doi:10.11959/j.issn.1000-436x.2017129

基于高效信息缓存的位置隐私保护方案

李璐璐[1]，华佳烽[2]，万盛[2]，朱辉[2]，李凤华[2]
(1. 西安电子科技大学通信工程学院，陕西 西安 710071；2. 西安电子科技大学网络与信息安全学院，陕西 西安 710071)

摘 要：随着移动互联网的高速发展与智能终端的迅速普及，基于位置的服务（LBS, location-based service）已在导航、社交等领域得到广泛应用。但由于个人位置信息的敏感性，移动对象位置相关的数据隐私保护已经成为 LBS 中的研究热点。综合考虑用户所处区域背景信息的开放性特征，引入信息缓存机制，提出一种虚假位置选择算法，有效减少用户与不可信服务器间的交互次数，并结合 *k*-匿名技术实现了用户位置信息的高效隐私保护。相对于现有技术，该方案不依赖任何可信第三方，可实现抵御拥有背景信息攻击者的推理攻击。实验结果证明了所提方案的有效性和高效性。

关键词：基于位置的服务；隐私保护；背景信息；缓存机制
中图分类号：TN929.5 **文献标识码**：A

Achieving efficient location privacy protection based on cache

LI Lu-lu[1], HUA Jia-feng[2], WAN Sheng[2], ZHU Hui[2], LI Feng-hua[2]
(1. School of Telecommunications Engineering, Xidian University, Xi'an 710071, China;
2. School of Cyber Engineering, Xidian University, Xi'an 710071, China)

Abstract: With the development of mobile Internet and the popularization of intelligent terminals, location based services(LBS) has been widespread in navigation, social network and other fields. Due to the sensitivity of personal location information, data privacy protection related to location of mobile objects has become the hotspots of research. Considering the openness of the background information, and based on the information cache mechanism, a dummy selection algorithm was proposed to reduce the number of interactions between the user and the untrusted server and combine the *k*-anonymity to achieve efficient location privacy. Without relying on trusted third party, the scheme can prevent the attackers owned background information from inference attack, and the detail simulation results indicate its effectiveness and efficiency.

Key words: location-based service, privacy protection, background information, cache mechanism

1 引言

网络技术、信息技术的持续快速发展，改变了文化教育、医疗卫生、公共交通等领域的服务模式和人们的学习、工作、生活方式，催生出各类新模式、新服务。基于位置的服务为用户实时提供位置情境信息服务，极大地改善了人们的生活，已经成为移动互联网的主要功能之一。

在整个基于位置的服务过程中，通常存在位置隐私（location privacy）和查询隐私（query privacy）这 2 类信息泄露。其中，位置隐私是指与用户位置信息相关的敏感信息，包括实时位置、用户访问过的位置、由位置信息推断出的其他敏感信息；查询隐私则是指与查询内容相关的敏感信息，包括用户

收稿日期：2017-01-13；**修回日期**：2017-03-10
通信作者：李凤华，lfh@iie.ac.cn
基金项目：国家自然科学基金资助项目（No.61672411）；国家自然科学基金委—广东联合基金资助项目（No.U1401251）；陕西省自然科学基金资助项目（No.2016JM6007）；北京市自然科学基金资助项目（No.4152048）
Foundation Items: The National Natural Science Foundation of China (No.61672411), The Key Program of the National Natural Science Foundation of China-Guangdong Union Foundation (No.U1401251), The Natural Science Foundation of Shaanxi Province (No.2016JM6007), The Natural Science Foundation of Beijing (No.4152048)

偏好和需求等信息。

目前，位置隐私保护已引起了学者的广泛关注，并提出了一系列隐私保护方案，主要分为 3 种结构：基于 TTP（trusted third party）的集中式网络结构[1~3]、P2P 网络（peer-to-peer network）结构[4~6]和独立结构[7]。其中，基于 TTP 结构的隐私保护方案利用 TTP 将原始的精确位置信息模糊化，以实现用户的位置隐私保护，但该结构中的中心服务器可信度无法得到保证，且容易遭受单点攻击而成为系统瓶颈；P2P 网络结构的隐私保护方案借助近邻用户位置信息实现位置隐私保护，具有良好的容错性和拓展性，无需可信第三方，但是 P2P 资源管理与通信开销控制等问题仍然是个重大挑战；独立结构的隐私保护方案利用自身的能力和知识进行位置隐私保护，具有结构简单，容易与其他技术结合的特点，但是对客户端要求较高，增加客户端负担。

基于上述分析，本文综合考虑用户所处区域背景信息的开放性特征，引入信息缓存机制，提出一种虚假位置选择算法，并基于该算法设计一种不依赖可信第三方的隐私保护方案。该方案通过减少用户与不可信服务器间的交互次数，并结合 *k*-匿名技术，以实现用户位置信息的高效隐私保护。

2　相关工作

近年来，针对 LBS 中的各种隐私信息主要采用基于隐私保护策略、密码学、隐私保护模型等手段来设计隐私保护方案。其中，基于隐私保护策略的方案[8~10]能够解决特定情境下的隐私场景，但不同场景的 LBS 应用系统的隐私保护策略缺乏通用性；基于密码学的方案[11~15]无需匿名服务器，虽然隐私保护效果较好，但计算开销太大，并不适用于资源受限的移动终端，实用性较差；基于隐私保护模型的方案通过在特定模型和算法基础上将用户真实位置信息模糊化或虚假化，以实现用户位置信息保护[16]，具有计算开销较小、实用性强的优点，已成为位置隐私保护的主要手段，其主要采用空间匿名、位置偏移、模糊或伪造虚假位置、群组协作等方法和 *k*-匿名[17]、SpaceTwist[18]等模型。

为实现连续场景下的位置隐私保护，Chow 等[19]提出了一种基于 P2P 的解决方案，通过用户间相互协作，利用近距离通信技术（Wi-Fi 或蓝牙）实现信息交换，并在考虑用户的最大移动距离基础上，实现连续场景下的 *k*-匿名，但由于用户的移动模式和通信距离限制，使用户真实的位置会以较大概率落在匿名集的中心区域。Kido 等[20]则通过将用户真实位置和基于随机运动所产生的多个虚假位置一起发送给服务提供商，使服务提供商无法区分用户真实位置和虚假位置，从而实现隐私保护，但该方案忽视了攻击者所掌握的背景信息（如查询概率）和虚假位置的选取质量问题，容易遭受关联攻击。Niu 等[21]以每个位置的查询概率作为攻击者所掌握的背景信息，设计了一种基于虚假位置和用户协作的隐私保护方案，然而该方案的通信开销和存储开销均较高。

为提升隐私保护方案的效率，霍峥等[22]设计了一种缓存访问机制为签到序列建立前缀树，实现了假名用户的轨迹隐私保护，且签到位置损失较少，有效保证了用户体验。Amini 等[23]通过预先缓存某一特定区域的服务数据，当用户使用基于位置的服务时，直接从缓存中获取服务信息而无需通过服务提供商，但该方案存储开销随区域大小线性增长，可扩展性较差。Shokri 等[24]提出利用移动终端中的缓存装置存储其他用户的查询请求信息和服务提供商返回的应答结果，当缓存装置中的数据无法满足需求时，用户向其他用户发送查询请求，如果仍无法获取服务内容，才向服务提供商发送查询请求，该方案减少了用户与服务提供商间的交互次数，但缓存命中率较低，且容易遭受推理攻击。Zhu 等[25]提出通过增大移动终端中的缓存容量，以提高缓存命中率，但该方案忽略了实际场景中背景信息的因素，攻击者可通过推理攻击获取用户的真实位置。Niu 等[26]提出当用户无法在自身或 ad hoc 网络中其他用户的缓存装置中获取服务数据时，可通过执行“空间隐匿算法”逐条将查询请求发送给某个虚拟请求者，由该虚拟请求者向服务提供商发起查询请求，从而保护用户的真实位置信息，但通信开销较大。

综上所述，现有的基于缓存机制的隐私保护方法没有充分考虑攻击者的背景信息和缓存命中率等问题。本文旨在提出一种基于高效信息缓存的隐私保护方案，通过定性分析影响隐私保护效果的因素，结合 *k*-匿名和信息复用技术，增大缓存命中率，提高隐私保护效果。

3　预备知识

3.1　背景信息

背景信息[27]是指用户在某一具体位置或特定

区域发送 LBS 查询请求的概率。

具体来说，在一幅包含 $N\times N$ 个位置单元的地图中，位置单元 l_i 的查询概率 q_i 可以表示为所有用户 u 对该位置单元的查询次数 n_i 与整个地图上总的查询次数 m 的比值，即

$$q_i = \frac{n_i}{m}$$

其中，i=1, 2, …, N^2，且满足 $\sum_{i=1}^{N^2} q_i = 1$。

3.2 位置熵

在不考虑背景信息的情况下，当用户利用 k-匿名技术直接向服务提供商发送查询请求时，服务提供商推断出用户真实位置的概率为 $\frac{1}{k}$，其中，k 表示 k-匿名技术中位置集合的大小。假设 p_i 表示位置单元 loc_i 是用户真实位置的概率，则

$$p_i = \frac{q_i}{\sum_{i=1}^{k} q_i}$$

其中，i=1, 2, …, k，且满足 $\sum_{i=1}^{k} p_i = 1$。

定义 1 位置熵。即攻击者从匿名位置集合中推断出用户真实位置的平均不确定度。如式(1)所示，计算 H 作为位置熵，其中，i=1, 2, …, k。

$$H = -\sum_{i=1}^{k} p_i \operatorname{lb} p_i \tag{1}$$

一个系统越混乱无序，熵就越大，故位置熵 H 越大，代表系统隐私保护效果越好。当所有的 p_i 相等时，位置熵 H 最大。

3.3 遗传算法

遗传算法可以不依赖问题的领域和种类，构造出求解复杂系统优化的通用框架，以适应度函数为依据，通过对群体中的个体施加遗传操作，实现群体内部结构重组。其最优解算法的构造步骤如下。1) 确定决策变量及其约束条件；2) 建立优化模型；3) 确定表示可行解的染色体编码方法；4) 确定解码方法；5) 确定个体适应度的量化评价方法；6) 设计遗传算子；7)确定遗传算法的有关运行参数。其中，可行解的编码方法、适应度函数和遗传算子的设计是构造遗传算法时需要考虑的核心问题。

4 系统模型与安全需求

4.1 系统模型

如图 1 所示，在基于缓存的 LBS 应用场景中，以缓存机制为基础，用户利用移动智能终端获取所处的位置信息，当需要位置服务时，若移动终端的缓存中有能够满足用户查询请求的数据，则直接返回用户所需的服务数据，用户无需向不可信的服务

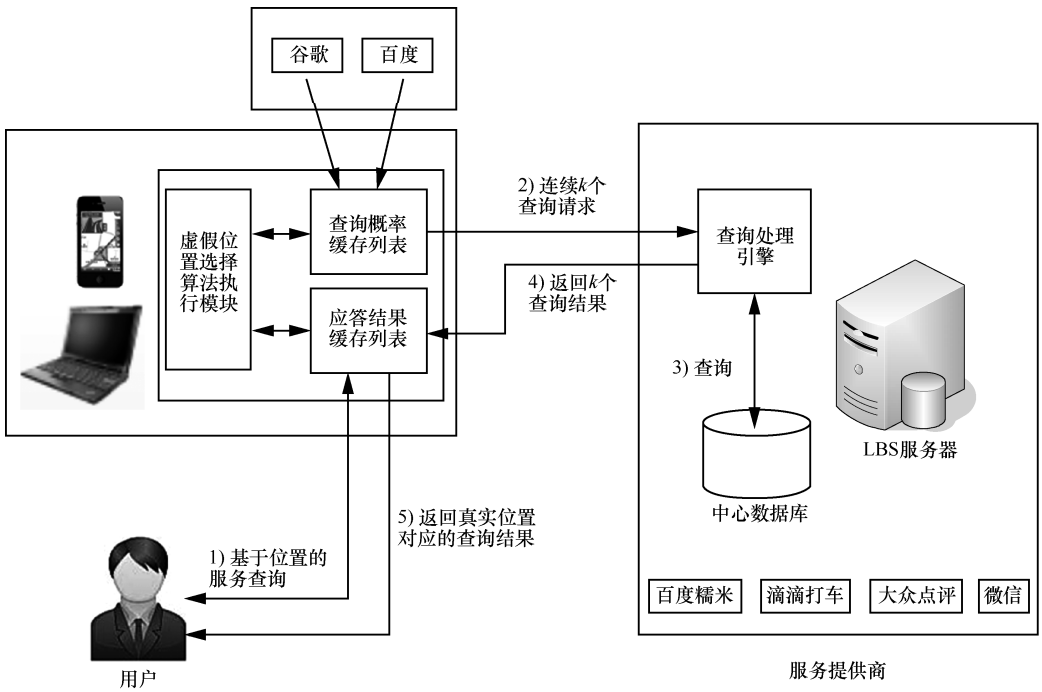

图 1 基于缓存的 LBS 应用场景

提供商发送查询请求；若移动终端的缓存中没有能够满足用户查询请求的数据，则智能终端利用 *k*-匿名技术构造匿名位置集合，并向服务提供商发起查询请求，服务提供商根据查询请求返回相应的应答结果；移动终端存储应答结果，同时提取所需信息返回给用户。

具体来说，在连续场景下，用户基于不同位置频繁发出查询请求，尤其是在某些时间段（如午餐时间），同一区域中的大量用户同时发送基于位置的服务请求（如查询附近餐厅）。对用户而言，其希望在得到高质量服务的同时，保护自己的位置隐私；对服务提供商而言，在为用户提供高效位置服务的同时，其通过分析用户位置信息而获取巨大的商业价值。因此，从用户角度出发，服务提供商是不可信的。

4.2　攻击模型

由于背景信息的开放性，攻击者所掌握的背景信息量不少于正常用户。根据文献[21]给出的攻击者描述对攻击者进行以下定义。假定不可信的服务提供商为攻击者 *A*。*A* 具有轻易获取背景信息和用户隐私保护机制的能力，可利用概率分布攻击、同源攻击和位置相似性攻击等手段获取用户位置信息。其具体攻击思路描述如下。

概率分布攻击。攻击者通过收集用户的查询请求，由此计算出其在不同区域的查询概率。当用户发送查询请求时，若隐匿区域内概率分布不均匀，则攻击者可以根据查询概率分布推断出用户的真实位置。

同源攻击。攻击者得知用户在同一位置多次发送查询请求，通过收集查询信息中的匿名位置集合，结合位置隐私保护机制，攻击者可计算出最有可能产生这一系列请求的用户真实位置。

位置相似性攻击。攻击者通过分析匿名位置集合中的语义信息，若该集合仅包含一种语义信息（如医院或学校等），则攻击者可推断出用户的行为。

4.3　设计思路

典型的基于信息缓存的位置服务场景具有信息价值高、用户查询请求频繁及同一区域查询请求密集海量等特点，通过对上述特点进行分析，可以发现用户运动轨迹易泄露、服务提供商不可信、LBS 服务器效率较低等安全隐患和缺陷。为解决上述问题，在综合考虑用户所处区域背景信息开放性特征的基础上，假设不可信服务提供商为攻击者 *A*，利用缓存机制，减少用户与不可信服务提供商间的交互次数，降低隐私泄露风险。

相较于基于同态/半同态的位置隐私保护技术，*k*-匿名技术由于计算开销较小，比较适用于资源受限的移动终端。然而，尽管已有的 *k*-匿名方案也通过引入缓存机制以降低通信开销，但其大多存在易遭受推理攻击和缓存数据复用率低的问题。因此，本文聚焦于匿名位置集合的选取优化和缓存访问的命中率提升这 2 个方面，综合考虑背景信息、归一化距离、数据新鲜度等特征，结合遗传算法思想，提出一种高效的虚假位置选择算法，并基于该算法设计一种不依赖任何可信第三方的隐私保护方案，实现有效减少用户与不可信服务提供商间的交互次数和抵御拥有背景信息攻击者的推理攻击。

5　位置隐私保护方案

本节首先给出衡量缓存机制隐私保护效果的度量标准的相关定义，通过对隐私保护效果进行定性分析，找出影响隐私保护效果的因素；然后结合遗传算法思想，提出一种高效的虚假位置选择算法；最后，基于虚假位置选择算法设计一种基于高效信息缓存的位置隐私保护方案（ELPPC, efficient location privacy protection based on cache）。

5.1　隐私保护效果度量标准定义

根据文献[27]中对 LBS 隐私保护效果的描述，给出下述定义。

定义 2　位置集合 *S* 表示从地图中随机选取|*S*|个位置单元，构成位置集合，其中，|*S*|表示集合 *S* 的大小。

定义 3　缓存命中率表示由缓存机制为移动用户提供 LBS 的概率。将由缓存提供的 LBS 次数记为 Q_{cache}，由服务提供商提供的 LBS 次数记为 Q_{server}，则缓存命中率为

$$\gamma = \frac{|Q_{\text{cache}}|}{|Q_{\text{cache}}| + |Q_{\text{server}}|} \tag{2}$$

缓存命中率越高，用户与服务提供商间的交互次数越少，基于缓存机制的位置隐私保护效果越好。

定义 4　缓存贡献度 δ 表示位置单元 *loc* 的查询概率 *q* 对缓存命中率的影响程度。

$$\delta = qg \tag{3}$$

其中，*g* 是一个调节参数。如果位置单元 *loc* 已在缓存中，再将该位置单元加入缓存列表中不会提高

缓存命中率，故此时 g=0；否则 g=1。位置单元的查询概率 q 越大，其贡献度 δ 也就越高。缓存贡献度将用于评判所选取的匿名位置集合的优劣。

定义 5 隐私保护效果 λ 表示引入缓存机制后整个系统的隐私保护效果。

$$\lambda=\frac{\sum_{q\in Q_{\text{server}}}H_q}{|Q_{\text{cache}}|+|Q_{\text{server}}|}+\gamma \text{lb}N^2 \tag{4}$$

其中，H_q 表示匿名位置集合的位置熵。由 λ 可知，可通过增大匿名位置集合的位置熵和缓存命中率这 2 种途径提高系统的隐私保护效果。为保证缓存机制的隐私保护效果最好，所选取的匿名位置集合需同时保证 H_q 和 γ 最大。

根据归一化距离（normalized distance）的定义，匿名位置集合中所选取的 k−1 个虚假地址对缓存的影响可表示为

$$D=\prod_{i=1}^{k-1}\sqrt{2\pi}\frac{d_i}{d(loc_t,loc_i)} \tag{5}$$

其中，d_i 表示用户真实位置 loc_t 和匿名集合中第 i 个虚假位置 loc_i 间的归一化距离，$d(loc_t, loc_i)$表示 loc_t 和 loc_i 间的物理距离。

根据数据新鲜度（data freshness）的定义，匿名位置集合中位置单元的平均数据新鲜度 F 可记为

$$F=\frac{\sum_{i=1}^{lk}f_i}{lk} \tag{6}$$

其中，f_i 表示第 i 个位置单元 loc_i 的新鲜度，l 表示匿名位置集合与缓存中位置单元的重复个数。

5.2 虚假位置选择算法

为实现缓存命中率最大，结合遗传算法的思想，提出一种基于信息缓存的虚假位置选择算法。

首先根据位置单元的背景信息，选取和用户真实位置 loc_t 查询概率 q_t 相近的 4k 个位置单元；为保证算法随机性，再从 4k 个位置单元中随机选取 2k 个位置单元构成位置集合 Q，其中，$|Q|$=2k；为使缓存命中率最大，需从 Q 中随机选取 k−1 个位置单元构成位置集合 P，选取最优的一组 $P_{\max}=\{loc_1, loc_2, \cdots, loc_{k-1}\}$与用户真实位置 loc_t 构成匿名位置集合 $S_{\max}$，以满足

$$S_{\max}=\arg\max\left(\sum_{i=1}^{k}\delta_i\right)(1-D)(1-F)$$

其中，$|P|=k-1$，$|S_{\max}|=k$，δ_i 表示位置单元 loc_i 的缓存贡献度，为便于算法步骤的描述，结合遗传算法的思想，对算法中的交叉和变异操作给出以下定义。

定义 6 交叉操作。在随机选取的位置集合 P_i、P_j 中随机选定一个交叉点，将交叉点两侧数据进行交叉拼接，产生新的集合 P_s，其中，$i, j, s\in\{1, 2, \cdots, C_{2k}^{k-1}\}$。如选择交叉点为 P_i、P_j 中第 r 个位置单元 loc_{ir} 和 loc_{jr}，经过交叉拼接后生成 2 个新的位置集合 P_i' 和 P_j'分别为$\{loc_{j1}, loc_{j2}, \cdots, loc_{j(r-1)}, loc_{ir}, loc_{i(r+1)}, \cdots, loc_{i(k-1)}\}$和$\{loc_{i1}, loc_{i2}, \cdots, loc_{i(r-1)}, loc_{jr}, loc_{j(r+1)},\cdots, loc_{j(k-1)}\}$。

定义 7 变异操作。随机选取 2 个位置集合 P_m 和 P_n，从集合 P_m 中随机选择一个位置单元 loc_e，选取集合 P_n 中位置单元 loc_e 之前的 $step$ 个位置单元或之后的 $step$ 个位置单元以替换集合 P_m 中相应的位置，形成新的位置集合 P_m'。其中，$m, n\in\{1, 2, \cdots, C_{2k}^{k-1}\}$，$e\in\{1, 2, \cdots, k-1\}$，$loc_e\in P_m, P_n$，$step$ 为算法变异步长(在实验部分可以根据计算量和算法准确度的需求进行调整)。

从位置集合 Q 中选取最优位置集合 $P_{\max}$ 的具体过程概括如下。

1) 备选位置集合初始化。从位置集合 Q 中随机选取 $popsize$ 组位置集合$\{P_1, P_2, \cdots, P_{popsize}\}$与用户真实位置 loc_t 构成备选位置集合$R=\{S_1, S_2, \cdots, S_{popsize}\}$，该过程称为备选位置集初始化。其中，$P\subset Q$，$|P|=k-1$，$S=\{loc_1, loc_2, \cdots, loc_{k-1}, loc_t\}$，$|R|=popsize$（$popsize$ 可根据不同参数下的算法需求设定）。

2) 位置集合筛选。计算备选位置集合 R 中每组位置集合 S_i 的总贡献度

$$scores(S_i)=\left(\sum_{i=1}^{k}\delta_i\right)(1-D)(1-F)$$

其中，i=1, 2, …, $popsize$。选择贡献度较高的前 $popsize\cdot e$ 个位置集合，e 为预设的存活率，淘汰剩下的位置集合。

3) 交叉、变异操作。根据预设的交叉概率 p_c 和变异概率 p_m，运用交叉和变异这 2 种操作方式，对步骤 2)中经筛选后剩下的位置集合进行改变，产生新的位置集合，并使位置集合的个数重新达到 $popsize$。

4) 算法收敛准则。算法预设定循环的轮数为 $maxiter$，若没有达到 $maxiter$ 次，则返回步骤 2)继续循环；若达到 $maxiter$ 次，则进行步骤 5)。

5) 选取一组满足 argmax($scores(S_i)$)的位置集合作为最优位置集合 $S_{\max}$。

由于 $S_{\max}$ 中所有位置单元的查询概率 q 满足 $q_1 \approx q_2 \approx \cdots \approx q_{k-1} \approx q_t$，故该匿名位置集合能同时保证位置熵最大。

该算法通过选取最优虚假位置集合，实现匿名位置集合的缓存命中率和位置熵同时最大，达到最优的隐私保护效果。

5.3 方案设计

基于虚假位置选择算法，提出一种基于信息缓存的隐私保护方案，主要包括初始化、基于位置的服务查询、服务应答及缓存更新这 3 个部分。

5.3.1 初始化

本节主要对缓存列表中的兴趣点信息和查询概率进行初始化。

1) 兴趣点信息缓存列表初始化

令 DB_{rec} 代表兴趣点信息缓存列表，$DB_{rec}=\{rec_1, rec_2, \cdots, rec_n\}$，其中，$rec_i$ 代表 DB_{rec} 中的第 i 条记录，$i=1, 2, \cdots, n$。

记录 rec 在 DB_{rec} 中的存储格式为 $rec=\{id, name, loc, info, f, timestamp\}$，其中，$id$ 代表记录在缓存列表中的序号；$name$ 代表兴趣点名称；loc 代表兴趣点的地理坐标；$info$ 代表兴趣点的详细服务信息；f 表示记录的数据新鲜度；$timestamp$ 代表记录的存入时间。

2) 背景信息缓存列表初始化

根据地图兴趣点库可获取用户所在城市地图中所有兴趣点的查询概率，并将其存储在缓存列表 DB_{poi} 中，$DB_{poi}=\{poi_1, poi_2, \cdots, poi_m\}$，其中，$m$ 为兴趣点的个数，poi_j 表示 DB_{poi} 中的第 j 条记录，$j=1, 2, \cdots, m$，DB_{poi} 以时间 T_{poi} 为自动更新周期。

兴趣点 poi 在 DB_{poi} 中的存储格式为 $poi=\{id, name, loc, p, timestamp\}$，其中，$id$ 代表兴趣点在缓存列表中的序号；$name$ 代表兴趣点名称；loc 代表兴趣点的地理坐标；p 代表兴趣点的查询概率；$timestamp$ 代表兴趣点的存入时间。

5.3.2 基于位置的服务查询

为便于对基于位置的服务查询过程进行描述，假定通信链路安全。定义通用的基于位置的服务查询请求 $query$ 的格式为 $query=\{user, loc, content, timestamp\}$，其中，$user$ 代表用户身份标识；loc 代表用户所处位置的地理坐标；$content$ 代表查询内容；$timestamp$ 代表查询时间。应答结果 $result$ 的格式为 $result=\{user, loc, info, timestamp\}$，其中，$user$ 代表用户身份标识；loc 代表用户所处位置的地理坐标；$info$ 代表兴趣点的详细服务信息；$timestamp$ 代表服务器的查询时间。

用户在使用基于位置的服务时，主要进行用户与缓存列表间的查询操作和用户与服务提供商间的查询操作。

1) 首先在终端的 DB_{rec} 中进行查询，若 DB_{rec} 中存储的某条记录 rec_i 能够满足用户查询请求 $query$，则直接向用户返回记录 rec_i 中的 $info$ 部分作为应答结果，用户不用向不可信的服务提供商发送查询请求。

2) 若 DB_{rec} 中的任一记录 rec_j 均不满足用户查询请求 $query$，智能终端首先根据所在区域背景信息，从 DB_{poi} 中选取和用户真实位置 loc_t 查询概率相近的 $4k$ 个位置单元$\{loc_1, loc_2, \cdots, loc_{4k}\}$；再从 $4k$ 个位置单元中随机选取 $2k$ 个位置单元$\{loc_1, loc_2, \cdots, loc_{2k}\}$构成位置集合 Q；然后根据 5.2 节中虚假位置选择算法的步骤，选取包含用户真实位置 loc_t 的最优位置集合 $S_{\max}=\{loc_1, loc_2, \cdots, loc_{k-1}, loc_t\}$作为匿名位置集合，并向服务提供商连续发送 k 个服务查询请求$\{query_1, query_2, \cdots, query_k\}$，其中，任意服务查询请求 $query_i=\{user, loc_i, content_i, timestamp\}$。

5.3.3 服务应答及缓存更新

服务提供商根据服务查询请求返回 k 个相应的应答结果$\{result_1, result_2, \cdots, result_k\}$。

1) 应答结果提取

移动终端将应答结果与包含用户真实位置 loc_t 的服务查询请求进行匹配，选取应答结果中满足 $loc_i=loc_t$ 的 $result_i$ 作为服务应答，并提取 $result_i$ 中的 $info_i$ 作为应答结果返回给用户。

2) 缓存更新

智能终端将服务提供商返回的应答结果$\{result_1, result_2, \cdots, result_k\}$存入 DB_{rec} 中。

若任意 $result_i=\{user, loc_i, info_i, timestamp\}$与记录 $result_j=\{id, name, loc_j, info_j, f, timestamp\}$均满足 $loc_i \neq loc_j$，其中，$i=1, 2, \cdots, k$，$j=1, 2, \cdots, n$，则将该应答结果 $result_i$ 存入 DB_{rec} 中作为记录，并将其数据新鲜度 f 设置为 1。

若应答结果 $result_i=\{user, loc_i, info_i, timestamp\}$与记录 $result_j=\{id, name, loc_j, info_j, f, timestamp\}$满足 $loc_i=loc_j$，$info_i=info_j$，其中，$i \in (1, 2, \cdots, k)$，$j \in$

(1, 2, ⋯, n)，则更新记录 rec_j 中的数据新鲜度 f 和存入时间 *timestamp*。

若应答结果 $result_i$={*user*, loc_i, $info_i$, *timestamp*}与记录 $result_j$={*id*, *name*, loc_j, $info_j$, *f*, *timestamp*}满足 loc_i=loc_j，$info_i$≠$info_j$，其中，i∈(1, 2, ⋯, k)，j∈(1, 2, ⋯, n)，则更新记录 rec_j 中兴趣点的详细服务信息 $info_j$、数据新鲜度 f 和存入时间 *timestamp*。

本文方案利用缓存机制，减少用户与不可信服务提供商间的交互次数，并利用虚假位置选择算法，优化所选取的匿名位置集合，提高缓存命中率，实现对用户位置信息 k-匿名级别的隐私保护。

6 安全性分析

由于缓存列表存储在用户的移动终端中，故可假定由缓存机制提供位置服务的过程是安全的。本节主要针对不可信服务提供商，从概率分布攻击、同源攻击和位置相似性攻击这 3 种主要攻击方式进行安全性分析。

针对概率分布攻击，用户利用虚假位置选择算法，选取缓存命中率最高的一组位置集合 $S_{\max}$={loc_1, loc_2, ⋯,loc_{k-1}, loc_t}作为匿名位置集合，并向服务提供商连续发送 k 个服务查询请求{$query_1$, $query_2$, ⋯, $query_k$}。在这个过程中，不可信服务提供商可从查询请求中提取到匿名位置集合 $S_{\max}$，并根据背景信息从 $S_{\max}$ 中获得各位置单元的查询概率{q_1, q_2, ⋯, q_{k-1}, q_t}，由于 $q_1 \approx q_2 \approx \cdots \approx q_{k-1} \approx q_t$，故位置单元 loc_i 为用户真实位置的概率为 $\frac{1}{k}$，其中，$q_i \approx q_t$，loc_i∈$S_{\max}$，即不可信服务提供商根据 $S_{\max}$ 推断出用户真实位置的概率仍为 $\frac{1}{k}$，无法根据背景信息提高推断出用户真实位置的概率。

针对同源攻击，本文方案利用 k-匿名技术和缓存机制，将基于位置的 k 个应答结果{$result_1$, $result_2$, ⋯, $result_{k-1}$, $result_t$}存储在 DB_{rec} 中。当用户在同一位置 loc_t、时间间隔Δt 内连续发送 m 个位置查询请求{$query_1$, $query_2$, ⋯, $query_m$}，由于任意 $query_i$ 中的地理坐标 loc_i=loc_t，其中，i∈(1, ⋯, m)，且当 DB_{rec} 中数据的生存周期 T>Δt 时，在时间间隔Δt 内的任意查询请求 $query_i$ 均可由缓存 DB_{rec} 中的记录 $result_t$ 作为应答结果；而当 DB_{rec} 中数据的生存周期 T<Δt 时，由于考虑数据新鲜度因素，在选取最优位置集合 $S_{\max}$ 时，尽量选取数据新鲜度较低的位置单元及时更新 DB_{rec}，提高信息复用率。因此能够有效减少用户在不可信 LBS 服务器中的查询次数，降低隐私泄露风险。

针对位置相似性攻击，本文方案引入归一化距离，使所选取的位置集合 $S_{\max}$ 中任意虚假位置与用户真实位置 loc_t 间的距离不太接近，从而保证虚假位置集合所形成的匿名区域中包含多种语义信息，即使不可信服务提供商根据查询请求获取到匿名位置集合 $S_{\max}$ 的所有位置单元，也无法根据语义特征推断出用户的隐私行为。

因此，本文方案能够有效地抵御拥有背景信息的攻击者的同源攻击、概率分布攻击和位置相似性攻击。

7 性能与仿真

本节首先对所提出的虚假位置选择算法 ELPPC 进行了算法复杂度分析，然后对所提出的 ELPPC 方案从通信开销和缓存命中率 2 个方面与 enhanced-CaDSA[27]、Baseline[27]和 enhanced-DLS[21]方案进行对比分析。

7.1 算法复杂度分析

基于缓存的虚假位置选择算法需经过多次循环操作，在每一轮循环中，都需要计算备选位置集合中每个集合的缓存贡献度，并按缓存贡献度大小进行排序。假设位置备选集的大小为 *popsize*，则排序算法计算复杂度为 O(*popsize* lb*popsize*)，缓存贡献度的计算复杂度为 O(*popsize*)，而每一轮循环中的交叉操作和变异操作计算复杂度为 O(1)。

由于后续每轮的备选集都是在前一轮的基础上进行迭代，故除第一轮循环外，后续的循环轮数并不需要 O(*popsize* lb*popsize*) 的计算量，因此，只需计算更新部分的缓存贡献度，并对更新部分进行堆排序插入到现有备选集队列中。在经过 *maxiter* 次循环后，新产生的位置集合数量为 ω=(*maxiter*−1)(1−e)*popsize*。再加上初始备选集大小 *popsize*，故整个算法的计算复杂度为 $O((\omega + popsize)\mathrm{lb}(\omega + popsize))$。

由于 enhanced-CaDSA 方案在从位置集合 Q 中选取匿名位置集合时，采用遍历的方法选取出缓存贡献度较大的前 k−1 个位置单元，其计算复杂度为 $O(C_{2k}^{k-1})$。如表 1 所示，当 k=20，备选方案数量为 $C_{40}^{19} \approx 1.3\times10^{11}$ 时，设置 *maxiter*=100，e=0.2，*popsize*=50，即可选出最优的位置集合[28]，其计算

复杂度为 4.8×10^4。显然，与 enhanced-CaDSA 相比，ELPPC 极大地减少了计算量。

表 1　　算法复杂度对比

算法	计算复杂度（计算式）	计算复杂度（k=20）
enhanced-CaDSA	$O(C_{2k}^{k-1})$	1.3×10^{11}
ELPPC	$O((\omega+popsize)\mathrm{lb}(\omega+popsize))$	4.8×10^4

7.2　性能仿真与分析

本文设定仿真环境为：将 8 km×8 km 的地图区域等分为 50 m×50 m 的位置单元，初始查询概率可由谷歌地图获知。每分钟随机选取 10 个位置单元发起位置服务请求，每个位置单元被选中的概率与其查询概率成正比。定义 t 表示仿真时间，m 表示移动终端向应用服务器发送的查询次数。

7.2.1　通信开销

本节主要分析移动终端与应用服务器间的通信开销。由于通信开销主要由移动终端向应用服务器发送的数据分组大小和查询次数决定，假定查询请求数据分组大小固定，则通信开销主要由移动终端向应用服务器发送的查询次数 m 决定。通过仿真实验，图 2 和图 3 分别表示移动终端向应用服务器发送的查询次数 m 随匿名位置集合大小 k、仿真时间 t 的变化趋势，具体描述和分析如下。

在图 2 中，由于 ELPPC、enhanced-CaDSA、enhanced-DLS 和 Baseline 方案均采用 k-匿名技术，故应用服务器完成一次位置服务需要返回 k 次应答。k 越大，匿名位置集合所含位置单元也越多，故 m 也越大。而在图 3 中，随着 t 的增大，缓存列表中存储的应答结果越多，由缓存列表提供服务的概率越大，故 m 逐渐减小。

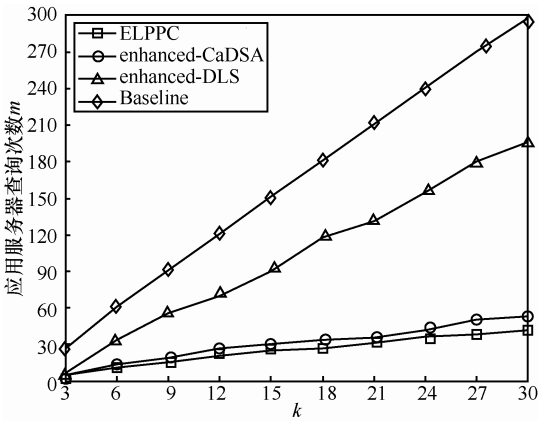

图 2　查询次数 m 随匿名位置集合大小 k 的变化趋势

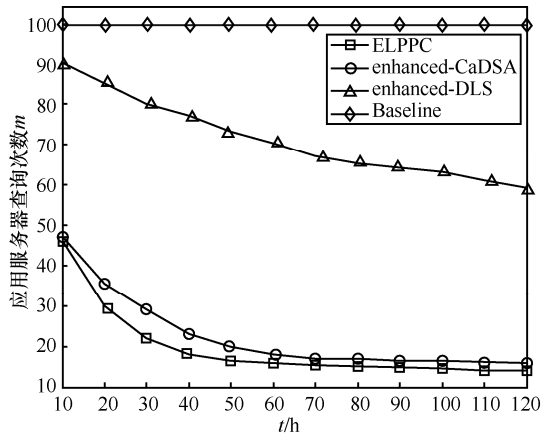

图 3　查询次数 m 随仿真时间 t 的变化趋势

由于 Baseline 方案中未采用任何缓存机制，移动终端向应用服务器发送的查询次数 m 随匿名位置集合大小 k 值的变化而呈线性增长，不随时间 t 的改变而变化；而在 enhanced-DLS、enhanced-CaDSA 和 ELPPC 方案中，通过缓存机制存储应用服务器返回的应答结果以提高信息复用率，可有效减少移动终端向应用服务器发送的查询次数。其中，enhanced-DLS 方案未引入缓存命中率因素，信息复用率较低；而与 enhanced-CaDSA 方案相比，由于 ELPPC 在选取匿名位置集合时进行了全局优化，缓存命中率更高，信息复用率更大，故在同等条件下 ELPPC 方案中的移动终端向应用服务器发送的查询次数会更少。

上述实验结果表明，与 Baseline、enhanced-DLS 和 enhanced-CaDSA 方案相比，ELPPC 方案能够有效减少移动终端向应用服务器发送的查询次数，降低通信开销。

7.2.2　缓存命中率分析

本节主要针对 ELPPC 方案中的缓存命中率进行分析。通过仿真实验，图 4 和图 5 分别表示缓存命中率随匿名位置集合大小 k 与仿真时间 t 的变化趋势。具体描述与分析如下。

在图 4 和图 5 中，由于 Baseline 方案未采用任何缓存机制，其缓存命中率始终为 0；在 ELPPC、enhanced-CaDSA 和 enhanced-DLS 方案中，随着 k 和 t 的增大，缓存列表中存储的应答结果数据逐渐增多，故缓存命中率逐渐增大；其中，enhanced-DLS 方案在匿名位置集合选取时未引入任何提高缓存命中率的因素，故缓存命中率始终低于 ELPPC 和 enhanced-CaDSA 方案；enhanced-CaDSA 方案在选取匿名位置集合时通过引入数据新鲜度和归一化

距离的因素，提高了其缓存命中率；而 ELPPC 方案在 enhanced-CaDSA 方案的基础上，结合遗传算法思想，从位置集合中选取出满足缓存命中率最大的一组位置集合作为最终的匿名位置集合，因此，在 4 个对比方案中拥有最高的缓存命中率。

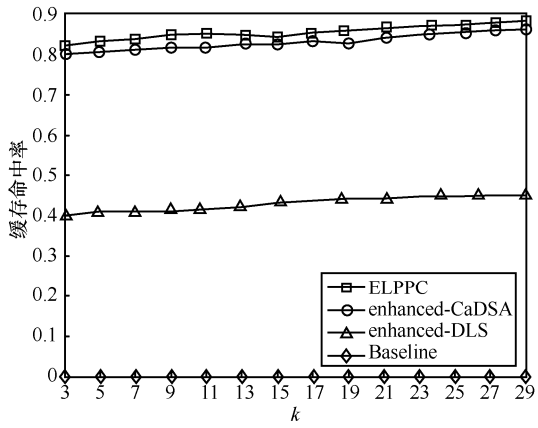

图 4　缓存命中率随匿名位置集合大小 k 的变化趋势

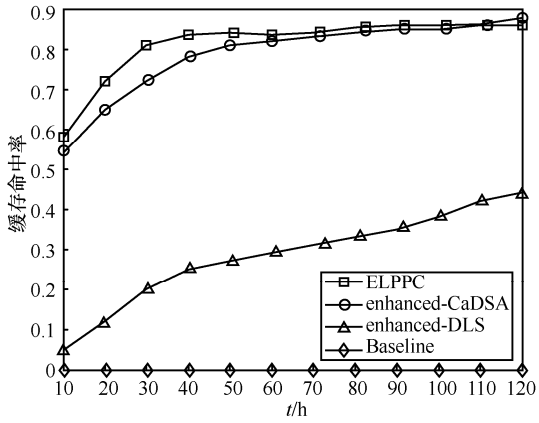

图 5　缓存命中率随仿真时间 t 的变化趋势

上述实验结果表明，与 Baseline、enhanced-DLS 和 enhanced-CaDSA 方案相比，ELPPC 方案能够有效提高缓存命中率，最大程度提升整个系统的隐私保护效果。

8　结束语

本文针对基于位置服务中移动终端用户的隐私保护问题，综合考虑用户所处区域背景信息的开放性特征，引入信息缓存机制，结合 k-匿名技术和遗传算法思想，提出一种虚假位置选择算法，并基于该算法设计了一种不依赖任何可信第三方的隐私保护方案，实现减少用户与不可信服务器间的交互次数和抵御拥有背景信息攻击者的多种推理攻击。安全性分析和实验结果进一步验证了所提出方案的有效性和高效性。下一步将着重分析不同场景、网络、终端等因素对缓存命中率的影响，并构造更高效的位置隐私保护方案。

参考文献：

[1] GRUSTER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]// International Conference on Mobile Systems. 2013:31-42.

[2] MOKBEL M F, CHOW C Y, AREF W G. The new casper: query processing for location services without compromising privacy[C]// The 32nd International Conference on Very Large Data Bases. Seoul, South Korea, 2006: 763-774.

[3] XU T, CAI Y. Feeling-based location privacy protection for location-based services[C]//ACM Conference on Computer & Communications Security(CCS). 2009:348-357.

[4] CHOW C, MOKEL M F, LIU X. A peer to peer spatial cloaking algorithm for anonymous location based service[C]//The 14th ACM International Symposium on Geographic Information Systems. 2006: 171-178.

[5] PAVEL S, FRANK D, KURT R. Map-aware position sharing for location privacy in non-trusted systems[C]//The 10th International Conference Pervasive Computing. Newcastle England, 2012: 388-405.

[6] 徐建, 黄孝喜, 郭鸣, 等. 动态 P2P 网络中基于匿名链的位置隐私保护[J]. 浙江大学学报(工学版), 2012, 46(4):712-718.
XU J, HUANG X X, GUO M, et al. Location privacy through an anonymous chain in dynamic P2P network[J]. Journal of Zhejiang University(Engineering Science), 2012, 46(4):712-718.

[7] CHENG R, ZHANG Y, BERTINO E, et al. Preserving user location privacy in mobile data management infrastructures[C]//Privacy Enhancing Technology Workshop (PET' 06). Cambridge, United Kingdom, 2006: 393-412.

[8] ANDERSEN M S, KJAERGAARD M B. Towards a new classification of location privacy methods in pervasive computing[J]. Social Informatics and Telecommunications Engineering, 2012, 104: 150-161.

[9] TRONG N P, TRAN K D. A novel trajectory privacy-preserving future time index structure in moving object databases[C]//The 3th International Conference on Computer and Computational Intelligence. 2012: 124-134.

[10] 魏志强, 康密军, 贾东宁, 等. 普适计算隐私保护策略研究[J]. 计算机学报, 2010, 33(1):128-138.
WEI Z Q, KANG M J, JIA D N, et al. Research on privacy protection policy for pervasive computing[J]. Chinese Journal of Computers, 2010, 33(1):128-138.

[11] YI X, PAULET R, BERTINO E, et al. Practical approximate k nearest neighbor queries with location and query privacy[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, (99):1-14.

[12] PAULET R, KAOSAR M G, YI X, et al. Privacy-preserving and content-protecting location based queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(5):1200-1210.

[13] GHINITA G, KALNIS P, KHOSHGOZARAN A, et al. Privacy queries in location based services: anonymizers are not necessary[C]//The 27th ACM Conference on Management of Data(SIGMOD'08). 2008: 121-132.

[14] HU H, LU R X, HUANG C, et al. An efficient privacy-preserving

location based services query scheme in outsourced cloud[J]. IEEE Transactions on Vehicular Technology, 2016, 65(9): 7729-7739.

[15] ZHU H, LIU F, LI H. Efficient and privacy-preserving polygons spatial query framework for location-based services[J]. IEEE Internet of Things Journal, 2015, (99): 1-1.

[16] 毛典辉, 曹健, 蔡强, 等. 情景感知的位置隐私保护方法研究进展[J]. 通信学报, 2013, 34(Z1): 230-234.
MAO D H, CAO J, CAI Q, et al. Survey of the context-aware location privacy-preserving techniques[J]. Journal on Communications, 2013, 34(Z1): 230-234.

[17] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//The 1st International Conference on Mobile Systems Applications and Services. USA, 2003: 31-42

[18] YIU M L, JENSEN C S, HUANG X, et al. SpaceTwist: managing the trade-offs among location privacy, query performance, and query accuracy in mobile services[C]//The IEEE International Conference on Data Engineering. 2008: 366-375.

[19] CHOW C Y, MOKBEL M F, LIU X. A peer-to-peer spatial cloaking algorithm for anonymous location-based services[C]//The 14th ACM International Symposium on Advances in Geographic Information Systems(ACM-GIS'06). 2006: 171-178.

[20] KIDO H, YANGISAWA Y, SATOH T. An anonymous communication technique using dummies for location-based services[C]//The 2nd IEEE International Conference on Pervasive Services(ICPS'05). 2005: 88-97.

[21] NIU B, LI Q H, ZHU X Y, et al. Achieving *k*-anonymity in privacy-aware location-based services[C]//The IEEE 33th International Conference on Computer Communications (INFOCOM'14). 2014: 754-762.

[22] 霍峥, 孟小峰, 黄毅. PrivateCheckIn: 一种移动社交网络中的轨迹隐私保护方法[J]. 计算机学报, 2013, 36(4): 716-726.
HUO Z, MENG X F, HUANG Y. PrivateCheckIn: trajectory privacy-preserving for check-in services in MSNS[J]. Chinese Journal of Computers, 2013, 36(4): 716-726.

[23] AMINI S, LINDQVIST J, HONG J, et al. Cache: caching location-enhanced content to improve user privacy[J]. ACM Sigmobile Computing & Communication Review, 2010, 14(3): 19-21.

[24] SHOKRI R, THEODORAKOPOULOS G, PAPADIMITRATOS P, et al. Hiding in the mobile crowd: location privacy through collaboration[J]. IEEE Transactions on Dependable and Secure Computing, 2013, 11(3): 266-279.

[25] ZHU X Y, CHI H T, NIU B, et al. Mobicache: when *k*-anonymity meets cache[C]//IEEE Global Communication Conference (GLOBECOM). 2013: 820-825.

[26] NIU B, ZHU X Y, LI W H, et al. EPcloak: an efficient and privacy-preserving spatial cloaking scheme for LBSs[C]//The 11th IEEE International Conference on Mobile Ad Hoc and Sensor Systems(MASS'14). 2014: 398-406.

[27] NIU B, LI Q H, ZHU X Y, et al. Enhancing privacy through caching in location-based services[C]//The IEEE 34th International Conference on Computer Communications (INFOCOM'15). 2015: 1017-1025.

[28] 西格兰·托比. 集体智慧编程[M]. 北京: 电子工业出版社, 2009: 89-91.
SEGRAN T. Programming collective intelligence[M]. Beijing: Publishing House of Electronics Industry, 2009: 89-91.

作者简介：

李璐璐（1990-），女，山西晋城人，西安电子科技大学硕士生，主要研究方向为信息安全、隐私保护。

华佳烽（1989-），男，湖北黄冈人，西安电子科技大学博士生，主要研究方向为信息安全、隐私保护。

万盛（1987-），男，江苏南通人，西安电子科技大学博士生，主要研究方向为网络安全与隐私保护。

朱辉（1981-），男，河南周口人，博士，西安电子科技大学教授、博士生导师，主要研究方向为数据安全及隐私保护、虚拟化技术与云计算安全、安全信息系统。

李凤华（1966-），男，湖北浠水人，博士，中国科学院信息工程研究所副总工程师、研究员、博士生导师，主要研究方向为网络与系统安全、隐私计算、信息保护。