个性化搜索中一种基于位置服务的隐私保护方法

张 强[①] 王国军*[②]

[①](中南大学信息科学与工程学院 长沙 410083)

[②](广州大学计算机科学与教育软件学院 广州 510006)

摘 要：在基于位置服务的个性化搜索中，利用可信第三方服务器以及对等节点是保护用户隐私的主要方法，但在现实生活中，它们却是不完全可信的。为了解决这一问题，该文提出一种个性化搜索中基于位置服务的隐私保护方法。该方法通过转换用户的位置信息，并根据用户的查询类型生成用户模型，进而形成带有用户位置信息的查询矩阵，然后利用矩阵加密用户的查询，隐藏查询矩阵中的用户信息，最后根据安全内积计算返回相关性得分最高的前K个查询文件给用户。安全性分析表明该方法能有效地保护用户的查询隐私和位置隐私，通过分析与实验表明，该方法大幅度地缩短了索引构建时间，降低了通信开销，同时为用户提供了基于位置的个性化搜索结果，一定程度上解决了移动设备屏幕小带来的弊端。

关键词：隐私保护；个性化搜索；位置转换；安全内积计算

中图分类号：TP393 **文献标识码**：A **文章编号**：1009-5896(2018)08-1998-08

DOI: 10.11999/JEIT171137

Privacy Preserving Method Based on Location Service in Personalized Search

ZHANG Qiang[①] WANG Guojun[②]

[①](*School of Information Science and Engineering*, *Central South University*, *Changsha* 410083, *China*)

[②](*School of Computer Science and Educational Software*, *Guangzhou University*, *Guangzhou* 510006, *China*)

Abstract: For personalized search based on location service, the trusted third-party server and peer node are used as the main method for privacy preserving. However, entirely trusted third-party server or peer node does not exist in real life. In order to address this problem, a method of privacy preserving on the location of mobile users is proposed when using personalized search. The method is used to convert the user's location information into distance information and generate the user model according to the user's query type, forming a query matrix with user location information, then the matrix is used to encrypt the user's query and conceal the user information in the query matrix. Finally, according to the calculation of the security inner product, the K file with the highest relevance score is returned to the user. It is evident from the security analysis that the proposed method can effectively protect the user's query privacy and location privacy. The analysis and experimental results show that the proposed method can greatly shorten the time of index construction and reduce the communication overhead. While providing users with location based personalized search results, the method is able to remedy the defects of small-screen mobile devices.

Key words: Privacy preserving; Personalized search; Location conversion; Security inner product

收稿日期：2017-12-04；改回日期：2018-04-20；网络出版：2018-06-07

*通信作者： 王国军 csgjwang@gmail.com

基金项目：国家自然科学基金(61632009, 61472451)，广东省自然科学基金(2017A030308006)，广东省高等教育高层次人才计划(2016ZJ01)，中南大学中央高校基本科研业务费专项资金 (2017zzts141)

Foundation Items: The National Natural Science Foundation of China (61632009, 61472451), The Guangdong Provincial Natural Science Foundation (2017A030308006), The High-Level Talents Program of Higher Education in Guangdong Province (2016ZJ01), The Fundamental Research Funds for the Central Universities of Central South University (2017zzts141)

1　引言

近年来，随着无线通信技术、移动智能终端以及定位技术的快速发展，使得基于位置的服务迅速发展并得到了广泛的应用[1,2]。用户可以通过基于位置的服务(Location Based Service, LBS)获得附近的兴趣点[3]，然而对于用户的决策而言，仅仅知道兴趣点的位置信息是远远不够的，比如，当用户在寻找附近的餐馆时，用户并不能完全根据周边餐馆的距离选择最适合自己的餐馆。正因于此，研究者们提出了各种各样的用户模型构建方法[4,5]，以期全方位的描述用户的兴趣、行为等特征，进而使用户获得更精准的个性化搜索结果。同时，在获得LBS服务的同时，用户隐私信息泄露的风险也不容小觑，比如攻击者可以根据位置信息，获得用户单位以及家庭的具体方位等敏感信息[6]。因此，研究者们提出通过加入可信第三方服务器(Trusted Third Party, TTP)对用户的隐私进行保护[7,8]，虽然这种方法解决了位置服务器的部分隐私泄露问题，但同时又使得TTP成为了众矢之的，即通过该方法只是把隐私保护问题从服务器端转移到了TTP，其不能从根本上保证隐私不被泄露。为了克服TTP所带来的问题，研究者们提出用对等节点代替TTP，即用户之间通过协作的方式进行查询以期保护用户的隐私，但此方法同样假设对等节点是完全可信的[9]。

为了解决TTP以及对等节点完全可信这一假设在现实中是不存在的问题，研究者们通过加密用户的查询试图达到保护用户隐私的目的。文献[10]提出了一种隐私信息检索的办法，其通过加密用户的查询使得该方法具有高强度的隐私保护效果。但该方法开销巨大，严重的影响了用户体验，同时，该文献并没有利用用户的位置信息，用户得不到基于位置的搜索结果。文献[11]使用同态加密技术对用户的位置信息进行加密，同时实现了细粒度的用户查询。但同态加密技术开销大，对于移动应用环境而言，其实用性成为了最大的挑战。文献[12]采用矩阵加密的方式保护了用户的隐私。但该方法并没有考虑用户的位置信息，导致用户得不到基于位置的个性化搜索结果。

为了解决移动环境下基于位置服务的个性化搜索中用户隐私保护的难题，本文提出了个性化搜索中基于位置服务的隐私保护方法。结合矩阵加密和安全内积计算技术，首先，在用户端，用户将自身的位置信息进行转换以构建用户模型。然后用户转换查询使其带有用户的位置信息，并通过矩阵加密技术加密转换后的查询，并将其发送到云服务器。最后，在云服务器上使用安全内积计算对结果进行相关性排序，并返回相关性得分最高的前K个结果给用户。在整个查询的过程中，云服务器不需要完全可信，其并不能根据所获得的信息推测出用户的位置隐私、查询隐私。同时，由于只需要返回前K个最相关的搜索结果，这大大地节省了带宽，也在一定程度上解决了移动终端屏幕小引起搜索体验差的痛点[13,14]。

2　系统模型和相关定义

2.1　系统模型

如图1所示，基于位置服务的用户隐私保护模型由3类实体组成，分别为数据拥有者、用户和云服务器。数据拥有者负责生成密钥SK并将其与文件密文解密密钥Key通过安全通道传送给用户，他还负责构建文件的可搜索索引并将可搜索索引$\boldsymbol{I}$以及加密文件C上传到云服务器中。在用户端，用户通过位置转换的方法构建用户模型，进而生成用户查询，并将加密后的查询以及参数K发送到云服务器。云服务器得到用户的搜索请求后，根据安全内积计算用户查询与每个文件索引的得分，并返回得分最高的前K个文件给用户。该方法通过位置转换并加密的方式在保护用户隐私的同时保证了用户位置信息的可用性与准确性，避免了基于位置服务中需要可信匿名器的弊端，同时实现了用户个性化搜索与隐私保护的需求。同时，该方法能够大大地节省通信开销，提升用户的使用体验。

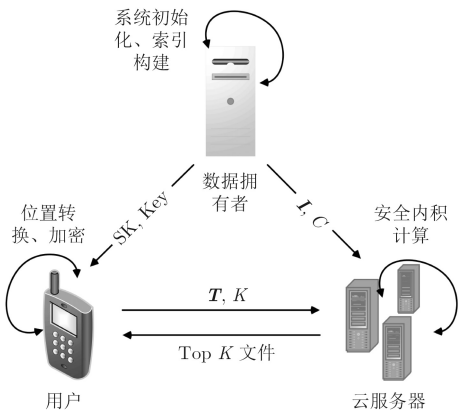


图 1 基于位置服务的用户隐私保护模型

2.2　相关定义

定义1　用户与查询点的距离　2元组(x, y)表示用户位置的经纬度，(x_i, y_i)表示查询点i的经纬度，为了简单起见，可以假设用户和查询点位于2维平面内，则它们之间的距离$d(i)$为

$$d(i)=\sqrt{(x_i-x)^2+(y_i-y)^2} \tag{1}$$

定义2 相关性得分 相关性得分用来度量$\boldsymbol{T}$与一个文件的相关性程度，我们用两个行矩阵的点积表示相关性得分，相关性得分越高，说明这两个行矩阵越相关，相应的文件也应该优先返回给用户。

定义3 安全内积计算[15] 为了保护用户与文件索引的隐私，需要加密用户的查询以及索引矩阵并计算两者之间的相关性得分，而安全内积计算能够实现这一目标，它允许云服务器在不知道$\boldsymbol{p}(i,:)$与$\boldsymbol{q}$的情况下，通过计算两者的密文$E(\boldsymbol{p}(i,:))$与$E(\boldsymbol{q})$，使得

$$E(\boldsymbol{p}(i,:))\cdot E(\boldsymbol{q})=\boldsymbol{p}(i,:)\cdot\boldsymbol{q} \tag{2}$$

即通过计算两个行矩阵加密后的内积可以得到两者加密前的内积，从而达到保护$\boldsymbol{p}(i,:)$与$\boldsymbol{q}$的目的。

2.3 安全模型

诚实而好奇模型(Honest-But-Curious, HBC)[16,17]：此模型中，攻击者严格遵守协议的整个流程，但为了某些目的，其希望从已知的信息中挖掘用户更多的敏感信息(例如：通过用户的网购记录推测用户的收入水平，或者通过用户的位置信息推测用户的家庭住址)，在这里，云服务器是诚实而好奇的。由于隐蔽式安全性是非常危险且愚蠢的，假设云服务器不仅知道密文，而且知道加密算法以及解密算法。为了更好地评价安全性，根据云服务器知道的信息，将其分为3个等级：

第1级 云服务器只知道密文信息C，加密后的索引$\boldsymbol{I}$以及加密后的查询矩阵$\boldsymbol{T}$。

第2级 云服务器不仅知道密文信息C，加密后的索引$\boldsymbol{I}$以及加密后的查询矩阵$\boldsymbol{T}$，而且知道一部分明文索引$\boldsymbol{p}_A$。

第3级 云服务器不仅知道密文信息C，加密后的索引$\boldsymbol{I}$以及加密后的查询矩阵$\boldsymbol{T}$，而且知道一部分明文索引$\boldsymbol{p}_A$及对应的密文值$\boldsymbol{I}_A$。

3 个性化搜索中基于位置服务的隐私保护方法

本方法包括5个阶段，分别是系统初始化阶段、索引构建阶段、用户模型构建阶段、用户查询生成阶段、索引搜索阶段。文中相关的符号描述如表1所示。

表 1 该文中的相关符号描述

符号	描述	符号	描述
SK	密钥	K	用户提交的参数K
$\boldsymbol{p}$	明文索引	a	大于0的随机数
$\boldsymbol{I}$	加密后的索引	$\boldsymbol{R}(i,:)$	与第i个文件的相关性得分
C	加密后的文件	$\boldsymbol{s}$	分裂指示器
$\boldsymbol{q}$	用户模型或用户查询	h	字典中的总关键词数
$\boldsymbol{U}$	用户模型	m	文件数量
$\boldsymbol{T}$	加密后的查询矩阵	$\boldsymbol{G}$	查询点的综合评分矩阵
n	真实的关键词数	t	随机生成的关键词数

个性化搜索中基于位置服务的隐私保护方法流程如图2所示。通过本文方法能够在提供基于位置服务的个性化搜索的同时保护用户的隐私不被泄露。

3.1 系统初始化阶段

在系统初始化阶段，数据拥有者随机生成$(n+t)$ bit的行矩阵$\boldsymbol{s}$，两个$(n+t)\times(n+t)$的可逆矩阵$\boldsymbol{M}_1,\boldsymbol{M}_2$作为密钥SK，因而，密钥SK是一个3元组$\{\boldsymbol{s},\boldsymbol{M}_1,\boldsymbol{M}_2\}$。

3.2 索引构建阶段

如表2的算法所示，在这一阶段，数据拥有者首先构建一个$n\times n$的对角矩阵$\boldsymbol{p}$，其中每一个对角元素代表了该文件表示查询点的综合评分(如餐馆的评分)。然后，将每一个明文索引$\boldsymbol{p}(i,:)$扩展为

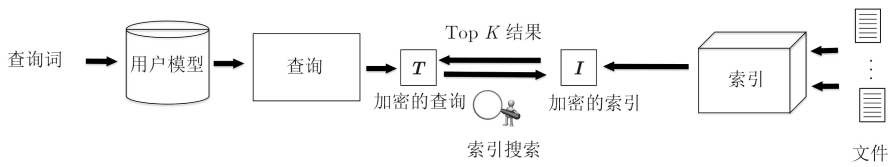


图 2 基于位置服务的隐私保护方法流程

$(n+t)$维并记为$\boldsymbol{p}^*(i,:)$$\boldsymbol{p}^*(i,:)$中$(n+g)\,(g\in[1,t])$维的位置键入随机数字。接下来，将$\boldsymbol{p}^*(i,:)$分裂为两个随机的$(n+t)$维行矩阵，并记为$\boldsymbol{p}'(i,:)$和$\boldsymbol{p}''(i,:)$。$\boldsymbol{s}$是分裂指示器，当$\boldsymbol{s}(j)$等于0时，$\boldsymbol{p}'(i,j)$和$\boldsymbol{p}''(i,j)$都等于$\boldsymbol{p}^*(i,j)$。当$\boldsymbol{s}(j)$等于1时，$\boldsymbol{p}'(i,j)$和$\boldsymbol{p}''(i,j)$是满足它们的和等于$\boldsymbol{p}^*(i,j)$的随机数。经过以上的步骤，加密的索引矩阵$\boldsymbol{I}=[\boldsymbol{p}'\boldsymbol{M}_1^{\mathrm{T}},\boldsymbol{p}''\boldsymbol{M}_2^{\mathrm{T}}]$。最后，云服务器保存来自数据拥有者的索引$\boldsymbol{I}$以及加密的文件$C$。

表 2　索引构建过程

算法 索引构建过程

输入： $\boldsymbol{G},s,m,\boldsymbol{M}_1^{\mathrm{T}},\boldsymbol{M}_2^{\mathrm{T}}$

输出： $\boldsymbol{I}$

(1) $p=\mathrm{diag}(\boldsymbol{G})$;
(2) $h=n+t$;
(3) $\boldsymbol{tp}=\mathrm{rand}(1,1)*\mathrm{ones}(1,t)$;
(4) **for** $i=1:m$ **do**
(5)　$\boldsymbol{p}^*(i,:)=[\boldsymbol{p}(i,:)\boldsymbol{tp}]$;
(6)　$\boldsymbol{r}=\mathrm{rand}(1,h)$;
(7)　**for** $j=1:h$ **do**
(8)　 **if** $\boldsymbol{s}(j)==1$ **then**
(9)　　$\boldsymbol{p}'(i,j)=\boldsymbol{r}(1,j)$;
(10)　　$\boldsymbol{p}''(i,j)=\boldsymbol{p}^*(i,j)-\boldsymbol{p}'(i,j)$;
(11)　　**else**
(12)　　$\boldsymbol{p}'(i,j)=\boldsymbol{p}''(i,j)=\boldsymbol{p}^*(i,j)$;
(13)　　**end if**
(14)　　**end for**
(15) **end for**
(16) $\boldsymbol{I}=[\boldsymbol{p}'\boldsymbol{M}_1^{\mathrm{T}},\boldsymbol{p}''\boldsymbol{M}_2^{\mathrm{T}}]$
(17) **return** $\boldsymbol{I}$

3.3 用户模型构建阶段

用户在查询前，用户端存储了资源中每个文件所代表查询点的类型以及位置信息{类型, (x_i,y_i)}，用户在提交查询关键词ω(如“餐馆”)的同时，会通过带有定位功能的移动设备获取自己当前的地理位置(x,y)。当用户提交的查询关键词ω与查询点的类型不相同时，查询点的权重$\boldsymbol{q}(i)=0$，否则，根据式(1)将用户的位置信息转换为用户与查询点之间的距离，进而根据式(3)计算查询点i的权重：

$$\boldsymbol{q}(i)=\frac{1}{d(i)} \tag{3}$$

最后，用户模型可以表示为

$$\boldsymbol{U}=(l_1:\boldsymbol{q}(1),\cdots,l_i:\boldsymbol{q}(i),\cdots,l_n:\boldsymbol{q}(n)) \tag{4}$$

若只用值来表示用户模型，则式(4)可以记为

$$\boldsymbol{q}=(\boldsymbol{q}(1),\cdots,\boldsymbol{q}(i),\cdots,\boldsymbol{q}(n)) \tag{5}$$

3.4 用户查询生成阶段

步骤1　用户查询的转换：当用户点击或提交查询关键词时，如“餐馆”，“超市”等，系统通过读取用户设备中的位置信息，生成相应的用户模型，并令其等于转换后的查询矩阵。即转换后的查询矩阵：

$$\boldsymbol{q}=(\boldsymbol{q}(1),\cdots,\boldsymbol{q}(i),\cdots,\boldsymbol{q}(n)) \tag{6}$$

步骤2　查询加密：首先，将用户查询矩阵$\boldsymbol{q}$扩展到$(n+t)$维，在$(n+1)$到$(n+t)$的位置用随机数填充，并保证这t个数的和为0，将其记为$\boldsymbol{q}^\star$。为了迷惑云服务器，本文将$\boldsymbol{q}^\star$的每一个位置的值乘$a(a>0)$。接下来，将$\boldsymbol{s}$作为分裂指示器，当$\boldsymbol{s}(j)$等于1时，$\boldsymbol{q}'(j)$和$\boldsymbol{q}''(j)$都等于$\boldsymbol{q}^*(j)$。当$\boldsymbol{s}(j)$等于0时，$\boldsymbol{q}'(j)$和$\boldsymbol{q}''(j)$是满足它们的和等于$\boldsymbol{q}^*(j)$的随机数。经过以上的步骤，加密的查询矩阵$\boldsymbol{T}=[\boldsymbol{q}'\boldsymbol{M}_1^{-1},\boldsymbol{q}''\boldsymbol{M}_2^{-1}]$。最后，用户将$\boldsymbol{T}$以及参数$K$上传到云服务器以获取个性化的查询结果。

3.5 索引搜索阶段

当用户将$\boldsymbol{T}$以及参数K上传到云服务器后，云服务器利用安全内积计算并根据式(7)计算每一个文件索引与$\boldsymbol{T}$的相关性得分。最后，云服务器将相关性得分最高的前K个密文返回给用户。

$$\boldsymbol{R}(i,:)=\mathrm{dot}(\boldsymbol{I}(i,:),\boldsymbol{T})=\boldsymbol{I}(i,:)\cdot\boldsymbol{T}=a\boldsymbol{p}(i,:)\cdot\boldsymbol{q} \tag{7}$$

用户在获得这K个密文后，只需用文件解密密钥Key进行解密，即可获得相应的明文信息，从而获得个性化的搜索结果。

4　安全性分析

本节，主要分析本文方法抵抗诚实而好奇的攻击者云服务器。具体分析如下

挑战 云服务器管理所有的密文、加密后的索引以及用户的查询历史，并希望从这些数据中挖掘出用户的某些隐私信息，从而获悉用户的兴趣以及位置等。同时，云服务器也希望知道文件的明文或索引信息。如果云服务器可以知道用户的兴趣、位置或者文件的明文、索引，那么云服务器将赢得这个游戏。

定理 1　本文方法能抵御云服务器的第1级攻击。

证明 对于第1级攻击，云服务器仅知道密文信息C，加密后的索引$\boldsymbol{I}$，加密后的查询矩阵$\boldsymbol{T}$以及加密解密算法，其并没有密钥，没有足够的信息破解经过加密的明文，索引以及查询矩阵，因此云服务器不可能知道用户的兴趣、位置，也不可能将文件的密文解密成明文。　　证毕

定理 2 本文方法能抵御云服务器的第2级攻击。

证明 对于第2级攻击，云服务器知道密文、加解密算法，但没有密钥，因此不可能将文件的密文解密成明文。但云服务器知道一部分明文索引$\boldsymbol{p}_A$，因而其可能依据用户查询与每个文件索引的相关性得分对用户的隐私信息进行揣测，本方法通过引入随机数a，即使查询矩阵$\boldsymbol{q}$相同，用户每次查询与同一文件索引的相关性得分也不一样。证明如下：

$$\begin{aligned}\boldsymbol{R}_1(i,:) &= \mathrm{dot}(\boldsymbol{I}(i,:),\boldsymbol{T}_1)=\boldsymbol{I}(i,:)\cdot\boldsymbol{T}_1\\ &= a_1\boldsymbol{p}(i,:)\cdot\boldsymbol{q}_1 \end{aligned}\tag{8}$$

$$\begin{aligned}\boldsymbol{R}_2(i,:) &= \mathrm{dot}(\boldsymbol{I}(i,:),\boldsymbol{T}_2)=\boldsymbol{I}(i,:)\cdot\boldsymbol{T}_2\\ &= a_2\boldsymbol{p}(i,:)\cdot\boldsymbol{q}_2 \end{aligned}\tag{9}$$

$$a_1\neq a_2,\boldsymbol{q}_1=\boldsymbol{q}_2 \tag{10}$$

由式(8)—式(10)可得$\boldsymbol{R}_1(i,:)\neq\boldsymbol{R}_2(i,:)$，得证。因此，云服务器不可能推断出用户的位置等隐私信息。

定理 3 本方法能抵御云服务器的第3级攻击。

证明 相对于云服务器的第2级攻击，在第3级攻击中云服务器知道一部分明文索引$\boldsymbol{p}_A$及其对应的密文值$\boldsymbol{I}_A$，这样云服务器可能采用暴力攻击的方法。本方法通过引入分裂指示行矩阵$\boldsymbol{s}$，将用户查询矩阵以及每个文件的索引分裂为两个行矩阵，云服务器要破解则必须知道分裂指示矩阵$\boldsymbol{s}$，为了增加本方法的安全性，我们还对查询矩阵与索引进行了维度扩展，即将n维扩展成$(n+t)$维。由于$\boldsymbol{s}$是一个二值向量，即$\boldsymbol{s}$中的每一个值有两种选择，云服务器要破解$\boldsymbol{s}$，需要尝试$2^{(n+t)}$次，如果令$n+t$=100，云服务器每秒能够尝试10^{12}次，也需要将近4×10^{10}年的时间，因此本文方法能够抵御云服务器的暴力攻击，所以本文方法对云服务器的第3级攻击是安全的，也就是说本文方法达到了挑战明文攻击安全，保证了用户与索引的隐私不被泄露。

5 实验及结果分析

实验主要从索引加密、用户查询生成以及索引搜索几个方面分析本方法的性能，并将其与MRSE[18]以及PRSE[12]方法进行比较。采用Yelp数据集中的“business”与“review”数据作为本实验的数据集。实验的硬件环境为：2.6 GHz Intel (R) Core (TM) i7-6700HQ CPU, 16.00 GB内存，操作系统为Microsoft Windows 10，采用Matlab R2016b实现，并使用OriginPro 2017对实验数据进行仿真。

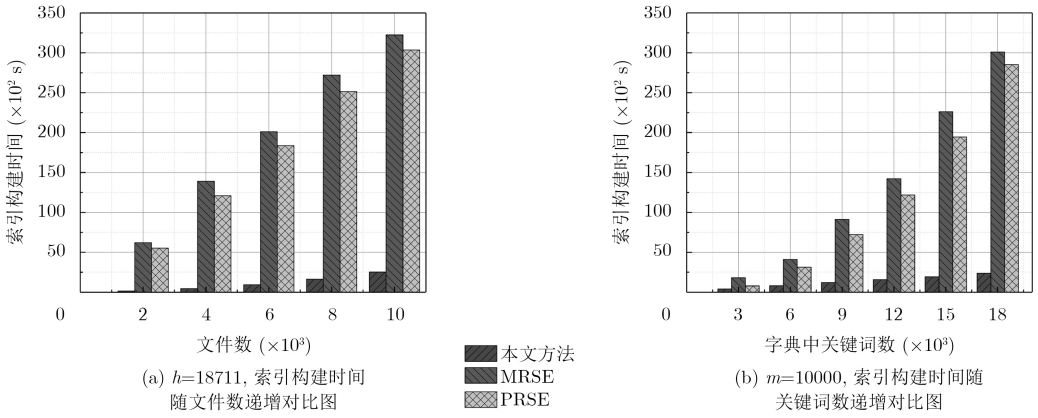
5.1 精确度

用户在搜索的过程中，尤其是在移动搜索的场景中，迫切地希望获得自己最需要的搜索结果。本文方法通过用户提交的查询关键词以及用户的位置信息优化查询结果，确保了用户查询的精确度。为了证明本文方法的精确性，我们随机地选取了10个用户，结果显示，有9人对返回结果满意，这也从侧面反映了本文方法是行之有效的。云服务器会根据参数K返回前K个文件给用户，如果设置K=10，而实际上与查询相关的文件有100个，那么召回率会很低，因此对于本文方法而言，讨论召回率是没有意义的。

5.2 构建索引

本文对比了采用矩阵加密构建索引的两种主要方法：MRSE[18]和PRSE[12]。为了实现个性化搜索以及隐私保护的目标，数据拥有者为每个文件构建可搜索索引并用矩阵加密，然后将索引$\boldsymbol{I}$以及加密的文件C上传到云服务器中。

图3(a)说明构建索引的时间随着文件数的增加而增加，这是很容易理解的。从图3(a)也可以看出，当字典中的关键词数为18711时，文件数从2000增加到10000的过程中，本文方法索引构建时间从145 s增加到2525 s, MRSE方法从6211 s增加到32243 s, PRSE方法从5531 s增加到30360 s。可



(a) h=18711, 索引构建时间随文件数递增对比图　(b) m=10000, 索引构建时间随关键词数递增对比图

图 3 基于位置服务的隐私保护方法索引构建时间模型图

以看出本文方法索引的构建时间明显低于MRSE和PRSE方法。从算法1可以看出，本文方法采用了先将每一个文件索引分裂为两个亚索引后，再对全部索引进行矩阵加密的方法，本文方法大幅度地缩短了构建索引的时间，减轻了数据拥有者的负担。

图3(b)说明索引的构建时间随着字典中关键词数的增加而增加。从中可以看到，当文件数为10000时，字典中关键词数从3000增加到18000的过程中，本文方法索引构建时间从411 s增加到2360 s，并且索引的构建时间随着字典中关键词数的增加呈线性增加。而在这一过程中，MRSE方法的索引构建时间从1812 s增加到了30111 s，其变化过程呈现2次曲线关系。在这一过程中，PRSE方法的索引构建时间与字典中关键词数呈现了抛物线的变化规律，从789 s增加到了28517 s。因而字典中关键词数越大，本文方法的优势越明显，这对拥有大量数据的数据拥有者来说选择本文方法的索引构建方法将达到事半功倍的效果。

5.3 构建用户模型

利用位置转换来构建用户模型是同时实现个性化搜索和隐私保护的根本，本文方法依据用户查询的兴趣类型，如“餐馆”、“超市”、“娱乐”，将用户的位置信息转换为与每个查询点的距离信息。

图4表示在用户端基于位置转换构建用户模型的执行时间，从中可以看出用户模型的构建时间与用户查询的兴趣类型中的查询点数息息相关，而与查询的兴趣类型无关，对应的查询点数越多，执行时间越长。

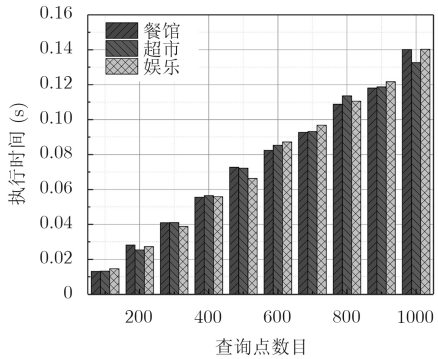


图 4 基于位置服务的隐私保护方法用户模型构建图

5.4 用户查询生成

将用户模型转换为用户的查询后，为了保证用户的隐私，需要对查询进行加密，对查询加密的方法与对一个文件的索引加密的方法类似，最大的不同是需要引入一个大于0的迷惑数a，由于a的值在每次查询之前都会随机生成，因此云服务器不可能恢复出用户查询与每个文件的相关性得分，也不可能根据之前的历史记录来推测将来查询的结果。因而通过引入迷惑数a，我们能够较好地保护用户的隐私不被泄露。

图5表示用户查询生成时间与字典中关键词数的关系，从图中可以看出用户查询生成时间随着字典中关键词数的增加而增加。当字典中的关键词数从3000增加到18000时，本方法用户查询生成时间从0.0409 s增加到0.3034 s，在这一过程中，MRSE方法从0.0212 s增加到0.2916 s，PRSE方法从0.0213 s增加到了0.2932 s。可以看出，对比MRSE与PRSE方法，本文方法的用户查询生成时间不占优势，但从中也可以看出，3个方法的差别不大，并且用户查询生成所花费的时间都非常小。

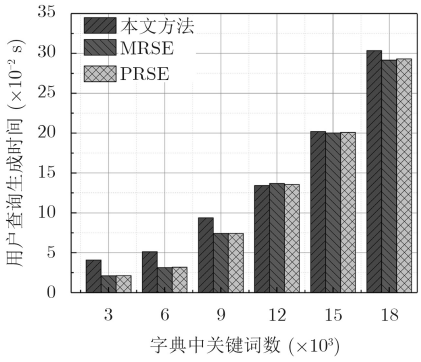


图 5 基于位置服务的隐私保护方法查询加密图

5.5 索引搜索

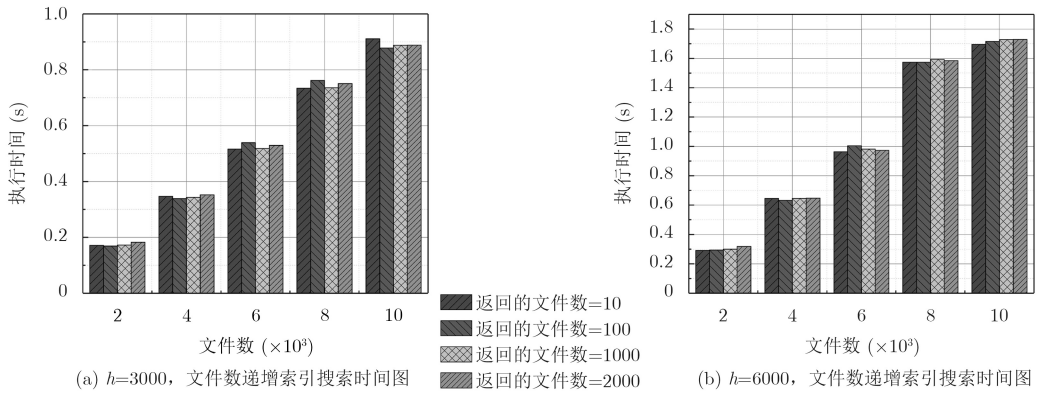
如何得到个性化的搜索结果，这关键的一步由云服务器完成，其在接收加密的查询矩阵$\boldsymbol{T}$以及参数K后。通过安全内积计算$\boldsymbol{I}$与$\boldsymbol{T}$之间的相关性得分，并将得分按从高到低排序，最后返回得分最高的前K个密文文件给用户。

图6(a)表明随着文件数的增加，索引搜索的时间也相应地增加，但索引搜索的执行时间与返回给用户的文件数，即K值没有关系。本文方法中，当返回文件数为10，文件总数从2000增长到10000时，索引搜索的时间从0.1715 s增加到了0.91104 s。

图6(b)与图6(a)对比表明，索引搜索时间随字典中关键词数的增加而增加。本文方法中，当返回文件数为10，文件总数为2000，字典中关键词数由3000增加到6000时，索引搜索时间从0.1715 s增加到了0.29065 s。

6 结束语

本文通过位置转换并结合矩阵加密和安全内积计算，在提供用户个性化搜索结果的同时，实现了用户信息的隐私保护，并证明了该方法的安全性。



(a) h=3000，文件数递增索引搜索时间图　(b) h=6000，文件数递增索引搜索时间图

图 6 基于位置服务的隐私保护方法索引搜索图

在构建索引时，本文方法采用了先将每一个文件索引分裂为两个亚索引后，再对全部索引进行矩阵加密的方法，构建时间与字典中关键词数呈线性关系，大大地减轻了数据拥有者的负担。同时，云服务器只需要返回相关性得分最高的前K个文件给用户，大大地减少了通信开销。并且，用户能够快速地找到自己最需要的个性化信息，这在一定程度上解决了移动设备屏幕小、操作不方便的弊端。当然该方法还有需要改进的地方，例如用户获得的个性化查询结果主要取决于用户查询的兴趣类型与位置信息，并不能够完全地反映用户的搜索意图，因此在下一步工作中，尝试将用户的搜索历史等个人信息融入到该方法中，从而获得更精确的个性化搜索结果。

参 考 文 献

[1] LU Rongxing, LIN Xiaodong, LIANG Xiaohui, *et al*. A dynamic privacy-preserving key management scheme for location-based services in VANETs[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2012, 13(1): 127–139. doi: 10.1109/TITS.2011.2164068.

[2] YU Rong, KANG Jiawen, HUANG Xumin, *et al*. MixGroup: Accumulative pseudonym exchanging for location privacy enhancement in vehicular social networks[J]. *IEEE Transactions on Dependable and Secure Computing*, 2016, 13(1): 93–105. doi: 10.1109/TDSC.2015.2399291.

[3] 张少波, 刘琴, 王国军. 基于网格标识匹配的位置隐私保护方法[J]. 电子与信息学报, 2016, 38(9): 2173–2179. doi: 10.11999/JEIT160350.

ZHANG Shaobo, LIU Qin, and WANG Guojun. The method of location privacy protection based on grid identifier matching[J]. *Journal of Electronics & Information Technology*, 2016, 38(9): 2173–2179. doi: 10.11999/JEIT160350.

[4] DU Qing, XIE Haoran, CAI Yi, *et al*. Folksonomy-based personalized search by hybrid user profiles in multiple levels[J]. *Neurocomputing*, 2016, 204(C): 142–152. doi: 10.1016/j.neucom.2015.10.135.

[5] ZHOU Dong, WU Xuan, ZHAO Wenyu, *et al*. Query expansion with enriched user profiles for personalized search utilizing folksonomy data[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2017, 29(7): 1536–1548. doi: 10.1109/TKDE.2017.2668419.

[6] PENG Tao, LIU Qin, and WANG Guojun. Enhanced location privacy preserving scheme in location-based services[J]. *IEEE Systems Journal*, 2017, 11(1): 219–230. doi: 10.1109/JSYST.2014.2354235.

[7] PAN Xiao, XU Jianliang, and MENG Xiaofeng. Protecting location privacy against location-dependent attacks in mobile services[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2012, 24(8): 1506–1519. doi: 10.1109/TKDE.2011.105.

[8] HWANG R H, HSUEH Y L, and CHUNG H W. A novel time-obfuscated algorithm for trajectory privacy protection[J]. *IEEE Transactions on Services Computing*, 2014, 7(2): 126–139. doi: 10.1109/TSC.2013.55.

[9] NIU Ben, LI Qinghua, ZHU Xiaoyan, *et al*.. Enhancing privacy through caching in location-based services[C]. Proceedings of the IEEE Conference on Computer Communications (INFOCOM). Hong Kong, China, 2015: 1017–1025. doi: 10.1109/INFOCOM.2015.7218474.

[10] CHOR B, GOLDREICH O, KUSHILEVITZ E, *et al*.. Private information retrieval[C]. Proceedings of the 36th Annual Symposium on Foundations of Computer Science, Washington, USA, 1995: 41–50. doi: 10.1109/SFCS.1995.492461.

[11] LI Xiangyang and JUNG T. Search me if you can: Privacy-preserving location query service[C]. Proceedings of the IEEE International Conference on Computer Communications (INFOCOM). Turin, Italy, 2013: 2760–2768. doi: 10.1109/INFCOM.2013.6567085.

[12] FU Zhangjie, REN Kui, SHU Jiangang, *et al*. Enabling personalized search over encrypted outsourced data with efficiency improvement[J]. *IEEE Transactions on Parallel &*

Distributed Systems, 2016, 27(9): 2546–2559. doi: 10.1109/TPDS.2015.2506573.

[13] ZHAO Feng, YAN Fengwei, JIN Hai, *et al*. Personalized mobile searching approach based on combining content-based filtering and collaborative filtering[J]. *IEEE Systems Journal*, 2017, 11(1): 324–332. doi: 10.1109/JSYST.2015.2472996.

[14] LEUNG W T, LEE D L, and LEE W C. PMSE: A personalized mobile search engine[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2013, 25(4): 820–834. doi: 10.1109/TKDE.2012.23.

[15] WONG Waikit, CHEUNG W L, KAO Ben, *et al*. Secure kNN computation on encrypted databases[C]. ACM SIGMOD International Conference on Management of Data. ACM, Providence, USA, 2009: 139–152. doi: 10.1145/1559845.1559862.

[16] 罗恩韬, 王国军. 移动社交网络中一种朋友发现的隐私安全保护策略[J]. 电子与信息学报, 2016, 38(9): 2165–2172. doi: 10.11999/JEIT151479.

LUO Entao and WANG Guojun. A novel friends matching privacy preserving strategy in mobile social networks[J]. *Journal of Electronics & Information Technology*, 2016, 38(9): 2165–2172. doi: 10.11999/JEIT151479.

[17] ZHANG Qiang, LIU Qin, and WANG Guojun. A privacy-preserving hybrid cooperative searching scheme over outsourced cloud data[C]. Security, Privacy, and Anonymity in Computation, Communication, and Storage. Zhangjiajie, China, 2016: 265–278. doi: 10.1007/978-3-319-49148-6_23.

[18] CAO Ning, WANG Cong, LI Ming, *et al*. Privacy-preserving multi-keyword ranked search over encrypted cloud data[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2014, 25(1): 222–233. doi: 10.1109/TPDS.2013.45.

张　强：男，1988年生，博士生，研究方向为隐私保护、个性化搜索.

王国军：男，1970年生，教授，博士生导师，研究方向为云计算、大数据、隐私保护.