbol{DB}_{\text{ori}}$ 的记录数，l 为每条记录的长度（比特），每条记录是{0，1}^{l} 向量。Rec[i]和 Rec'[i]分别被分为若干个 32 bit 的数据块，Rec[i]=(b_{i1},…b_{im})，Rec'[i] =(b'_{i1},…,b'_{im})，其中，$m=\lfloor l/32 \rfloor+1$。图 3 示意了 DB_{ori} 加密过程。实现加密的伪代码见 En 算法。

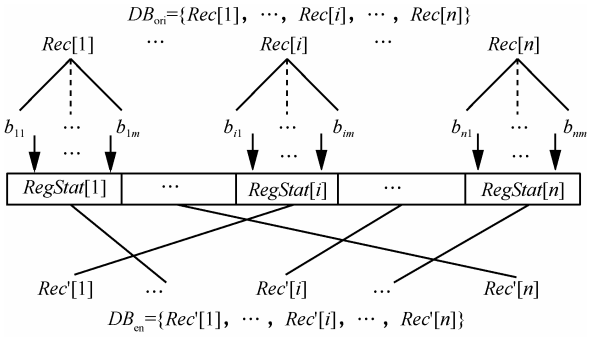


图 3　DB_{ori} 加密过程

图 4 描述了数据库置换协议。在 SC 内建立记录映射表 map[t]，DB_{ori} 记录和 DB_{en} 记录的对应关系记载 map[t]中；SC 以 32 bit 随机数为种子生成 n×32 bit 长伪随机数序列；SC 读入 DB_{ori} 的数据，与伪随机数序列相应位做模 2 运算，结果存入寄存器内；SC 按照 map[t]重新排序寄存器，并将寄存器内数据依次写入 DB_{en}。

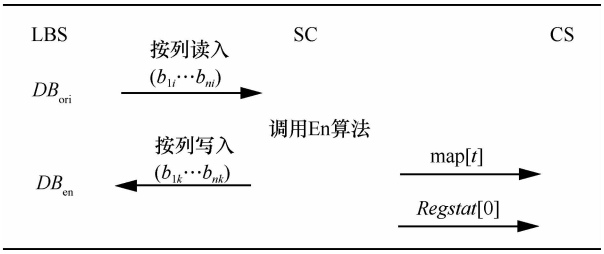


图 4 数据库置换协议

En 算法

输入：DB_{ori}

输出：DB_{en}

1) **for** $t \leftarrow 1$ to n **do**
2) map[t]$\leftarrow$ random{1, 2, … , n}
// 1 到 n 之间产生不重复的随机整数
3) t=t+1
4) **repeat**
5) **for** $i \leftarrow 1$ to m do
6) *RegStat*[0] $\leftarrow$ random{0, 1}^{32}
//生成 32 bit 随机数存入 32 bit 寄存器内。
7) **if** *RegStat*[0] == 0
8) **then** goto 1)
9) **else**
10) 以 *RegStat* [0]为种子，生成 n×32 bit 伪随机数序列 $PRS[n\times 32]=(p_1,p_2,\cdots,p_n)$
11) **for** $j \leftarrow 1$ to n **do**
12) *RegStat* [j] $\leftarrow b_{ji} \oplus p_j$
13) j=j+1
14) **repeat**
15) *RegStat* [map[i]] $\leftarrow$ *RegStat* [i]
//按照映射表对寄存器重新排序
16) **for** $k \leftarrow 1$ to n **do**
17) 追加 *RegStat* [k]到 DB_{en} 的第 k 条记录上
18) k=k+1
19) **repeat**
20) i=i+1
21) **repeat**

3.2 查询

查询包括 2 个步骤：U 与 CS 之间的交互过程和 CS 与 LBS 之间的交互过程，用户发送服务请求给 CS，CS 代理用户查询所需信息，结果反馈给用户。用户 u 用 CS 的公钥加密 LQ（u），得到 E_p（LQ（u）），并发送给 CS；CS 用私钥解密 E_p（LQ（u）），$D_s(E_p(LQ(u)))$= LQ(u)，提取 *loc*、*c*；CS 调用 Cal 算法计算加密记录序号 *E-Id*；CS 向 LBS 发出查询 Q_E(*E-Id*)；LBS 检索数据库 DB_{en}，结果 R_E 发送给 CS；CS 调用 De 算法，结果 R_D 返回给 u。

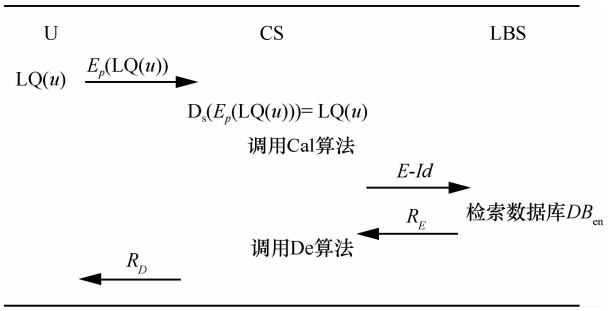


图 5 服务信息查询协议

3.3 计算

计算过程在 CS 内完成，包括 2 部分：加密记录号计算和明文计算。加密记录号计算过程：首先，CS 以用户的查询内容 c 为关键字检索数据库 DB_{in}，结果集为 R；然后，分别计算用户坐标 *loc* 与 R 中每个元素坐标的欧氏距离，距离最短的就是距离用户最近 PoI，确定记录号；最后，查询 map[t]，确定最近 PoI 在加密数据库中的序号 *E-Id*。Cal 算法具体描述了计算加密记录号的伪代码。

Cal 算法

输入：*loc*，*c*，DB_{in}

输出：*E- Id*

1) $d \leftarrow \infty$，*index* $\leftarrow$ 0
2) **for** $i \leftarrow 1$ to n **do**
3) **if** *Rec′*[i].*Name*== c
//读入 DB_{in} 的当前记录并比较
4) **then** *Rec′*[i] insert R
5) i++
6) **repeat**
7) **for** $j \leftarrow 1$ to $|R|$ **do**
8) **if** Ed（R_j.*Loc*，*Loc*）＜ d
// Ed 为计算欧氏距离函数
9) **then** $d \leftarrow$ Ed（R_j.*Loc*，*Loc*）
10) *index* $\leftarrow R_j$.*Id*
11) **repeat**
12) 查询映射表 map[t]，确定加密数据库中的索引号 *E- Id*。
13) **return** *E-Id*

明文计算是 CS 利用预处理阶段接收到的随机

种子 *RegStat*[0]生成伪随机序列矩阵 ***PRS'***[*n*×*m*]

$$\boldsymbol{PRS}'[n\times m]=\begin{bmatrix} p'_{11} & p'_{12} & \cdots & p'_{1m} \\ p'_{21} & p'_{22} & \cdots & p'_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p'_{n1} & p'_{n2} & \cdots & p'_{nm} \end{bmatrix}$$

根据 *index* 确定伪随机码矩阵中所在的行，该行上的伪随机码与 R_E 作异或运算，得到明文信息 R_D，De 算法具体描述了明文计算的伪代码。

De 算法

输入：*index*，R_E

输出：R_D

1) $\boldsymbol{i}\leftarrow index$

2) $R_D \leftarrow p'_{i1}\, p'_{i2}\cdots p'_{im}\oplus R_E$

3) **return** R_D

4　方案分析

4.1　隐私安全性分析

方案中 Per 协议和 Que 协议是 LPPM 的关键技术，它们的安全性直接影响 LPPM 的隐私安全属性。本节详细证明了 Per 协议和 Que 协议在 LPPRPS 中的隐私安全性。

定义 10　完美匿名。全部移动用户组成的集合 $U=\{u_1, u_2,\cdots, u_n\}$，任意查询 q，$P_q(u_i)=1/n$，其中 $P_q(u_i)$ 代表查询 q 由用户 u_i 发出的概率。

定义 11　位置盲查询。全部兴趣点组成的集合 $PoI=\{poi_1, poi_2, \cdots, poi_n\}$，任意查询 q，$P'_q(u)=1/n$，其中，$P'_q(u)$代表查询内容与用户 u 相互关联的概率。

定义 10 是 *k*-匿名的一个特例，完美匿名使查询用户与所有用户不可区分。定义 11 将查询兴趣点集合与结果集扩展为 LBS 的整个数据库。

命题 1　每次查询是相互独立的，即

$$p(q_1,\cdots,q_k)=p(q_1)p(q_2)\cdots p(q_k)$$

命题 2　每次查询的熵是最大的，即

$$H(q_1)=H(q_2)=\cdots=H(q_k)=\log n$$

定理 1　给定协议 Que 和算法 En，通过 Que 执行查询 $q_1, \cdots, q_k$, Que 具有私有安全性当且仅当 $q_1, \cdots, q_k$ 的联合信息熵是最大的。

证明　使用数学归纳法证明，首先考虑 *k*=1 的情况，其次考虑 *k*=2 的情况，然后假设 *k*=*K* 的情况命题成立，证明 *k*=*K*+1 的情况成立。

当 *k*=1 时，仅有一次查询执行协议，LBS 记录顺序被随机的扰乱，加密记录无法对应到原始记录。$p(q_1=1)=\cdots=p(q_1=n)=1/n$, $H(q_1)=H_{\max}(q_1)=\log n$。命题 2 得证。

当 *k*=2 时，因为 LBS 读取一条加密记录与 $q_1=q_2$ 或 $q_1\neq q_2$ 是无关的，所以对于 LBS 来说，q_1 和 q_2 是独立变量。因此，$p(q_1,q_2)=p(q_1)p(q_2)$; $H(q_1)=H(q_2)=\log n$。命题 1 得证。

假设 $p(q_1,\cdots q_K)=p(q_1)p(q_2)\cdots p(q_K)$; $H(q_1)=\cdots=H(q_K)=\log n$。考虑第 *K*+1 次查询，由命题 1 和命题 2，有 $p(q_1,\cdots,q_K,q_{K+1})=p(q_1)p(q_2)\cdots p(q_K)p(q_{K+1})$; $H(q_1)=\cdots=H(q_K)=H(q_{K+1})=\log n$。定理得证。

定理 2　Que 协议实现了完美匿名，满足 LPPM 的不可追踪性。

证明　由于 Que 协议在 LPPM 下执行时，U 和 CS 是可信的，而且 U 与 CS 之间的通信是加密的。所以，U 与 CS 之间交互的隐私安全性能够得到保障。

因为 LBS 不能得到 LQ(*u*)，进而，LBS 无法获得 *u* 的 *loc*。

所以，Que 协议实现了完美匿名，即用户集合为 $U=\{u_1, u_2, \cdots, u_n\}$，用户 u_i 的匿名度为 *k*，有 *k*=*n*。

又因为，每次查询 LBS 确定 u_i 的位置为 *loc* 概率为 1/*n*，信息熵为 log*n*，信息熵达到了最大值。

所以，从 LBS 的角度分析，用户位置信息是杂乱无序的，进而系统具备不可追踪性。

定理 3　Que 协议实现了位置盲查询，满足 LPPM 的不可关联性。

证明　因为 $DB_{\text{ori}}=\{Rec[1],Rec[2],\cdots,Rec[n]\}$，$DB_{\text{en}}=\{Rec'[1],Rec'[2],\cdots,Rec'[n]\}$ 是向量集合，*Rec*[*i*]，*Rec*'[*j*]是 DB_{ori}、DB_{en} 中的随机向量，从信息论的角度分析，LBS 正确猜中 *Rec*[*i*]和 *Rec*'[*j*]对应关系的概率是 1/*n*，每次查询所产生的信息熵为

$$H(q)=\sum_{i=1}^{n}p(x_i)\log\frac{1}{p(x_i)}=\sum_{i=1}^{n}\frac{1}{n}\log n=\log n \quad (1)$$

所以，由式(1)可知，单次查询信息熵达到了最大值。

又因为每次查询是相互独立的，LBS 无法从查询中得到任何信息，则 *m* 次查询的联合信息熵为

$$H(q_1\cdots q_m)=\sum_{i=1}^{m}H(q_i)=m\log n \quad (2)$$

所以，由式(2)可知，多次查询的联合信息熵达到了最大值。

由定理 1，Que 协议具有私有安全性，进而证明了 Que 协议实现了位置盲查询，满足不可关联性。

定理 4 实现完美匿名和位置盲查询当且仅当 LPPM 具有不可追踪性和不可关联性。

4.2 效率分析

本文比较 IntervalCloak、Hilbert cloak、Casper 和 LPPRPS 的 *k*-ASR。首先，固定用户数量 *n*=10 000，匿名度为 *k*，*k*-ASR 大小用占整个空间的百分比表示。图 6(a)显示每次查询的平均 *k*-ASR，很明显，LPPRPS 实现了完美匿名。图 6(b)显示固定 *k*=100，*n* 为用户数量，随着 *n* 值变大，增加了用户的分布密度，IntervalCloak、Hilbert cloak 和 Casper 的 *k*-ASR 相应减小，LPPRPS 的 *k*-ASR 没有变化，所以，LPPRPS 不受用户数量和分布密度的影响。

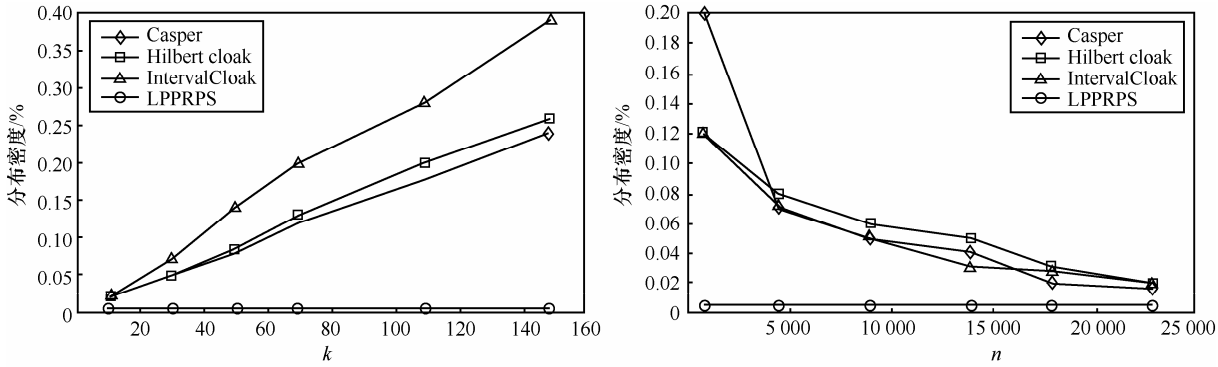
图 7 显示了不同的 *k* 和 *n* 时，构建 ASR 的时间。LPPRPS 中不需要构建 ASR，因为在 LBS 隐私保护研究中，多数方法采用最近邻匿名技术，欧氏空间距离计算是必不可少的，所以 Cal 算法并没有增加计算量。

图 8 显示了不同的 *k* 和 *m*（PoI 的数量）时，查询结果候选集的变化。随着数据库规模变大和 *k* 值的增加，IntervalCloak、Hilbert cloak 和 Casper 的查询候选集也变大。LPPRPS 通过盲查询得到唯一的记录，不会受到 LBS 数据库规模和 *k* 值的影响。

4.3 对比分析

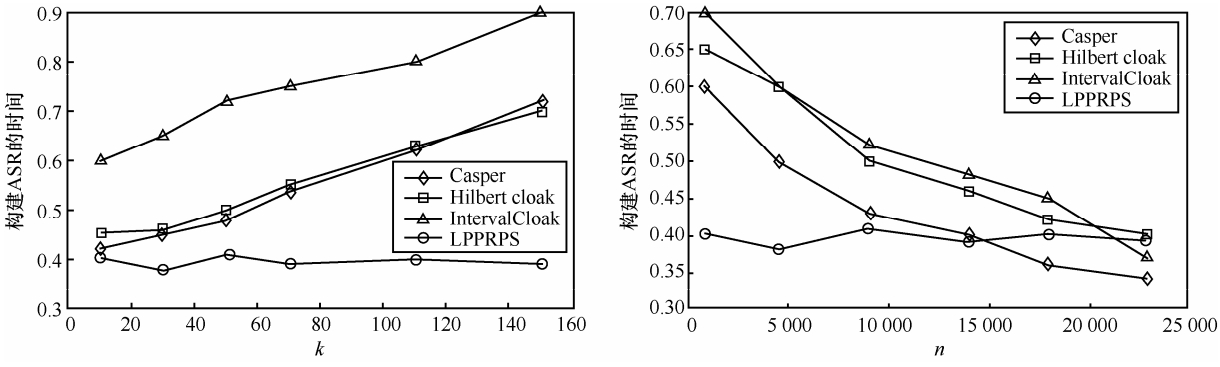
基于 *k*-匿名思想的隐私保护方法通常使用匿名框代替精确位置，匿名算法和查询算法均比较简单，计算量和通信量相对比较小。但是，*k*-匿名技术本身的不足容易遭受推理攻击，隐私保护度受到攻击者背景知识的影响。定理 2 证明了 LPPRPS 实现了完美匿名。

基于 PIR 技术的位置隐私保护方法在隐私保护理念上强于基于 *k*-匿名技术方法。在文献[17]中，每一个查询不可避免地线性扫描整个数据库，这样就带来了巨大的计算花费和通信花费，每次查询的通信复杂度为 $\sqrt{n}$，*n* 代表数据库的规模，在用户数量多的场景下，服务器无法承受私有信息检索带来的计算量和通信量。LPPRPS 类似于文献[18]，主要



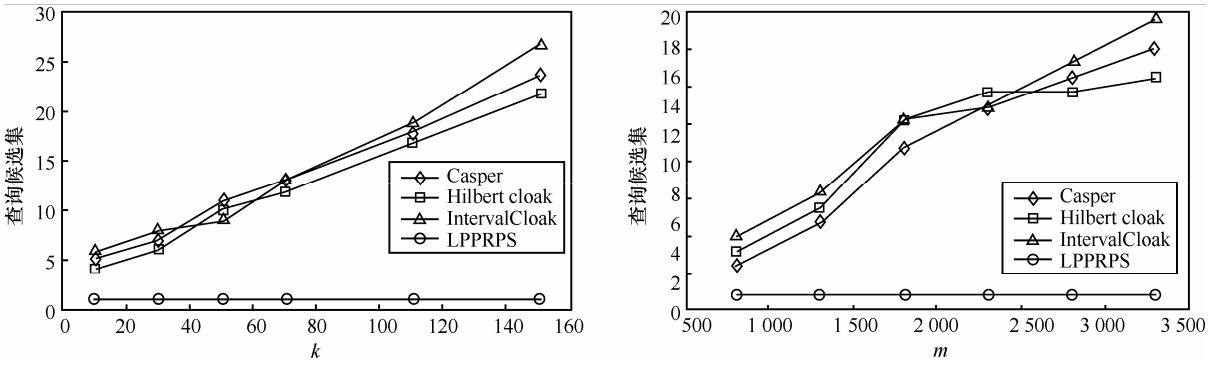
(a) *k* 可变，*n*=10 000　　(b) *n* 可变，*k*=100

图 6 *k*-ASR 大小



(a) *k* 可变，*n*=10 000　　(b) *n* 可变，*k*=100

图 7 *k*-ASR 构建时间



(a) k 可变，m=2 000　　(b) m 可变，k=100

图 8　查询结果候选集的变化

区别是文献[18]中的 SC 参与每次查询，由于 SC 计算和存储能力受限，查询效率会大大折扣。

LPPRPS 中的一些复杂性工作（如数据库加密）可以在初始化阶段离线进行，计算复杂度为 $O(cn)$，优于现有模型的计算复杂度 $O(n^{3/2})$[21]，SC 不会对查询效率产生任何影响，De 算法的复杂度与查询结果的位数长度是线性关系，计算复杂度为 $O(n)$，n 为查询结果的位数长度。定理 3 从信息论角度证明了 LPPRPS 没有削弱隐私保护的安全性，符合定义 8 的要求。

5　结束语

中心服务器结构下，基于 k-匿名的位置隐私保护方法有以下问题：1）需要用户定期更新位置信息，然而，定期更新位置信息成为中心服务器结构的根本瓶颈，随着移动用户数量的增加，加重了中心服务器的通信负担和存储负担；2）匿名算法的复杂性造成匿名失败或者时间延迟，无法保证服务质量；3）更大的匿名区域和匿名集虽然能够取得更好的位置模糊和身份匿名的效果，但是，结果集的求精过程增加中心服务器的计算负担。本文针对以上问题，提出了一种位置隐私保护模型和设计了一种基于伪随机置换的位置隐私保护方案，包括 2 个协议和 3 个算法。文中证明了此方案实现了完美匿名和位置盲查询，并且满足 LPPM 的不可追踪性和不可关联性。另外，还对方案的安全性和效率进行了分析，安全性优于基于 k-匿名技术方案，在部分场景下不低于基于 PIR 技术方案。

进一步研究工作包括：1）多 LBS 协同计算问题，保证多个 LBS 相互勾结也不能推断出来用户隐私数据，同时抵抗恶意节点的主动攻击；2）用户认证问题，如果没有合理的认证机制，LBS 数据库资源会受到黑客的肆意攻击，黑客也可能假冒 LBS 窃取用户隐私数据；3）用户细颗粒度查询时候算法优化问题，例如，用户查询“离我最近加油站的价格信息”，有待于在本文研究的基础上可以进一步探讨研究；4）LBS 数据库信息频繁更新问题，在一些特定的场景下，例如，基于位置的广告精准投放，LBS 数据库信息经常变动，本文方案的计算效率和安全性均会受到影响，下一步研究工作再深入分析 PIR 技术，寻求安全性和效率的平衡点及度量的方法。

参考文献：

[1] HARRISON B, DEY A. What have you done with location-based services lately[J]. IEEE Pervasive Computing, 2009, 8(4): 66-70.

[2] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[A]. Proceedings of the 1st International Conference on Mobile Systems, Applications and Services[C]. A Francisco, California, 2003. 31-42.

[3] BAMBA B, LIU L, PESTI P, *et al*. Supporting anonymous location queries in mobile environments with privacygrid[A]. Proceedings of the 17th International Conference on World Wide Web[C]. Beijing, China, 2008.237-246.

[4] LIU F, HUA K A, CAI Y. Query l-diversity in location-based services[A]. Proceedings of the 10th International Conference on Mobile Data Management: Systems, Services and Middleware[C]. 2009. 436-442.

[5] LEE B, OH J, YU H, *et al*. Protecting location privacy using location semantics[A]. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. 2011. 1289-1297.

[6] KALNIS P, GHINITA G, MOURATIDIS K, *et al*. Preventing location-based identity inference in anonymous spatial queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(12): 1719-1733.

[7] PINGLEY A, YU W, ZHANG N, *et al*. A context-aware scheme for privacy-preserving location-based services[J]. Computer Networks,

2012, 56(11): 2551-2568.

[8] NI W, ZHENG J, CHONG Z. HilAnchor: location privacy protection in the presence of users' preferences[J]. Computer Science and Technology, 2012, 27(2):413-427.

[9] CHOW C, MOKBEL M F, LIU X. A peer to peer spatial cloaking algorithm for anonymous location based services[A]. Proceedings of the ACM Symposium on Advances in Geographic Information Systems[C]. Arlington, VA, 2006.171 - 178.

[10] 黄毅,霍峥,孟小峰. CoPrivacy:一种用户协作无匿名区域的位置隐私保护方法[J].计算机学报,2011, 34(10): 1976-1985.
HUANG Y, HUO Z, MENG X F. CoPrivacy : a collaborative location privacy-preserving method without cloaking region[J]. Chinese Journal of Computers, 2011, 34(10): 1976-1985.

[11] HU H B, XU J L. Non-exposure location anonymity[A]. IEEE 25th International Conference on Data Engineering[C].2009.1120-1131.

[12] HASHEM T, KULIK L. “Don't trust anyone”: privacy protection for location-based services[J]. Pervasive and Mobile Computing, 2011, 7(1): 44-59.

[13] NERGIZ M E, ATZORI M, SAYGIN Y, *et al*. Towards trajectory anonymization: a generalization-based approach[J]. Transactions on Data Privacy, 2009, 2(1): 47-75.

[14] MEYEROWITZ J, CHOUDHURY R R. Hiding stars with fireworks: location privacy through camouflage[A]. Proceedings of the 15th Annual International Conference on Mobile Computing and Networking[C]. Beijing, China, 2009.345-356.

[15] BERESFORD A R, STAJANO F. Location privacy in pervasive computing[J]. IEEE Pervasive Computing, 2003, 2(1): 46-55.

[16] HOH B, GRUTESER M, XIONG H, *et al.* Preserving privacy in GPS traces via uncertainty-aware path cloaking[A]. Proceedings of the 14th ACM Conference on Computer and Communications Security[C]. Alexandria, Virginia, USA, 2007. 161-171.

[17] GHINITA G, KALNIS P, KHOSHGOZARAN A, *et al*. Private queries in location-based services: anonymizers are not necessary[A]. Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data[C]. Vancouver, Canada: ACM, 2008.121-132.

[18] KHOSHGOZARAN A, SHAHABI C, SHIRANI-MEHR H. Location privacy: going beyond *k*-anonymity, cloaking and anonymizers[J]. Knowledge and Information Systems, 2011, 26(3):435-465.

[19] ASHOURI-TALOUKI M, BARAANI-DASTJERDI A, SELCUK A. GLP: a cryptographic approach for group location privacy[J]. Computer Communications, 2012, 35(12): 1527-1533.

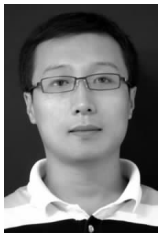
[20] 林欣,李善平,杨朝晖. LBS 中连续查询攻击算法及匿名性度量[J].软件学报,2009,20(4):1058-1068.
LIN X, LI S P, YANG Z H. Attacking algorithms against continuous queries in LBS and anonymity measurement[J]. Chinese Journal of Software, 2009,20(4):1058-1068.

[21] ASONOV D. Querying Databases Privately: a New Approach to Private Information Retrieval[M]. Springer, 2004.

作者简介：

杨松涛（1972-），男，黑龙江佳木斯人，哈尔滨工程大学博士生，佳木斯大学副教授，主要研究方向为物联网安全、隐私保护。

马春光（1974-），男，黑龙江双鸭山人，哈尔滨工程大学院教授、博士生导师，哈尔滨工程大学国家保密学院副院长，主要研究方向为密码学、信息安全、传感网与物联网。



周长利（1985-），男，黑龙江哈尔滨人，哈尔滨工程大学博士生，主要研究方向为物联网安全、隐私保护。