doi:10.11959/j.issn.1000-436x.2017120

路网环境下基于伪随机置换的 LBS 隐私保护方法研究

周长利[1]，田晖[1]，马春光[2]，杨松涛[2]
（1. 华侨大学计算机科学与技术学院，福建 厦门 361021；
2. 哈尔滨工程大学计算机科学与技术学院，黑龙江 哈尔滨 150001）

摘 要：针对路网连续查询用户的位置隐私和查询内容隐私保护问题，提出一种基于伪随机置换的隐私保护方法。首先，基于路网顶点（锚点）组织兴趣点（PoI）分布信息，以单个路网顶点为基本处理对象，构造基于伪随机置换的 LBS 服务端兴趣点记录置换方案，该方案以 32 bit 随机种子生成置换表，并对兴趣点记录进行加密和置换处理后存入数据库；然后，可信中心服务器代理用户以目标类型兴趣点记录号发起查询，LBS 服务器无法确定用户真实位置及查询内容，实现了保护隐私的秘密检索；最后，对查询准确性、数据分组量和处理时间进行了对比分析实验，性能分析证明了所提方法具有位置不可追踪性和查询内容不可关联性。
关键词：基于位置的服务；隐私保护；*K* 近邻查询；不可追踪性；不可关联性
中图分类号：TP311 **文献标识码**：A

Research on LBS privacy preservation based on pseudorandom permutation in road network

ZHOU Chang-li[1], TIAN Hui[1], MA Chun-guang[2], YANG Song-tao[2]
(1. School of Computer Science and Technology, Huaqiao University, Xiamen 361021, China;
2. School of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)

Abstract: A method of privacy preservation based on pseudorandom permutation was put forward for the issues of location privacy and query content privacy. Firstly, the distribution information of points of interest (PoI) based on the vertexes in the road network was organized, each single road vertex was taken as the foundational processing object. Based on the pseudorandom permutation, a permutation scheme of the point-of-interest records at the LBS server's end was put forward, a 32-bit random seed was adopted to generate a permuted table in the scheme, and the point-of-interest records were encrypted and permuted according to the table. These processed records were stored in the LBS database. Then a trusted intermediate server, replacing of the user, issued a query request with a record number instead of the query content to the LBS server. The LBS server could not determine which kind of PoI the user was interested in or which road section the user was locating on, and therefore the scheme achieved private information retrieval. Finally, the efficiency in the metrics of query accuracy, communication overhead and processing time was also analyzed. By the performance analysis and extensive experiments, the proposed scheme is proved to be location untraceable and query content uncorrelation.
Key words: location-based service, privacy preservation, *K* nearest neighbor query, untraceable, uncorrelation

收稿日期：2016-11-24；**修回日期**：2017-03-27
基金项目：国家自然科学基金资助项目（No.61370007, No.61472097, No.U1405254, No.U1536115）；福建省高校新世纪优秀人才计划基金资助项目（No.2014FJ-NCET-ZR06, No.MJK2016-23）；福建省自然科学青年基金资助项目（No.2016J05158）；福建省高校杰出青年科研人才培育计划基金资助项目（No.MJK2015-54）；华侨大学科研基金资助项目（No.15BS412）
Foundation Items: The National Natural Science Foundation of China (No.61370007, No.61472097, No.U1405254, No.U1536115), Program for New Century Excellent Talents in Fujian Province University (No.2014FJ-NCET-ZR06, No.MJK2016-23), The Natural Science Foundation for Youths of Fujian Province (No.2016J05158), Program for Outstanding Youth Scientific and Technological Talents in Fujian Province University(No.MJK2015-54), Scientific Research Funds of Huaqiao University (No.15BS412)

1 引言

基于位置的服务（LBS, location based service）成为移动用户智能终端中最为广泛的应用形式[1,2]，LBS将用户位置数据作为基本输入条件，为用户提供高质量的智能服务，如导航、兴趣点查询和广告推荐等，实体架构如图1所示。然而，在获取LBS过程中用户直接提交位置及其查询内容等私密信息，这会给用户带来严重的隐私安全问题，攻击者可以通过用户提交的位置、查询内容并结合其自身掌握的背景知识，推断出用户的深入隐私信息，如家庭住址、日常生活轨迹、健康状况及爱好习惯等[2]。由于现有诸多移动终端软件在运行过程中都会读取用户的位置数据等隐私信息，因此，用户在获取这些位置服务时，其隐私信息必须得到有效保护[3]。

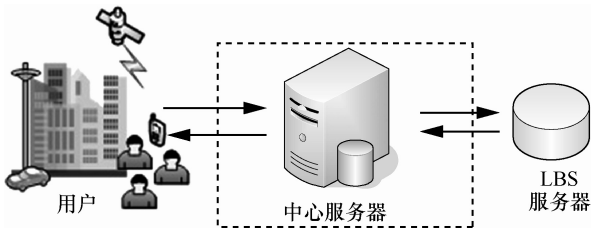


图1 LBS中的隐私保护架构

基于位置的兴趣点查询是LBS中最为典型的应用形式[4]，按照用户的查询形式可将LBS查询分为快照查询和连续查询，按照用户的查询范围可以分为范围查询和K近邻查询[5,6]。连续K近邻兴趣点查询是LBS中用户最常用的服务形式之一，在连续查询中移动用户根据自己变换的位置周期发起查询请求。一条LBS查询请求中包括多种隐私数据，其中，位置数据和查询内容直接反映用户的位置隐私及兴趣目标隐私，需要有效保护。典型的位置隐私保护方法可分为2类：构造匿名框（cloaking region）[7,8]和生成假位置（pseudo location）[9~12]。

匿名框以k个用户共享某个地理区域的形式泛化用户具体位置，用户以匿名框代替真实位置发起查询请求，以实现k匿名(k-anonymity)。但是这种方法需要LBS服务器具备处理匿名框查询的能力，同时，在移动用户发起连续查询时需要构造包括多数初始用户的连续匿名框[8]，然而参与用户移动方式的不确定性使构造连续匿名框成为难题，并且生成的连续匿名框容易相互关联，造成用户轨迹隐私的泄露。因此，匿名框方法不适用于LBS中的连续查询。

假位置方法是用户以一个或多个虚假的位置替代真实位置发起查询，具有构造灵活、易共享和易处理等优势。锚点技术是假位置技术的一种，能够有效实现位置隐私保护和精确的目标兴趣点查询。文献[10]方法（space twist）在最早提出时存在未实现k匿名、查询不均衡、锚点选用随机和基于欧氏空间设计等缺陷。针对未实现k匿名的问题，Gong等[11]提出了KAWCR，黄毅等[12]提出了Coprivacy等方法，这些方法多面向欧氏空间设计，无法直接应用于路网环境，并且锚点选取随机会给LBS服务器端带来频繁的数据库查询操作等问题。2015年，Ma等[13]针对锚点k匿名及查询不均衡等问题提出了HINN方法，针对锚点选取随机的问题提出了满足语义多样性的锚点选取方法[14]，提高了锚点选取的安全性。但是这些方法依然存在不适用于路网环境、锚点复用率低等缺陷。

另外，用户提交的查询内容直接反映了用户的查询兴趣，对查询内容隐私的保护也是LBS隐私保护的重要研究内容之一。一般地，用户要获取查询结果就必须提交查询目标，这必然会出现查询内容，隐私的泄露问题。为了既不泄露查询内容，又能获取查询结果，研究人员提出了私有信息检索（PIR, private information retrieval）[7,15~17]概念，但是一些PIR协议在查询过程中要依靠一个可信的安全处理器（SC，secure coprocessor）[16]对LBS数据库进行预处理，并且还存在查询计算开销巨大的问题[7,15]，因此，不适用于位置变换频繁的连续查询。2014年，Yang等[17]提出了一种基于加密置换的秘密检索隐私保护方法，能够有效提高查询效率，但是该方法面向欧氏空间设计，不适用于路网环境，且在查询时对兴趣点的处理方式不够高效，存在改进空间。

针对上述问题，本文研究内容及贡献如下。

1) 面向路网环境，针对LBS中连续查询用户的位置隐私和查询内容隐私的保护需求，提出了位置隐私保护模型及通过位置不可追踪性、查询内容不可关联性这2个指标来判断位置隐私和查询内容隐私保护程度，并分析了本文方法的安全性和效率。

2) 为实现上述位置隐私和查询内容隐私保护目标，提出了一种基于伪随机加密置换的LBS隐私保护方案。该方案以路网顶点（锚点）为基础组织兴趣点分布情况信息，同时，以锚点替代用户位置

发起查询保护连续查询中的位置隐私，降低数据库查询次数。以路网顶点为基本处理单元，通过加密及伪随机置换兴趣点记录的方式，实现了对路网目标兴趣点的秘密检索，LBS 服务器无法确定用户查询了哪类兴趣点，保护了用户的查询内容隐私。

2　路网 LBS 用户隐私保护模型

本文提出的用户隐私保护模型采用有中心服务器的 3 层实体架构，如图 2 所示。

定义 1　位置隐私保护模型。该模型可以用 3 元组 (E,L,M) 来描述，E 表示系统内实体集合，L 表示实体掌握的知识集合，M 表示评价方法及指标集合。

$E=\{U,CS,LBS,SC,Adv\}$，其中，U 为配备有移动智能终端的用户集合；CS 表示中心服务器集合，中心服务器代理用户发送查询信息，并为用户生成精确查询结果；LBS 表示位置服务器集合；SC 表示安全处理设备集合，负责对兴趣点原始数据记录进行加密置换；Adv 表示攻击者集合。

$L=\{A,S,P,B,\cdots\}$，实体知识集合由算法集 A、策略集 S、协议集 P 及背景知识集 B 等构成。

$M=\{Eva,Metr\}$，其中，Eva 和 $Metr$ 分别表示评价方法及指标集合。

图 2 描述了 LBS 用户隐私保护实体架构及查询流程。从可信性角度看，除自身外用户认为 CS 和 SC 是可信的，LBS 是半可信的，存在隐私泄露风险，或其本身就对用户隐私信息感兴趣。从计算能力和存储能力角度看，用户 U 计算能力较弱，而中心服务器 CS 计算存储能力较强，能为众多用户提供隐私保护服务和代理查询服务，LBS 具有超强的计算和存储能力，并具备统计分析能力。SC 集成了加密功能硬件设备，禁止外界访问其 RAM，其可发现对自身软、硬件的改动并及时响应。

定义 2　K 近邻查询请求。用户的查询请求可以用 $Q=<u_k,loc,time,K,C>$ 表示，其中，$u_k\in U$ 表示用户身份标识，为了不泄露真实身份及抵御身份标识关联，通常使用假名并在每次查询中更换，假名之间无关联；loc 表示用户位置；$time$ 表示查询时间戳；K 表示查询近邻兴趣点数量；C 表示查询内容，即用户感兴趣的目标兴趣点类型；其中，loc 和 C 是用户查询请求中直接涉及用户隐私的 2 个重要内容，本文重点研究对这 2 项内容的保护方法。

在用户提出查询请求之前，首先可信安全处理器 SC 调用加密算法 $Enc\in A$，对原始兴趣点记录进行加密、顺序置换，存入 LBS 数据库中，加密置换过程对 LBS 服务器保密，LBS 服务器不保留原始数据库；然后用户 u_k 发送查询请求 Q 给 CS，CS 调用算法计算出目标兴趣点在 LBS 数据库中的记录序号，进而利用该序号和路网锚点替换用户查询内容 C 和位置 loc，代理用户向 LBS 服务器发起查询；LBS 服务器检索加密数据库，将数据库中的某个兴趣点记录密文结果发送给 CS；最后 CS 调用解密算法 $Dec\in A$ 计算出明文结果 R，并将其发送给用户 u_k。

定义 3　位置 k 匿名。从攻击者角度来看，对于任意路段用户 $u_i\in U$ 都无法将其映射到本路段及邻近路段上 k 个用户位置中具体的某一个。

在本文方案中，多个路段上的用户共用同一个路网顶点作为查询锚点，攻击者无法确定究竟是哪个用户发起了该查询，实现单次位置 k 匿名。当用户在连续查询中均利用不同锚点发起查询时，与路网内其他用户共用连续锚点，实现连续位置匿名，用户连续位置之间相互独立。

定义 4　位置不可追踪性。位置不可追踪性是指攻击者依据用户查询请求中的位置信息，并结合掌握的背景知识推断出用户连续位置（轨迹）隐私的概率很低。

在连续查询中用户发起多次同类查询，因此，查询内容依然可以作为关联因素，推断出用户连续位置隐私，甚至以此推断出用户真实身份，因此，LBS 用户隐私保护还需要满足不可关联性。

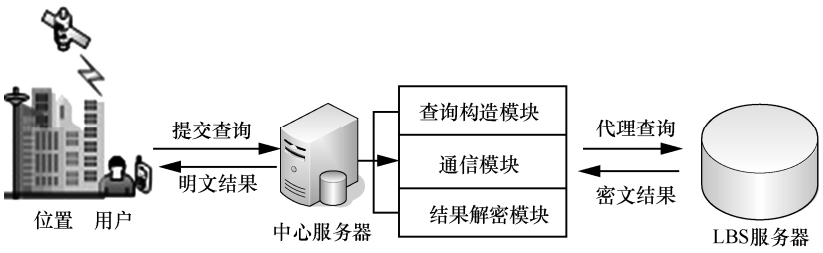


图 2　LBS 用户隐私保护实体架构及查询流程

定义 5 查询内容不可关联性。查询内容不可关联性是指连续查询中任意 2 个查询内容 C_m 与 C_n 之间相互关联的概率很低，即连续查询中单次查询之间相互独立。

位置和查询内容是用户获取 LBS 需要提交的 2 项重要隐私内容。位置不可追踪性和查询内容不可关联性是实现 LBS 隐私保护的重要指标，属于评价指标集合 *Metr* 成员。私有信息检索是实现不可追踪、不可关联的有效技术手段，私有信息检索是指 u_k 在数据库中秘密检索第 i 条记录，攻击者及 LBS 服务器均无法确定 u_k 检索了哪条记录，且连续相同内容检索之间无关联。

保护单次查询中的位置和查询内容，无法确保连续查询中用户隐私信息不泄露，因为攻击者可以通过追踪用户位置、关联查询内容来深入分析用户身份等其他隐私信息，因此，需要通过确保位置不可追踪性和查询内容的不可关联性，达到保护 LBS 连续查询中的用户隐私的目标。

定义 6 位置隐私。对于任意用户 u_k 发起的查询请求 $Q=<u_k',loc,time,K,C>$，如果存在方法 *func* 使式(1)成立，并且当 *loc* 为可用位置坐标时，则称 *loc* 为用户位置隐私。

$$func:(B_{Adv},<u_k',loc,time,K,C>,Asi)\rightarrow u_k \quad (1)$$

其中，B_{Adv} 为攻击者掌握的背景知识集，u_k 表示用户真实身份标识，查询中以假名 u_k' 替代，每次查询均使用无关联的不同假名，*Asi* 为辅助输入，辅助输入的添加取决于攻击者对方法 *func* 的构造能力，$func\in L$ 可以是攻击者自运行算法，也可以是与其他实体交互的协议或策略等。

可见，隐私泄露是指将某个可用私密数据映射到具体用户身份标识上。无法实现身份映射的私密数据对于攻击者来说没有实际价值，因此，隐私保护的基本思想是切断私密数据到用户身份标识的映射，由此可类推定义用户查询内容隐私。

3 基于伪随机置换的隐私保护方案

私有信息检索技术（PIR）是实现隐私保护的有效手段，但现有 PIR 实现方案存在计算开销大、不适用于路网环境和单次查询检索范围大等问题，针对上述问题，本文设计了路网环境下基于伪随机置换的 PIR 方案，该方案分为以下 2 个实施阶段：1) *SC* 离线初始化兴趣点记录数据库，以路网顶点为划分单位，对数据库记录加密，然后置换每条加密记录，并存储置换关系；2) *CS* 依据用户的查询请求，并结合兴趣点类型的查询记录号代理用户发起查询，并根据返回的结果计算出用户 *K* 近邻目标兴趣点发送给用户，实现对用户位置隐私和查询内容隐私的保护。

3.1 初始化阶段

为了适用于路网环境、提高查询效率和切断连续查询中内容关联的可能，*SC* 首先以各路网顶点为出发点构建兴趣点分布信息表，每个顶点对应的各类兴趣点按照与该顶点路网距离远近，分类升序排列相同兴趣点 $K_{\max}$ 近邻结果及对应描述信息，LBS 数据库中兴趣点存储信息如表 1 所示，表 1 中的 inf_i 表示对应的 loc_i 位置上某个兴趣点的详细描述信息，如某个宾馆的房间类型、价格等。

令 $\boldsymbol{DB}_{\text{ori}}=\{\boldsymbol{REC}_{v_1},\boldsymbol{REC}_{v_2},\cdots,\boldsymbol{REC}_{v_n}\}$ 表示原始数据库，其中，$\boldsymbol{REC}_{v_i}$ 表示以某个路网顶点 v_i 为起始的所有类型兴趣点 $K_{\max}$ 近邻记录集合，如表 1 所示。$\boldsymbol{REC}_{v_i}^{\text{enc}}$ 和 $\boldsymbol{REC}_{v_i}^{\text{idx}}$ 分别代表对 $\boldsymbol{REC}_{v_i}$ 的加密置换后的结果和数据库索引，$\boldsymbol{DB}_{\text{ori}}$ 中任意 $\boldsymbol{REC}_{v_i}$ 有

表 1 **LBS 数据库中兴趣点存储信息**

顶点记录号	路网顶点（锚点）	兴趣点记录号	兴趣点类型	$K_{\max}$ 近邻兴趣点集	兴趣点对应描述信息
$\boldsymbol{REC}_{v_1}$	v_1(锚点 1)	1	医院	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$	$\{inf_1,inf_2,\cdots,inf_{K_{\max}}\}$
		2	银行	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$	$\{inf_1,inf_2,\cdots,inf_{K_{\max}}\}$
		3	宾馆	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$	$\{inf_1,inf_2,\cdots,inf_{K_{\max}}\}$
		4	加油站	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$	$\{inf_1,inf_2,\cdots,inf_{K_{\max}}\}$
$\boldsymbol{REC}_{v_2}$	v_2(锚点 2)	1	医院	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$	$\{inf_1,inf_2,\cdots,inf_{K_{\max}}\}$
		2	银行	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$	$\{inf_1,inf_2,\cdots,inf_{K_{\max}}\}$
		3	宾馆	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$	$\{inf_1,inf_2,\cdots,inf_{K_{\max}}\}$
⋮	⋮	⋮	⋮	⋮	⋮

$\boldsymbol{REC}_{v_i}=\{\boldsymbol{rec}[1],\boldsymbol{rec}[2],\cdots,\boldsymbol{rec}[n]\}$，其中，$\boldsymbol{rec}[i]$ 表示 $\boldsymbol{REC}_{v_i}$ 中第 i 条记录，$1\leqslant i\leqslant n$；$\boldsymbol{REC}_{v_i}^{\text{enc}}=\{\boldsymbol{rec}'[7],\boldsymbol{rec}'[3],\cdots,\boldsymbol{rec}'[2]\}$，其中，$\boldsymbol{rec}'[i]$ 表示 $\boldsymbol{REC}_{v_i}$ 中某条加密后的记录，并且其位置经过置换，$1\leqslant i\leqslant n$；$\boldsymbol{REC}_{v_i}^{\text{idx}}=\{\boldsymbol{rec}''[1],\boldsymbol{rec}''[2],\cdots,\boldsymbol{rec}''[n]\}$，其中，$\boldsymbol{rec}''[i]$ 表示 $\boldsymbol{REC}_{v_i}$ 中第 i 条记录的索引，$1\leqslant i\leqslant n$。

这种以路网顶点为基本元的兴趣点分布信息组织方式不仅满足路网距离搜索，还能够有效控制数据库检索范围。为了降低用户数量多带来的服务瓶颈，*CS* 可分布式部署，某个 *CS* 负责管辖区域内路网用户查询请求代理，并保存区域内 K 近邻兴趣点分布索引，如表 2 所示，该索引由表 1 生成。

表 2　区域内 K 近邻兴趣点分布索引

路网顶点(锚点)	兴趣点标号	兴趣点类型	$K_{\max}$ 近邻兴趣点集
v_1 (锚点 1)	1	医院	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$
	2	银行	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$
	3	宾馆	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$
	4	加油站	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$
v_2 (锚点 2)	1	医院	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$
	2	银行	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$
	3	宾馆	$\{loc_1,loc_2,\cdots,loc_{K_{\max}}\}$
⋮	⋮	⋮	⋮

SC 调用加密算法 $Enc\in A$ 对 $\boldsymbol{REC}_{v_i}$ 中的每个 $\boldsymbol{rec}[i]$ 进行加密，并分别对 $\boldsymbol{REC}_{v_i}$ 中的 n 条记录 $\{\boldsymbol{rec}[1],\boldsymbol{rec}[2],\cdots,\boldsymbol{rec}[n]\}$ 顺序进行伪随机置换，且不同 $\boldsymbol{REC}_{v_i}\subseteq\boldsymbol{DB}_{\text{ori}}$ 的置换表不同。下面，以 $\boldsymbol{DB}_{\text{ori}}$ 中某个 $\boldsymbol{REC}_{v_i}$ 中的兴趣点记录置换过程为例进行说明。

设某个 $\boldsymbol{REC}_{v_i}$ 中共有 n 条不同类型兴趣点近邻记录 $\{\boldsymbol{rec}[1],\boldsymbol{rec}[2],\cdots,\boldsymbol{rec}[n]\}$，且每条记录是由二进制数表示的向量，长度为 l bit，按照 32 bit 长度将每条记录分为 m 个数据块，其中，$m=\left\lfloor\frac{l}{32}\right\rfloor+1$，即 $\boldsymbol{rec}[i]=(a_{i1},a_{i2},\cdots,a_{im})$，则 $\boldsymbol{REC}_{v_i}$ 和某次加密置换后的 $\boldsymbol{REC}_{v_i}^{\text{enc}}$ 可以分别表示为 $\{0,1\}^{n\times m}$ 矩阵。

$$\boldsymbol{REC}_{v_i}=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1m}\\a_{21}&a_{22}&\cdots&a_{2m}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\cdots&a_{nm}\end{bmatrix}$$

$$\boldsymbol{REC}_{v_i}^{\text{enc}}=\begin{bmatrix}a'_{71}&a'_{72}&\cdots&a'_{7m}\\a'_{31}&a'_{32}&\cdots&a'_{3m}\\\vdots&\vdots&\ddots&\vdots\\a'_{21}&a'_{22}&\cdots&a'_{2m}\end{bmatrix}$$

加密置换过程分为 3 步，如图 3 所示。1) *SC* 生成 32 bit 随机数，并以此为种子生成 n 个 32 bit 随机序列，存储在数组 $s[n]$ 中；2) *SC* 将 $\boldsymbol{REC}_{v_i}$ 中的 n 个记录与伪随机序列中对应的随机数按序做模 2 运算，得出的每条加密结果 $\boldsymbol{rec}'[i]$ 暂时存在寄存器中；3) *SC* 生成映射数组 $map_{v_i}[t]$，并按照数组重排寄存器中 n 条记录的顺序，生成伪随机置换的加密记录结果 $\boldsymbol{REC}_{v_i}^{\text{enc}}$，并将其存储在 LBS 数据库中。LBS 服务器并不掌握置换表，可信 *CS* 依据置

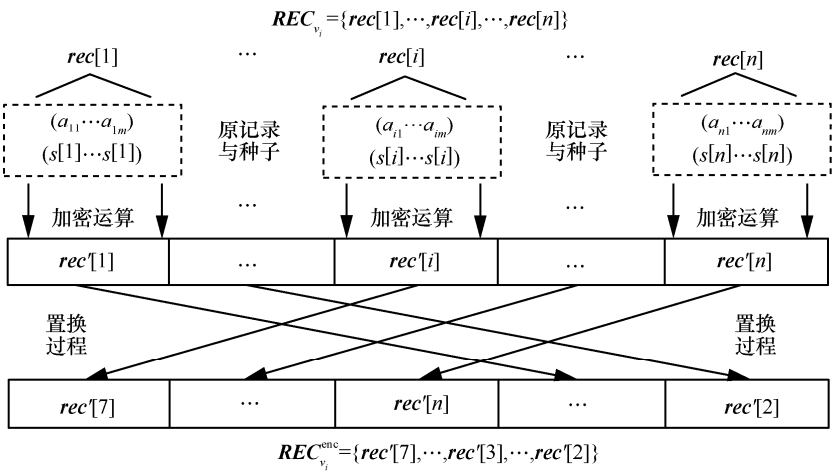


图 3　*SC* 端加密置换过程

换后记录号获取查询结果，因此，LBS 服务器无法确定 *CS* 查询了哪类兴趣点。加密置换算法如算法 1 所示。

算法 1 *SC* 端顶点 v_i 兴趣点记录 $\boldsymbol{REC}_{v_i}$ 加密置换算法

输入 以路网顶点 v_i 为出发点的近邻兴趣点记录 $\boldsymbol{REC}_{v_i}$

输出 加密并置换位置后的记录 $\boldsymbol{REC}_{v_i}^{\text{enc}}$

1) **Begin**

2) for $(t=1;t\leqslant n;t=t+1)$ do

3) 在 1~n 之间生成随机不重复整数按序存入 $map_{v_i}[t]$；

4) 生成 32 bit 随机种子存入数组 $s[0]$；

5) for $(i=1;i\leqslant n;i=i+1)$ do

6) 以 $s[0]$ 为种子生成随机 32 bit 随机序列存入 $s[i]$；

7) for $(j=1;j\leqslant m;j=j+1)$ do

8) 计算 $a_{ij}\oplus s[i]$ 并依次存入寄存器 $\boldsymbol{rec}'[i]$ 中；

9) 将 $\boldsymbol{rec}'[i]$ 按照随机置换映射数组 map_{v_i} 中的置换顺序得到 $\boldsymbol{REC}_{v_i}^{\text{enc}}$ 并重新存储到LBS数据库中；

10) return $\boldsymbol{REC}_{v_i}^{\text{enc}}$, map_{v_i}, s；

11) **End**

加密置换后的兴趣点记录结果存储在 LBS 端数据库中，*SC* 保存兴趣点记录置换关系表 map_{v_i} 及加密种子数组 s。通过加密和置换操作，攻击者及 LBS 服务器对存储的兴趣点记录内容及位置无法辨别。进而对于该 LBS 服务器负责区域内全部路网顶点的兴趣点记录均做同样的加密置换操作，并要根据 *SC* 周期性变换的置换表来定期更新加密数据库内容，进一步降低推断攻击的可能。

查询内容（目标兴趣点类型）以记录号替代，这样，LBS 服务器仅依据查询请求中的记录号返回密文记录，无法得知返回了哪个类型兴趣点的记录。由于不同路网顶点的兴趣点记录置换顺序是独立、随机的，因此，能够确保连续查询中内容无关联。

3.2 查询阶段

用户每次进入一个新路段后，将单次查询请求 $Q=<u_k',loc_{v_i},time,K,C>$ 发送给中心服务器 *CS*，其中，u_k'、loc_{v_i} 分别是用户假名和当前所在路段正方向顶点位置（锚点）。*CS* 据此向安全服务器 *SC* 获取兴趣点置换关系 map_{v_i}，根据用户查询内容 C 在索引表找到原记录顺序号，然后依据置换表 map_{v_i} 找到置换后的记录号 t，最后 *CS* 构建用户在本路段上的唯一一次代理查询请求 $Q_d=<CS_i,loc_{v_i},time,t>$，并发送给 LBS 服务器。其中，$CS_i$ 表示该代理中心服务器的标识符，该计算过程如算法 2 所示。

算法 2 *CS* 端生成代理查询 Q_d 过程

输入 用户查询请求 Q、兴趣点索引表、兴趣点置换表 map_{v_i}

输出 代理查询请求 Q_d

1) **Begin**

2) 提取 Q 中路网顶点 loc_{v_i} 及查询内容 C；

3) 依据 loc_{v_i} 在索引表中提取该顶点兴趣点索引表 $\boldsymbol{REC}_{v_i}^{\text{idx}}$；

4) for $\boldsymbol{REC}_{v_i}^{\text{idx}}$ 中每个记录 do

5) if 存在与查询内容 C 相同的兴趣点类型 then

6) 获取 $\boldsymbol{REC}_{v_i}^{\text{idx}}$ 中该类型兴趣点的记录号 x；

7) 向 *SC* 获取该路网顶点置换表 map_{v_i} 及加密种子 s；

8) 依据 map_{v_i} 将 x 置换成对应记录号 t；

9) else 返回查询出错；

10) 构造代理查询请求 $Q_d=<CS_i,loc_{v_i},time,t>$；

11) return Q_d；

12) **End**

CS 将代理查询请求 Q_d 发送给 LBS 服务器，LBS 检索数据库中路网顶点 v_i 的兴趣点记录集合中记录号为 t 的记录 $\boldsymbol{rec}'[t]$，并将其返回给 *CS*，*CS* 结合加密种子 s 对其内容解密，最后将 K 近邻查询结果返回给用户。解密过程如算法 3 所示。

算法 3 *CS* 端查询结果解密算法

输入 $\boldsymbol{rec}'[t]$，种子数组 s

输出 明文查询结果 $\boldsymbol{rec}[t]$

1) **Begin**

2) 在种子数组中找出第 t 个种子 $s[t]$；

3) 以 32 bit 为单位将 $\boldsymbol{rec}'[t]$ 分成 m 块；

4) for $(i=1;i\leqslant m;i=i+1)$ do

5) 将每块与种子 $s[t]$ 做异或运算，结果依次存入 $\boldsymbol{rec}[t]$；

6) return $\boldsymbol{rec}[t]$；

7) **End**

CS 可缓存部分置换表、加密种子集合以及曾经的 *K* 近邻查询结果，提高查询效率。

4　方案分析及实验

本节在分析本文所提方法的基础上，选取经典隐私保护方法与本文方案相比较，从安全性和工作效率 2 个主要方面对本文方法进行分析评价。实验在 Windows 7 平台上利用 Java 语言实现，数据地图采用美国地名委员会提供的地理数据集，并采用路网移动节点数据生成器 Thomas Brinkhoff 生成的模拟数据集，网络通信带宽为 3 Mbit/s，每次查询返回单个数据分组为 1 KB，每个兴趣点描述信息设置为 300 B，除去 40 B 分组首部，包含兴趣点个数为 $\frac{1\,024-40}{300}\approx 3$ 个。主要参数配置如表 3 所示。

表 3　　实验部分默认参数配置

参数名	取值范围	默认值
用户数量 U	[50 000,300 000]	100 000
用户查询兴趣点数 K	[10,60]	20
用户位置更新频率 f/s	[5,30]	15
兴趣点缓存时长 T_k /h	—	6
固定锚点使用率 r_c	[20%,100%]	90%
平均路段长度 S/m	[200,2 000]	1 000

4.1　安全分析及实验

在位置匿名和兴趣点查询过程中，确保位置不可追踪性和查询内容的不可关联性是防止攻击者推断用户深入隐私信息的重要指标，2 个评价指标的确保需要有以下命题成立。

命题 1　对于相关用户集 U_n 中某个用户的任意单次查询 Q_i，攻击者无法通过 Q_i 中的任意属性以超过 $\frac{1}{n}$ 的概率将其对应在该用户身份标识上，即有信息熵 $H(Q_1)=H(Q_2)=\cdots=H(Q_n)=\mathrm{lb}n$ 成立，此时，单次查询信息熵最大，称为单次查询不可识别。

命题 2　对于任意用户 u_k 的某次连续查询 $Q=(Q_1,Q_2,\cdots,Q_w)$ 中的任意 2 个查询是相互独立的，即有 $p(Q_1,Q_2,\cdots,Q_w)=p(Q_1)p(Q_2)\cdots p(Q_w)$ 成立，其中，$p(Q_i)$ 表示攻击者通过查询 Q_i 识别出用户的概率，称为连续查询相互独立。

定理 1　对于任意 LBS 连续查询用户 u_k，其隐私安全性在任意单次查询 $Q_i\in Q$ 无法识别(命题 1 成立)且连续查询 $Q=(Q_1,Q_2,\cdots,Q_w)$ 中任意 2 次查询相互独立(命题 2 成立)时实现。

证明　由数学归纳法可知，当用户发起查询，即 $z=1$ 时，*CS* 代理用户发起单次查询 $Q_i=<CS_i,loc_{v_i},time,t_i>$，在查询中以用户所在路段正方向顶点 loc_{v_i} 作为查询位置，该顶点为多个路段共用顶点，当多个路段内其他用户均采用其作为锚点查询时，攻击者无法通过锚点位置识别出该用户；同时，查询内容采用加密置换后的记录号 t_i 表示，攻击者无法通过查询内容识别出用户，实现了对目标类型兴趣点的秘密检索，有 $H(Q_1)=H(Q_2)=\cdots=H(Q_n)=\mathrm{lb}n$ 成立，由此可见，上述单次查询请求中的 2 个涉及用户位置、查询内容隐私的属性无明显识别特性，命题 1 得证。

当 $z=2$ 时，由于全局用户均采用锚点查询，所有锚点被全局用户共用，锚点并非用户为某次查询而构造，无用户从属性，因此，相邻时间域内任意 2 次查询请求位置 loc_{v_i} 和 loc_{v_j} 无关联；由于本文方法采用路网顶点(锚点)组织兴趣点分布信息，任意 2 个顶点的加密和置换过程均不同，使任意 2 个顶点相同查询内容的记录号 t_i 和 t_j 之间无关联。任意 2 次查询中涉及用户位置、查询内容隐私的 2 个属性无法相互关联，即有 $p(Q_1,Q_2)=p(Q_1)p(Q_2)$ 成立。

由此可推知，当 $z=w$ 时，有 $p(Q_1,Q_2,\cdots,Q_w)=p(Q_1)p(Q_2)\cdots p(Q_w)$ 成立。命题 2 得证。

当 $z=w+1$ 时，查询请求 Q_w、Q_{w+1} 相互独立，有 $p(Q_w)=p(Q_{w+1})$ 成立，使 $p(Q_1,Q_2,\cdots,Q_w,Q_{w+1})=p(Q_1)p(Q_2)\cdots p(Q_w)p(Q_{w+1})$ 成立，由此可得此次连续查询的联合信息熵最大，即

$$H(Q_1,Q_2,\cdots,Q_w)=\sum_{i=1}^{w}H(Q_i)=w\mathrm{lb}n \qquad (2)$$

由单次信息熵、连续信息熵最大可知，单次查询和连续查询的不确定性最高，即攻击者对单次查询不可识别、连续查询不可关联，定理 1 得证。

因此，由定理 1 可得，当连续查询请求相互独立时，用户的连续位置间无相互关联。相同查询内容在连续查询中以不同序号出现，查询内容相互独立无关联。因此，本文方法实现了用户位置的不可追踪性和查询内容的不可关联性。由于单次查询不可识别、释放的信息量少，连续查询的关联度低，因此，攻击者很难根据式(1)构造出合适方法 *func* 关联相关信息，推断出用户的真实身份信息。

同时，本文方法与另外 2 种经典方法匿名框方法[7]和文献[10]方法比较。在连续查询过程中假设路网顶点入边路段均为 4 条，路段上用户均以该顶点为锚点发起查询，在概率均等的条件下，处在顶点 v_n 的某条入边上的某个用户发起某次查询 Q_i，攻击者通过其查询内容等背景知识确定该用户所在路段的平均信息量为

$$H(Q_i)=-\sum_{i=1}^{4}p_i\mathrm{lb}p_i=-4\times\left(\frac{1}{4}\times\mathrm{lb}\frac{1}{4}\right)=2\ \mathrm{bit} \quad (3)$$

假设用户每次查询有效期内经过的路网顶点数量不超过 5 个，平均信息量累计至少 10 bit，以此时的平均信息量值为界限，如果用户在一次查询有效期内连续查询信息量低于 10 bit，则可认为此次查询存在隐私泄露的可能。高于 10 bit 查询占全部查询请求比例如图 4 和图 5 所示。

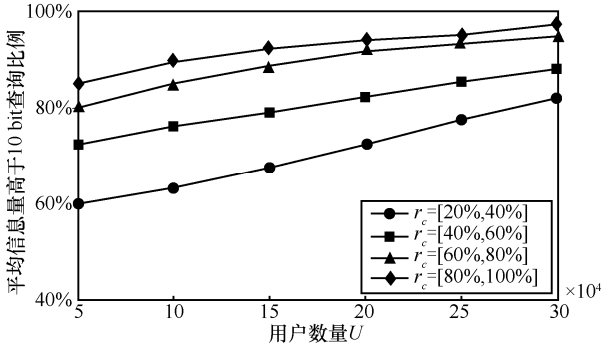


图 4 平均信息量大于 10 bit 查询比例分析

如图 4 所示，随着用户数量的增加，用户的匿名性增强，因此，平均信息量高于 10 bit 的查询比例逐渐增高，而在固定锚点共用比率高的实验中能够有更多的用户共用固定路网锚点查询，进一步增强了攻击者的不确定性，使隐私保护效果提高。

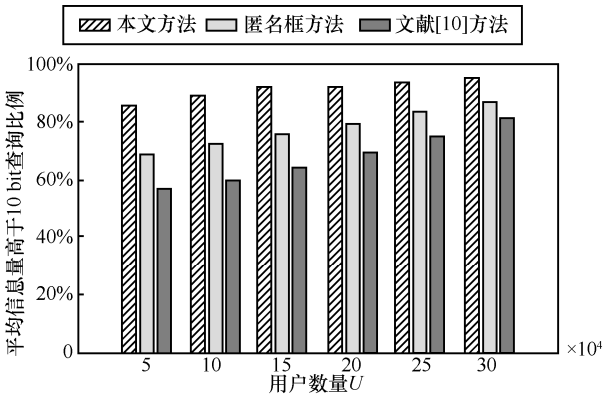


图 5 3 种方法平均信息量大于 10 bit 查询比例对比

如图 5 所示，与其他方法相比，本文方法能够确保多数查询具有良好的隐私保护性能，而匿名框方法虽然采用了 Paillier 密码系统处理查询请求，实现了秘密检索，但是由于该方法在欧氏空间下设计，实际路网中完全可以根据其构造的匿名框和路段的重合情况、距离远近以较高概率确定用户位置，因此，其实际应用的安全性能并不理想，而文献[10]方法锚点围绕用户随机选取，用户之间几乎不共用锚点，在路网环境中很容易根据锚点与路段距离推断出用户实际位置。

4.2 效率分析及实验

1) 查询准确率

首先，在 LBS 数据库预置存储以路网顶点为出发点的 $K_{\max}$ 近邻查询密文结果，然后，用户发起代理查询获取 $K(K\leqslant K_{\max})$ 近邻目标兴趣点结果集，在收集到的查询结果中采用大比例抽样、分组统计的方法，对比查询结果与实际情况的符合程度，以实际路网距离为衡量标准，核实查询结果中的 K 个兴趣点是否是距离用户路网最近的，考虑到用户实际移动会带来结果集的实时变化，影响统计分析，因此，在连续查询中除首次查询外，其他单次查询均以用户进入新路段为起始位置，最后，整理的结果显示，相比匿名框、随机锚点等欧氏空间查询方法，本文方法的查询结果更符合实际情况。

如图 6 所示，本文查询方法的查询准确率随用户数量增长迅速收敛，随用户数量增大而略有降低，这是由于用户数量增大给 LBS 服务器端带来了较重的查询负担，一些查询请求被丢弃，造成准确率下降，当 $K\leqslant 50$ 时，总体查询准确率可稳定在 80%以上。同理，查询兴趣点数 K 增大时，LBS 服务器的处理负担相应增大，查询准确呈下降趋势。

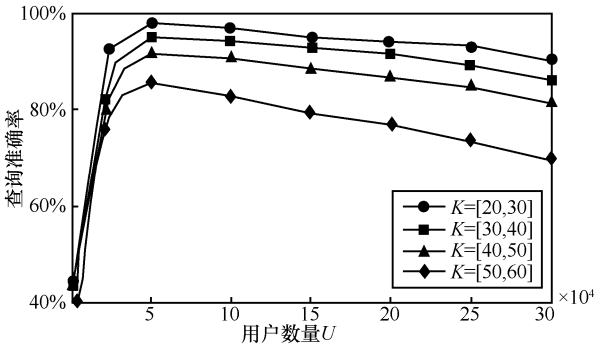


图 6 本文方法查询准确率

如图 7 所示，通过与其他 2 种隐私保护方法对比分析发现，随着用户数量的增长，本文方法的查询准确率相对稳定，这是由于本文方法采用了随机置换的方法，只要 *CS* 准确获得置换关系，即能够以较高概率计算出准确的查询结果。

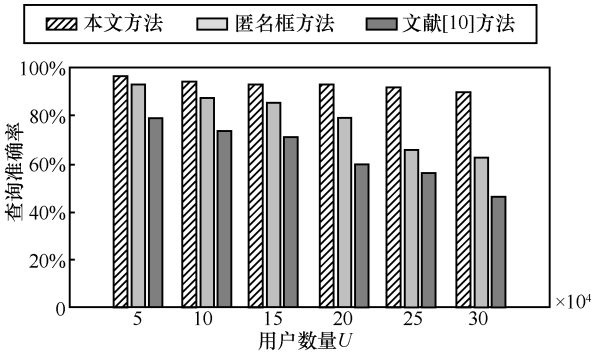


图 7　3 种方法查询准确率比较

匿名框方法在用户数量增长到一定程度时准确率下降较快，这是因为该方法采用同态加密处理查询请求，LBS 服务端处理此类查询请求比较耗时，当用户数量显著增大时，较多查询由于 LBS 服务端无法处理而被丢弃，因此，查询准确率下降。而文献[10]方法由于采用欧氏距离查询，查询结果准确率较低。

2) 平均数据分组量

用户出于保护自身位置需求用路网锚点提出查询时，由于兴趣点分布信息也是基于路网顶点组织的，因此，这种方法便于用户查询 *K* 近邻兴趣点。当路网内多数用户采用路网顶点作为锚点发起查询时，代理用户发起查询的 *CS* 可以缓存部分查询结果，相比使用随机假位置查询，固定锚点查询可以被不同用户反复使用，因此，在查询效率方面具有显著优势，降低 *CS* 向 LBS 数据库发起查询的比率，从而降低平均数据分组量和处理时间。

如图 8 所示，在用户数量增加的过程中，随着使用固定锚点用户比例 r_c 的增加，会有更多相同锚点的近邻查询结果被缓存下来。相比锚点使用率较少的情况，数据分组量明显较少，并且随着用户数量增长，整体数据分组量增长并不明显。

同时，相比另外 2 种方法，如图 9 所示，匿名框方法由于处理过程复杂，单次需要融合更多的处理数据，特别是在用户量较大时，LBS 服务端处理速度下降，因此丢弃部分查询请求，造成重复请求量增加，进而带来额外的数据分组量。此时文献[10]方法的数据分组量相对稳定。

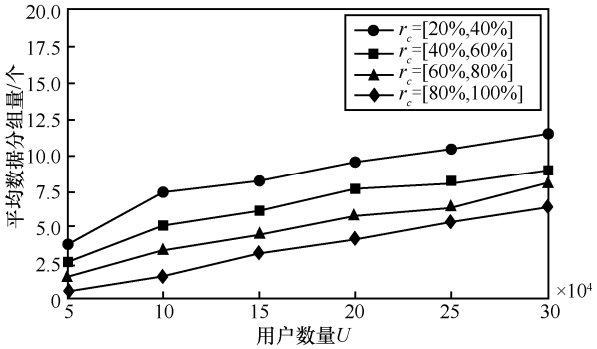


图 8　平均数据分组量随 *U* 和 r_c 变化

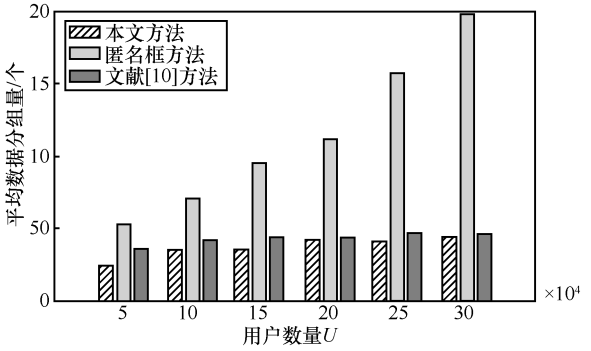


图 9　3 种方法平均数据分组量比较(*U* 变化)

当用户查询兴趣点数量 *K* 增加时，如图 10 所示，本文方法的平均数据分组量会有一定增长，但并不显著，这是由于固定锚点查询结果缓存及数据库端预置了各类兴趣点 $K_{\max}$ 近邻结果的原因。而另外 2 种方法每次查询都要直接向 LBS 数据库获取更多的查询结果，因此，平均数据分组量会有明显增长。

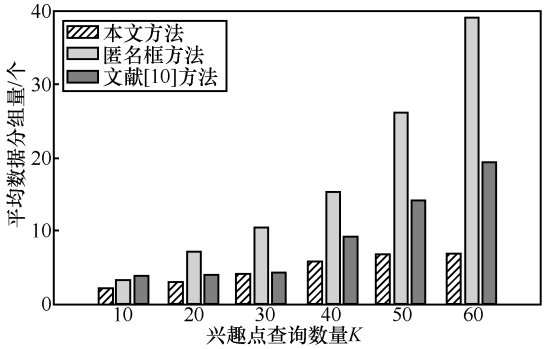


图 10　3 种方法平均数据分组量比较(*K* 变化)

当用户位置变化频率不同时，同样会影响数据分组量，如图 11 所示，当用户位置变化较为频繁

时，匿名框和文献[10]方法中用户会不断更新位置数据发起新的查询，因此，数据分组量随之升高，但是由于本文方法采用固定路网锚点发起查询，用户仅在变换路段时，以新路段正方向顶点发起一次查询，因此，不会因为用户位置频繁更新不断发起多次查询，带来分组量的增长。

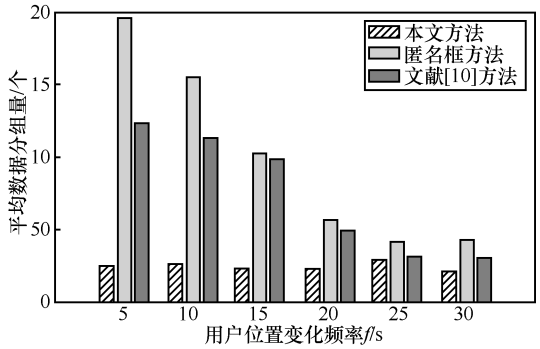


图 11　3 种方法平均数据分组量比较(f变化)

综上，本文方法在不同因素影响下，具有相对较少的数据分组量，同时查询处理速度相对较快。

3) 平均处理时间

本文方法采用路段顶点发起查询，用户仅在进入新路段时会发起一次查询，减少由位置频繁更新带来的查询处理数量，因此，平均路段长度会影响查询处理速度。如图 12 所示，在用户数量增长的过程中，平均路段长度 S 增大会使查询次数变少，在一定程度上降低 LBS 端分组量和平均处理时间。因此，可以采用剪枝的方法去掉一些较短路段，提高查询效率，但这种处理方法会带来查询准确率的下降。

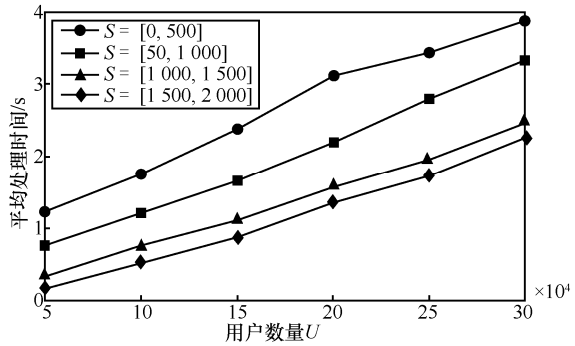


图 12　平均处理时间随 U 和 S 变化

同时，本文方法相对私有信息检索方法匿名框方法具有显著降低处理时间的优势，如图 13 所示，本文方法虽然也采用私有信息检索机制，但是实现方法采用的是随机加密置换技术，因此，用户端和 LBS 服务端均不需要复杂的处理运算。而匿名框方法由于采用的是同态加密处理查询请求，因此，当用户数量增大时其查询处理时间显著增加。

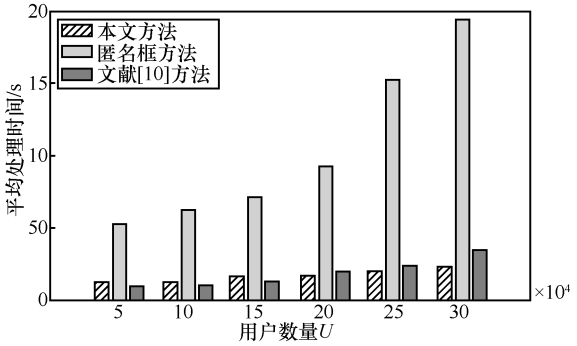


图 13　3 种方法平均处理时间比较

因此，本文方法相比同态加密实现的私有信息检索具有更为明显的处理速度优势，其平均处理时间能够达到与一般查询处理方法（文献[10]方法）同等的处理速度。

5　结束语

本文针对路网 LBS 用户兴趣点连续查询中的隐私保护问题，提出一种保护用户位置隐私和查询内容隐私的方法，该方法依据路网兴趣点组织特性，通过伪随机置换的方法实现了对用户 K 近邻兴趣点的私有信息检索，攻击者及 LBS 服务器无法知晓用户的真实位置及其查询内容（目标兴趣点类型）。通过理论分析及证明可知，本文方法能够确保攻击者无法通过单次查询识别出用户真实身份，也无法关联多个单次查询，实现了用户位置不可追踪、查询内容不可关联的 2 个目标。仿真实验表明，本文方法相比其他私有信息检索实现方法，具有较高的查询准确率、较少的数据分组量和处理时间，能够在保护位置隐私的同时具有较好的查询处理效率。

参考文献：

[1] GHINITA G. Privacy for location-based services[J]. Synthesis Lectures on Information Security, Privacy, & Trust, 2013, 4(1): 1-85.

[2] 张学军, 桂小林, 伍忠东. 位置服务隐私保护研究综述[J]. 软件学报, 2015, 9: 2373-2395.
ZHANG X J, GUI X L, WU Z D. Privacy preservation for location-based services: a survey[J]. Journal of Software, 2015, 9: 2373-2395.

[3] 王宇航, 张宏莉, 余翔湛. 移动互联网中的位置隐私保护研究[J]. 通信学报,2015,36(9):230-243.

WANG Y H, ZHANG H L, YU X Z. Research on location privacy in mobile internet[J]. Journal on Communications, 2015, 36(9): 230-243.

[4] 潘晓，郝兴，孟小峰. 基于位置服务中的连续查询隐私保护研究[J]. 计算机研究与发展, 2010, 47(1): 121-129.
PAN X, HAO X, MENG X F. Privacy preserving towards continuous query in location-based services[J]. Journal of Computer Research and Development, 2010, 47(1): 121-129.

[5] 高胜, 马建峰, 姚青松, 等. LBS 中面向协同位置隐私保护的群组最近邻查询[J]. 通信学报,2015,36(3):146-154.
GAO S, MA J F, YAO Q S, et al. Towards cooperation location privacy-preserving group nearest neighbor queries in LBS[J]. Journal on Communications, 2015,36(3):146-154.

[6] NI W, GU M, CHEN X. Location privacy-preserving k nearest neighbor query under user's preference[J]. Knowledge-Based Systems, 2016, 103: 19-27.

[7] 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, 28(6): 1546-1559.

[8] PAN X, XU J, MENG X. Protecting location privacy against location-dependent attacks in mobile services[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(8): 1506-1519.

[9] NIU B, ZHANG Z, LI X, et al. Privacy-area aware dummy generation algorithms for location-based services[C]// 2014 IEEE International Conference on Communications (ICC). 2014: 957-962.

[10] MAN L Y, JENSEN C. S, HUANG X G, et al. SpaceTwist: managing the trade-offs among location privacy, query performance, and query accuracy in mobile services[C]//IEEE 24th International Conference on Data Engineering.2008:366-375.

[11] GONG Z, SUN G Z, XIE X. Protecting privacy in location-based services using *k*-anonymity without cloaked region[C]//Mobile Data Management (MDM), 2010 Eleventh International Conference. 2010: 366-371.

[12] 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.
HANG 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.

[13] 马春光, 周长利, 杨松涛, 等. 基于 Voronoi 图预划分的 LBS 位置隐私保护方法[J]. 通信学报,2015,36(5):5-16.
MA C G, ZHOU C L, YANG S T, et al. Location privacy-preserving method in LBS based on Voronoi division[J]. Journal on Communications, 2015,36(5):5-16.

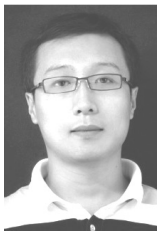
[14] 周长利, 马春光, 杨松涛. 基于敏感位置多样性的 LBS 位置隐私保护方法研究[J]. 通信学报, 2015, 36(4): 129-140.
ZHOU C L, MA C G, YANG S T. Research of LBS privacy preserving based on sensitive location diversity[J]. Journal on Communications, 2015, 36(4):129-140.

[15] YI X, PAULET R, BERTINO E, et al. Practical *k* nearest neighbor queries with location privacy[C]//2014 IEEE 30th International Conference on Data Engineering. IEEE, 2014: 640-651.

[16] MOURATIDIS K, YIU M L. Shortest path computation with no information leakage[J]. The VLDB Endowment, 2012, 5(8): 692-703.

[17] 杨松涛, 马春光, 周长利. 面向 LBS 的隐私保护模型及方案[J]. 通信学报, 2014, 35(8): 116-124.
YANG S T, MA C G, ZHOU C L. LBS-oriented location privacy protection model and scheme[J]. Journal on Communications, 2014, 35(8): 116-124.

作者简介：



周长利（1985-），男，黑龙江哈尔滨人，博士，华侨大学讲师，主要研究方向为位置隐私保护、网络与信息安全等。

田晖（1982-），男，湖北赤壁人，华侨大学副教授，主要研究方向为网络信息安全、大数据安全与隐私保护、多媒体内容安全等。

马春光（1974-），男，黑龙江双鸭山人，哈尔滨工程大学教授、博士生导师，主要研究方向为密码学、网络与信息安全等。

杨松涛（1972-），男，黑龙江佳木斯人，哈尔滨工程大学副教授，主要研究方向为网络与信息安全、位置隐私保护等。