

文章编号:1007-5321(2021)06-0103-06

DOI:10.13190/j.jbupt.2021-041

# 多值预测位置隐私保护机制

宋成, 金彤, 贺军义

(河南理工大学 计算机科学与技术学院, 焦作 454003)

**摘要:** 针对当前位置隐私保护方案中存在的安全和效率问题,基于马尔可夫链技术提出一种多值预测查询的位置隐私保护方案. 首先,根据状态转移矩阵对输入的多个查询值进行计算,并生成下一时刻的预测位置和查询内容;然后,基于布隆过滤器原理,建立兴趣点缓存机制. 安全分析结果表明,所提方案满足匿名性、不可伪造性和抵抗查询服务追踪等安全特性;仿真结果表明,所提方案与现有方案相比具有较高的执行效率和较低的通信开销,且有较高的缓存命中率,能有效减少与基于位置服务器间的交互次数.

**关键词:** 位置隐私; 马尔可夫链; 布隆过滤器; 缓存

**中图分类号:** TP309

**文献标志码:** A

## A Multi-Value Prediction Location Privacy Protection Mechanism

SONG Cheng, JIN Tong, HE Jun-yi

(School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454003, China)

**Abstract:** To solve the security and efficiency issues in the existing location privacy protection schemes, a location privacy protection scheme based on Markov chain technology for multi-value predictive query is proposed. First, the state transition matrix is used to calculate the input multiple query values, and generate the predicted position and the query content at the next moment. Then, Bloom filter technology is adopted to establish a caching mechanism for points of interest. Security analyses demonstrate that the proposed scheme satisfies security features such as anonymity, unforgeability, and is able to resist to query service tracking. Simulation results show that the proposed scheme has higher execution efficiency, lower communication costs, and a higher cache hit rate, which can effectively reduce the number of interactions with the location based service server.

**Key words:** location privacy; Markov chain; Bloom filter; cache

随着无线通信网络和移动智能终端技术的发展,基于位置服务(LBS, location based service)广泛应用于日常生活中,如地图导航、酒店住宿、餐厅推荐等. LBS 为用户带来便利的同时,也使用户面临泄露位置隐私的风险. 近年来,国内外学者针对LBS 隐私泄露的问题<sup>[1]</sup>做了大量研究,其中,假位置

和  $K$  匿名技术是应用较为广泛的隐私保护技术,但普遍存在查询结果冗余的缺陷. 假位置技术的原理是:通过提交若干虚假位置取代用户位置,使攻击者无法准确判断用户的位置坐标,但易遭到背景知识的攻击. 针对该问题,王洁等<sup>[2]</sup>提出计算假位置间的语义距离和查询概率,建立位置语义树,使其满足

收稿日期: 2021-03-19

基金项目: 国家自然科学基金项目(61872126, 61772159); 河南省科技攻关计划项目(192102210123, 182102110333)

作者简介: 宋成(1980—), 男, 副教授.

通信作者: 贺军义(1982—), 男, 副教授, E-mail: hejunyi@hpu.edu.cn.

语义差异性;而 Li 等<sup>[3]</sup>则提出通过构造属性层次树为地图标注属性信息,以保证属性的差异性;Guo 等<sup>[4]</sup>选择利用历史相邻位置替代用户的真实位置,但会降低查询准确度。 $K$ 匿名技术是另一种常用的隐私保护技术,其原理是构造包含  $k$  个用户的匿名区域,以取代用户的真实位置,使攻击者追踪真实用户的概率不超过  $1/k$ 。为克服用户分布不均匀的问题, Ni 等<sup>[5]</sup>在密集区域和稀疏区域分别构造匿名域。李璐璐等<sup>[6]</sup>根据遗传算法和信息熵选择假位置并建立缓存机制,但需要占用终端较多的存储空间;而 Zhang 等<sup>[7]</sup>将  $k$  个查询位置随机映射到不同的匿名器,解决了单点故障问题,但总体开销有所增加。其他的隐私保护技术主要有动态假名、轨迹抑制、差分隐私等。其中混合区域是一种常用的动态假名技术,其原理是用户进入混合区域后停止服务,离开时更换假名,避免攻击者追踪用户,如 Memon 等<sup>[8]</sup>根据位置的相关度构建多重混合区域,实现了用户多样性,但区域内的用户必须停止 LBS 服务,因此,频繁进入混合区域会影响服务质量。轨迹抑制是通过隐藏频繁访问的敏感位置点来减少轨迹暴露风险的一种技术,实现方法简单但会造成轨迹数据丢失。对此, Terrovitis 等<sup>[9]</sup>提出将原始轨迹分割成较小轨迹,对敏感数据进行局部抑制。而差分隐私方法主要是向原有轨迹添加噪声,但添加随机噪声不能控制噪声大小和可用性,如 Zhao 等<sup>[10]</sup>提出建立轨迹前缀树,先添加拉普拉斯噪声,使其满足差分隐私需求,再计算位置敏感度,限制噪声大小。

## 1 预备知识

### 1.1 多值预测位置隐私保护模型

隐私保护模型中,用户通过移动终端携带的卫星定位系统获取自身位置信息,并在地图范围内选择  $n$  个兴趣点,根据马尔可夫链生成预测位置和预测查询,将包含伪查询和预测查询在内的  $N$  个请求发送给 LBS 服务器,如图 1 所示。LBS 服务器完成查询后,将加密查询结果返回给移动终端。用户对返回结果进行解密计算,获得相关查询信息并建立本地缓存,供下次 LBS 请求前检索。系统架构主要包括移动终端和 LBS 服务器两部分实体。

1) 移动终端。负责生成预测位置和预测查询内容,向 LBS 服务器发送查询请求,接收 LBS 服务器返回的查询结果以及建立本地缓存机制。

2) LBS 服务器。负责用户的匿名化,提供位置

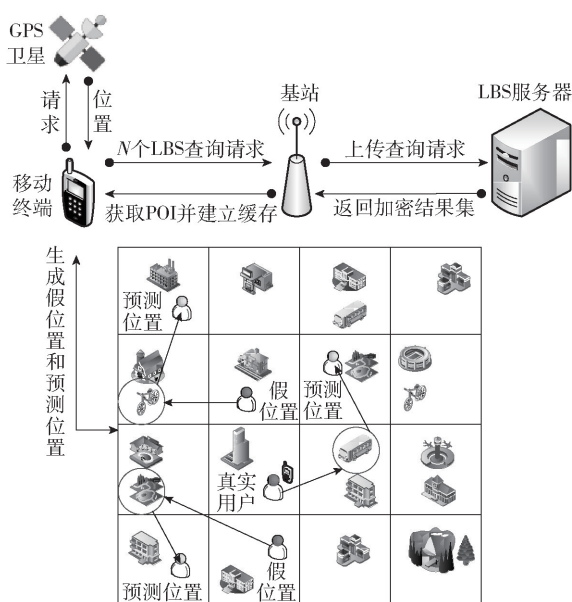


图1 多值预测位置隐私保护模型

服务查询和查询结果的加密。

### 1.2 马尔可夫链

马尔可夫过程是状态不断演变的过程,过程中每个状态的转移只依赖于之前的状态。对于一阶马尔可夫模型,第  $t+1$  时刻上的取值依赖且仅依赖于第  $t$  时刻的取值,则有  $X(t+1) = f(X(t))$ 。而时间和状态都是离散的马尔可夫过程称为马尔可夫链。

移动终端将获取的地图信息划分为  $M$  个大小相同的位置单元格,即  $\{C_1, C_2, \dots, C_M\}$ 。根据用户的移动模型,可通过马尔可夫链描述用户在不同时刻的空间位置关联性。地图信息中共有  $M$  个位置单元,即用户可能的位置状态有  $M$  种。用户移动过程中的每种位置状态转移都有一个由  $C_i$  向  $C_j$  移动的概率  $W_{ij}$  ( $i=1, 2, \dots, M, j=1, 2, \dots, M$ )。将所有的状态转移概率用状态转移矩阵表示为

$$\Delta = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1M} \\ W_{21} & W_{22} & \cdots & W_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ W_{M1} & W_{M2} & \cdots & W_{MM} \end{bmatrix}$$

假设  $t$  时刻地图上存在 4 名用户,分别位于位置单元  $C_1, C_2, C_4, C_5$ ,则  $t$  时刻对应的位置状态向量为  $P_t = \{0.25, 0.25, 0, 0.25, 0.25, \dots\}$ ,而  $t+1$  时刻的用户位置分布可通过  $P_{t+1} = P_t \Delta$  计算得到。

为了提高状态转移矩阵的准确性,可利用 Geolife 数据集提供的历史位置信息,根据  $P_{t+1} = P_t \Delta$  计算转移矩阵,即转移矩阵已知。

### 1.3 布隆过滤器

布隆过滤器是一种设计巧妙的概率型数据结构,由一个很长的二进制数组和一系列哈希函数组成,具有较低的空间复杂度和高效插入和查询的特征。其原理是插入一个元素时,首先通过  $k$  个哈希函数对其进行  $k$  次计算,得到  $k$  个哈希值,并在数组中将对应数组位的值置为 1;判断是否对布隆过滤器中的某个值进行  $k$  次哈希计算,根据所得的哈希值在位数组中判断每个元素是否为 1,若每个元素都为 1,说明该值存在于布隆过滤器中;否则,该值一定不存在于布隆过滤器中。因此,布隆过滤器返回的结果是概率性的,即判断一个值存在时,会存在误判的可能性。

设布隆过滤器的数组长度为  $l$ ,样本数量为  $\varphi$ ,哈希函数的个数为  $k = \ln 2 \frac{l}{\varphi}$ ,则其真实失误率为  $1 - e^{-\frac{ek}{l}}$ 。设样本数量不变,真实失误率随着数组长度的增大而降低。

## 2 多值预测查询位置隐私保护方案

方案包括系统初始化阶段、生成预测位置和查询内容阶段、匿名化阶段、位置服务请求阶段和布隆过滤器缓存 5 个阶段。

### 2.1 系统初始化阶段

**步骤 1** 选择 2 个阶为  $q$  的循环群  $G_1, G_2$ , 其中  $G_1$  为加法循环群,  $G_2$  为乘法循环群,  $q$  是一个大素数。  $e: G_1 \times G_2 \rightarrow G_2$  表示一个双线性映射。

**步骤 2** 定义 3 个哈希函数  $H_1, H_2$  和  $H_3$ 。其中  $H_1: \{0, 1\}^* \rightarrow G_1^*$ ,  $H_2: G_2 \rightarrow \{0, 1\}^z$ ,  $H_3$  为 SHA256 的哈希函数,  $z$  为一个整数,  $\{0, 1\}^*$  为任意长度的二进制串。

**步骤 3** LBS 服务器选取随机数  $s \in Z_q^*$  作为系统私钥, 计算公钥:  $PK = sp$ , 其中  $p$  为  $G_1$  的生成元。

**步骤 4** LBS 服务器保存系统私钥  $s$  和哈希函数  $H_3$ , 并公开系统公共参数:  $\{G_1, G_2, e, z, q, p, PK, H_1, H_2\}$ 。

### 2.2 生成预测位置和查询内容阶段

**步骤 1** 设用户的最大移动速度为  $V_{\max}$ , 连续查询的时间间隔为  $t_0$ , 以  $V_{\max} t_0$  为边长将地图信息划分成  $M$  个大小相同的方形位置单元, 记为  $\{C_1, C_2, \dots, C_M\}$ ,  $C_i = \{L_i, S_i\} = \{(x_i, y_i), S_i\}$ , 其中  $(x_i, y_i)$  表示位置单元格中心点的位置坐标,  $S_i$  表示兴趣点信

息。若位置单元格内的兴趣点不唯一, 则取与相邻单元语义类型差异性最大的兴趣点作为查询标识。

**步骤 2** 用户获取自身地理位置  $L_0 = (x_0, y_0)$ , 并在地图范围内选择一组用户关心的兴趣点  $\{S_1, S_2, \dots, S_n\}$ , 将其转化成对应的位置点  $\{L_1, L_2, \dots, L_n\}$ ; 然后取  $L_0$  所在位置单元的中心坐标  $F_0$  作为假位置代替用户真实位置, 并在相邻单元格内生成  $n-1$  个假位置  $\{F_1, F_2, \dots, F_{n-1}\}$ , 生成的假位置均满足  $\frac{\sqrt{2}}{2} V_{\max} t_0 \leq D \leq \frac{3}{2} V_{\max} t_0$ , 其中  $D$  为假位置到  $F_0$  的欧氏距离。

**步骤 3** 移动终端根据  $\{L_1, L_2, \dots, L_n\}$  对应的位置集合  $\{C_i\}$  建立  $t+t_0$  时刻的位置状态向量  $\mathbf{P}_{t+t_0}$ , 例如, 若  $t$  时刻的位置集合为  $\{C_1, C_2, C_6, C_8\}$ , 则有

$$\mathbf{P}_{t+t_0} = (0.25 \ 0.25 \ 0 \ 0 \ 0 \ 0.25 \ 0 \ 0.25 \ 0 \ \dots)$$

最后根据状态转移矩阵计算  $t+2t_0$  时刻的预测位置状态向量  $\mathbf{P}_{t+2t_0} = \mathbf{P}_{t+t_0} \mathbf{\Delta}$ , 并取其中状态转移概率较大的  $g$  个位置点位置作为预测位置单元集合  $\{C_{n+1}, C_{n+2}, \dots, C_{n+g}\}$ , 即得到对应的预测位置  $\{L_{n+1}, L_{n+2}, \dots, L_{n+g}\}$  和预测查询内容  $\{S_{n+1}, S_{n+2}, \dots, S_{n+g}\}$ 。

**步骤 4** 将  $L_\alpha \rightarrow (L_\beta, S_\beta)$  作为一个查询组合, 即构成 LBS 请求所需的位置信息和查询内容。移动终端首先取  $\alpha \in \{F_0\} \cup \{F_1, F_2, \dots, F_{n-1}\}$ ,  $\beta \in \{(L_1, S_1), (L_2, S_2), \dots, (L_n, S_n)\}$ , 交叉组合得到  $n^2$  个原始查询组合; 再取  $\alpha \in \{L_1, L_2, \dots, L_n\}$ ,  $\beta \in \{(L_{n+1}, S_{n+1}), (L_{n+2}, S_{n+2}), \dots, (L_{n+g}, S_{n+g})\}$ , 得到  $ng$  个预测查询组合, 即可生成  $N = n(n+g)$  个 LBS 请求。

### 2.3 匿名化阶段

**步骤 1** 移动终端用户向 LBS 服务器发送假名注册请求。LBS 服务器使用伪随机生成器生成盐值  $ID_{\text{salty}}$ , 并生成假名  $PID = H_3(ID_u + ID_{\text{salty}})$  和公私钥对  $U_{PK} = H_1(ID_u)$ ,  $U_{SK} = sU_{PK}$ , 然后将结果回传用户。

**步骤 2** 用户终端根据收到的注册结果计算  $U_{PK} = H_1(ID_u)$ , 并判断等式  $e(U_{SK}, p) = e(U_{PK}, PK)$  是否成立。若成立, 则注册成功; 否则, 返回步骤 1。

**步骤 3** 重复上述步骤直至成功生成  $N$  个假名。

### 2.4 位置服务请求阶段

**步骤 1** LBS 服务器随机选取  $d_1, d_2, \dots, d_j \in Z_q^*$ , 其中  $j \geq N$ , 分别计算  $p_1 = d_1 PK, p_2 = d_2 PK, \dots, p_j = d_j PK$ , 并将其作为选择基点公布。



**步骤2** 用户  $B$  随机选择  $a_1, a_2, \dots, a_N \in Z_q^*$ , 分别计算  $v_i = a_i p_i$ , 其中  $i = 1, 2, \dots, N$ .

**步骤3** 用户  $B$  根据假名查询组合构成查询集合  $\text{Msg} = \{(\text{PID}_1, L_1, Q_1, v_1), (\text{PID}_2, L_2, Q_2, v_2), \dots, (\text{PID}_N, L_N, Q_N, v_N)\}$ , 并发送给 LBS 服务器.

**步骤4** LBS 服务器收到位置服务请求后, 获取  $N$  个查询结果  $\{m_1, m_2, \dots, m_N\}$ , 然后随机选取临时会话私钥  $r \in Z_q^*$ , 并分别计算  $Y_0 = r\text{PK}$ ,  $Y_i = rv_i$  和  $c_i = m_i \oplus H_2(e(p_i + s\text{PK}, U_{\text{PK}}^i)^r)$ , 最后将加密结果集  $\{Y_0, (Y_1, Y_2, \dots, Y_N), (c_1, c_2, \dots, c_N)\}$  发送给用户  $B$ .

**步骤5** 用户  $B$  收到消息后, 判断等式  $e(v_i, Y_0) = e(Y_i, \text{PK})$  是否成立. 若成立, 用户  $B$  分别计算  $a_i$  的乘法逆元  $a_i^{-1} \in Z_q^*$ , 解密密钥  $V_i = a_i^{-1} Y_i$ ,  $m_i = c_i \oplus H_2(e(V_i, U_{\text{PK}}^i) e(Y_0, U_{\text{SK}}^i))$ , 其中,  $i = 1, 2, \dots, N$ , 从而获取查询结果并将该结果作为建立缓存的有效输入; 否则, 用户  $B$  放弃本轮查询结果, 返回步骤1 重新服务请求.

## 2.5 布隆过滤器缓存阶段

**步骤1** 定义一个整型数组  $\text{DB}[l]$  作为缓存检索表, 将所有位初始为 0, 并定义  $k$  个独立的哈希函数  $\{h_1, h_2, \dots, h_k\}$ .

**步骤2** 利用  $\{h_1, h_2, \dots, h_k\}$  对查询兴趣点  $S$  分别进行哈希计算, 得到一组哈希值序列  $\{h_1(S), h_2(S), \dots, h_k(S)\}$ . 根据该序列将对应的查询结果  $m_i$  加入缓存文件列表  $\text{DB}_{\text{res}}$  中, 生命周期为  $T = \sigma t_0$ ,  $\sigma \in \mathbf{N}^+$ , 并将数组的对应位加 1, 即

$$\text{DB}[h_i(S) \bmod l] + 1, i = 1, 2, \dots, k$$

**步骤3** 重复步骤2 直至将  $N$  个查询结果全部存入缓存文件的列表中.

**步骤4** 用户需要进行新一轮 LBS 请求时, 首先对欲查询组合进行同样的哈希计算, 判断所得哈希序列在  $\text{DB}[l]$  中对应位置的元素是否均不为 0. 若满足, 则从  $\text{DB}_{\text{res}}$  中获取对应的  $m_i$ ; 否则, 将缓存中所有文件的生命周期衰减  $t_0$ , 发起新的 LBS 请求.

**步骤5** 当  $\text{DB}_{\text{res}}$  中的文件生命周期衰减为 0 时, 从缓存中删除该文件, 并将对应数组位减 1.

## 3 安全性分析

### 3.1 匿名性

**定义1** 设攻击者  $A$  赢得游戏的优势为  $\text{Adv}(A) = |\Pr(A) - 1/N|$ , 其中:  $\Pr(A)$  表示攻击者

$A$  能够输出  $m'_u = m_u$  的概率, 即攻击者  $A$  能够获取用户真实查询结果的概率;  $1/N$  为攻击者  $A$  从  $N$  个消息中随机选择一个输出使得  $m'_u = m_u$  的概率.

**定理1** 方案中, 如果攻击者  $A$  无法取得随机概率  $1/N$  以外的优势破解用户  $B$  的消息, 则该方案满足匿名性.

**证明** 若攻击者  $A$  试图直接解密密文则攻击者  $A$  必须获取用户  $B$  选择的基点  $P_u$ 、服务器私钥  $s$  和临时随机数  $r$ . 由于基点信息由 LBS 服务器随机选择且不公开, 攻击者  $A$  只能随机猜测; 对攻击者  $A$  而言, 若想根据  $\text{PK} = sp$  和  $Y_0 = r\text{PK}$  求解  $\{r, s\}$ , 则求解该问题等价于求解椭圆曲线离散对数难题. 因此, 攻击者  $A$  能够获取用户  $B$  隐私信息  $m_u$  的概率为  $\text{Adv}(A) = |W_r(A) - 1/N|$ , 可以忽略不计, 即满足匿名性.

### 3.2 不可伪造性

**定义2** 设攻击者  $A$  伪造用户私钥使等式成立的概率为  $W_r(A)$ . 若攻击者  $A$  以无法忽略的  $W_r(A)$  伪造成功, 则攻击者  $A$  赢得游戏; 否则, 方案满足不可伪造性.

**定理2** 随机预言模型中, 若攻击者  $A$  只能以可忽略的概率  $W_r(A)$  在多项式时间内求解服务器私钥  $s$  和临时会话私钥  $r$ , 则方案满足不可伪造性.

**证明** 攻击者  $A$  在询问过程中无法直接获取系统私钥  $s$ . 若攻击者  $A$  欲通过  $\{p, \text{PK}\}$  和  $\text{PK} = sp$  求解  $s$ , 将面临求解椭圆曲线离散对数难题, 即  $W_r(A)$  可忽略不计. 同理, 位置服务请求阶段, 系统临时会话密钥  $r$  对攻击者  $A$  保密, 攻击者  $A$  无法伪造正确的  $Y'_u$  通过判别式验证. 若攻击者  $A$  欲通过  $Y_0 = r\text{PK}$  推导  $r$ , 难度等同于求解椭圆曲线离散对数难题. 综上所述, 攻击者  $A$  无法以不可忽略的概率  $W_r(A)$  赢得游戏, 该方案满足不可伪造性.

### 3.3 抵抗查询服务跟踪

**定义3** 攻击者  $A$  通过比较不同时刻匿名区域内用户集合的交集来推测用户  $B$  的身份信息, 该攻击方式称为查询服务跟踪.

**定理3** 设用户  $B$  在移动轨迹中连续查询  $Q$  次, 即产生  $Q$  个不同时刻的匿名域, 且每次查询发送  $N$  个 LBS 请求. 攻击者  $A$  根据用户假名  $\text{PID}_i$  求解  $\text{ID}_i$  的概率为  $W_r(\text{PID}_i)$ , 通过截取用户  $B$  和 LBS 服务器之间通信解密消息的概率为  $W_r(A)$ , 则在不考虑缓存命中的情况下, 连续查询过程中攻击者成功追踪用户身份的概率为

$$W = \prod_{i=1}^Q \frac{1}{N} W_r(A) W_r(PID_i)$$

如果攻击者追踪成功的概率可忽略不计,则该方案能抵抗查询服务追踪。

**证明** 所提方案中,连续请求 LBS 过程中,用户  $B$  每次发起请求前都会更换假名,其中,  $W_r(PID_i)$  等价于破解 SHA256 哈希函数的难度,且攻击者  $A$  至少需要破解连续  $Q$  个假名,其概率可以忽略。在位置服务传输阶段,攻击者  $A$  破解加密消息的概率  $W_r(A)$  等价于破解椭圆曲线的密码体制,在计算上不可行。与此同时,整个方案过程中采用了  $K$  匿名的思想,根据定理 1 的证明过程可知,攻击者  $A$  从每个时刻的匿名区中识别用户查询消息的概率不超过  $1/N$ 。此外,所提方案采用了缓存机制,当缓存生效时,用户无需向 LBS 服务器发送服务请求消息,即攻击者所截获的消息是不连续且不完全的,这将进一步阻止攻击者关联用户的真实身份。综上所述,攻击者  $A$  通过查询服务跟踪获取用户真实身份的概率可以忽略不计,即所提方案可抵抗查询服务跟踪。

## 4 仿真实验

基于位置隐私保护系统架构模型,从缓存命中率和系统性能两方面对方案进行仿真实验。仿真实验在 64 位的 Windows10 操作系统、2.80 GHz 英特尔 i7 CPU、8 GB 内存的 PC 机和 Matlab 仿真软件的环境下进行。实验选取 5 km × 5 km 的真实地理地图,设用户单位查询间隔时间内的最大移动距离为 500 m,将地图划分成 100 个位置单元格,即样本数量的最大取值为 100。默认实验参数  $l = 2\ 048$ ,  $k = 4$ ,  $n = 3$ ,  $g = 5$ ,实验中的主要影响参数为查询次数 ( $Q$ ),表示连续查询过程中终端请求 LBS 的总次数。

### 4.1 缓存命中率

缓存机制的性能取决于缓存命中率,缓存命中率是指缓存机制提供 LBS 查询结果的概率。定义连续查询过程中缓存机制提供 LBS 查询结果的次数为  $Q_c$ ,由 LBS 服务器提供查询结果的次数为  $Q_s$ ,总查询次数  $Q = Q_c + Q_s$ ,取值范围为  $[10, 80]$ ,则缓存命中率为  $\gamma = \frac{Q_c}{Q}(1 - \varepsilon)$ ,其中  $\varepsilon$  为缓存真实失误率。

如果将全部样本加入缓存,当缓存失误率最大时,真实失误率约为 0.001 3,可忽略不计。文献[2]方案未采用缓存机制,故缓存命中率始终为 0。随着查询

次数增加,所提方案的缓存命中率始终优于文献[6]方案,但优势逐渐减小,这是由于文献[6]方案在终端的缓存信息数较高,需要占用终端更多的存储空间。所提方案通过增大预测参数  $g$  或延长缓存生命周期  $T$  可以提高终端缓存数,进而提高缓存命中率。仿真结果如图 2 所示。

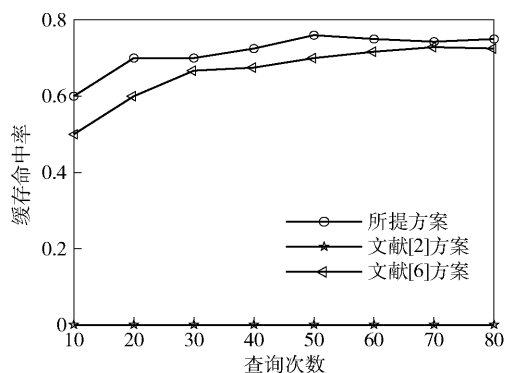


图2 查询次数与缓存命中率的关系

## 4.2 系统性能

### 4.2.1 通信开销

LBS 查询过程中,通信开销主要由 LBS 请求的数据大小和查询次数决定,所提方案由缓存提供 LBS 服务时,无需与 LBS 服务器通信,即通信开销为 0。假设各方案连续查询过程中查询请求的格式和大小固定,则通信开销的大小取决于查询次数,选取范围为  $[10, 80]$ 。由图 3 所示的仿真结果可见,查询次数较小时,所提方案 and 对比方案间无明显差距。随着查询次数增加,由于文献[2]方案未采用缓存机制,故其通信开销呈线性增长。所提方案和文献[6]方案缓存中的纪录逐渐增加,由缓存提供查询结果的概率提高,其通信开销增加幅度明显低于文献[2]方案。所提方案的缓存命中率高过文献[6]方案,但由于匿名化过程增加了注册成本,故通信开销方面的优势不明显,但安全性更高。

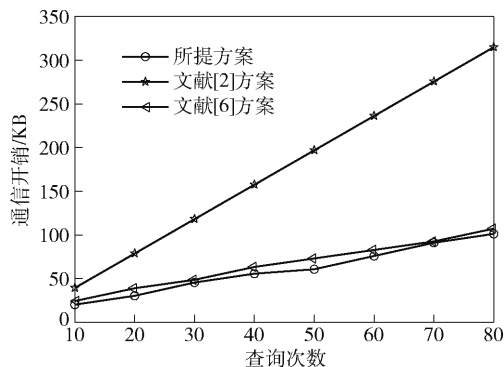


图3 通信开销与查询次数的关系

### 4.2.2 效率

系统效率指用户完成一次位置服务花费的时间。所提方案的执行时间主要由预测位置和查询内容生成阶段以及缓存阶段决定。当由缓存提供 LBS 服务时,无需调用预测位置和查询内容生成选取算法,只需要从缓存列表中匹配相应的兴趣点信息;缓存未命中时,则执行预测查询生成算法并建立缓存。连续查询过程中,预测位置和查询内容生成算法的执行时间取决于生成预测位置的数量和查询次数。假设各方案生成的混淆位置的数量相同且固定,则系统效率由查询次数决定,其选取范围为[10,80]。从图4所示的结果看,未采用缓存机制的方案<sup>[2]</sup>始终需要执行算法复杂度更高的假位置生成选取算法,故执行时间明显优于其他方案,且由于每次生成的假位置数量固定,假位置生成选取算法的平均执行时间变化不大,随查询次数的增加呈现线性增长。所提方案与文献[6]方案在查询次数较小时无明显差距,随着查询次数的增加,缓存在终端的兴趣点数量增加,缓存命中率趋于稳定,方案的效率优势逐渐体现出来。

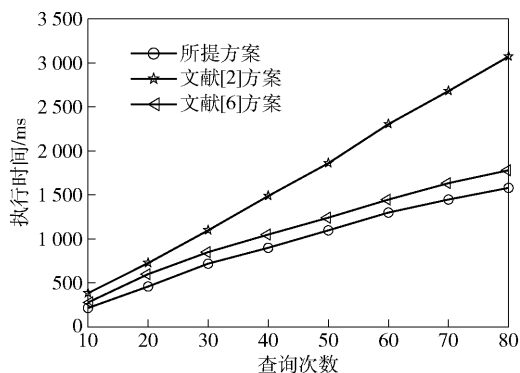


图4 执行时间与查询次数的关系

## 5 结束语

提出了一种多值预测查询位置隐私保护机制,通过马尔可夫链实现了多值预测查询,并设计了一种匿名查询组合方法,增强混淆性的同时,扩展了查询的可选择性。根据布隆过滤器的原理建立缓存机制,解决了连续匿名查询过程中的结果冗余问题,减少了终端和 LBS 服务器的交互次数。背景知识攻击对传统匿名化技术的安全性造成了冲击,而差分隐私的方法为弥补这一漏洞提供了解决思路。如何保证安全性的同时降低差分隐私对查询精度的影响将

会是今后的工作重点。

### 参考文献:

- [1] Peng Tao, Liu Qin, Wang Guojun, et al. Multidimensional privacy preservation in location-based services[J]. Future Generation Computer Systems, 2019, 93: 312-326.
- [2] 王洁, 王春茹, 马建峰, 等. 基于位置语义和查询概率的假位置选择算法[J]. 通信学报, 2020, 41(3): 53-61.  
Wang Jie, Wang Chunru, Ma Jianfeng, et al. Dummy location selection algorithm based on location semantics and query probability[J]. Journal on Communications, 2020, 41(3): 53-61.
- [3] Li Weihao, Li Chen, Geng Yeli. APS: attribute-aware privacy-preserving scheme in location-based services[J]. Information Sciences, 2020, 527: 460-476.
- [4] Guo Xueying, Wang Wenming, Huang Haiping, et al. Location privacy-preserving method based on historical proximity location[J]. Wireless Communications and Mobile Computing, 2020(8): 1-16.
- [5] Ni Lina, Tian Fulong, Ni Qinghang, et al. An anonymous entropy-based location privacy protection scheme in mobile social networks[J]. EURASIP Journal on Wireless Communications and Networking, 2019(1): 93-96.
- [6] 李璐璐, 华佳烽, 万盛, 等. 基于高效信息缓存的位置隐私保护方案[J]. 通信学报, 2017, 38(6): 148-157.  
Li Lulu, Hua Jiafeng, Wan Sheng, et al. Achieving efficient location privacy protection based on cache[J]. Journal on Communications, 2017, 38(6): 148-157.
- [7] Zhang Shaobo, Mao Xinjun, Choo K K R, et al. A trajectory privacy-preserving scheme based on a dual-K mechanism for continuous location-based services[J]. Information Sciences, 2020, 527: 406-419.
- [8] Memon I, Mirza H T, Arain Q A, et al. Multiple mix zones decorrelation trajectory privacy model for road network[J]. Telecommunication Systems, 2019, 70(4): 557-582.
- [9] Terrovitis M, Poulis G, Mamoulis N, et al. Local suppression and splitting techniques for privacy preserving publication of trajectories[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(7): 1466-1479.
- [10] Zhao Xiaodong, Pi Dechang, Chen Junfu. Novel trajectory privacy-preserving method based on prefix tree using differential privacy[J]. Knowledge-Based Systems, 2020: 105940.