doi:10.11959/j.issn.1000-436x.2017146

基于保序加密的网格化位置隐私保护方案

沈楠[1]，贾春福[1,2]，梁爽[1]，李瑞琪[1]，刘哲理[1]
（1. 南开大学计算机与控制工程学院，天津 300350；2. 中国民航大学信息安全测评中心，天津 300300）

摘 要：集中式可信第三方结构是基于位置的服务中常用的隐私保护结构。然而，一旦中心第三方服务器被攻破或不可信任，用户的位置隐私就有被泄露的风险。针对以上问题，提出一种用户自定义网格化的位置隐私保护方案，先将查询范围自动网格化处理，再结合保序加密技术，使用户的实时位置在隐藏状态下仍能进行比较。由于该方案在整个查询过程中，信息全程处于加密状态，服务器不知道任何用户的具体位置信息，增强了对用户位置隐私的保护；又由于该方案的中心第三方服务器只需要进行简单的比较操作，有效减少了它在处理大量数据时的时间开销。安全分析阐明了该方案的安全性，模拟实验结果表明该方案能够使中心服务器的时间开销明显降低。
关键词：可信第三方；基于位置的服务；位置隐私保护；用户自定义网格；保序加密
中图分类号：TP309 **文献标识码**：A

Approach of location privacy protection based on order preserving encryption of the grid

SHEN Nan[1], JIA Chun-fu[1,2],LIANG Shuang[1], LI Rui-qi[1], LIU Zhe-li[1]
(1.College of Computer and Control Engineering,Nankai University, Tianjin 300350, China;
2. Information Security Evaluation Center of Civil Aviation, Civil Aviation University of China, Tianjin 300300, China)

Abstract: The centralized structure of the trusted third party is a major privacy protection structure on location based services. However, if the central third party server can not be trusted or compromised, users have the risk of leakage of privacy location. Aiming at the above problems, location privacy protection approach based on a user-defined grid to hide location was proposed. The system first automatically converted the query area into a user-defined grid, and then the approach utilized order preserving encryption, which made the user's real-time position in the hidden state could still be compared. Because the information in the process of the approach was in a state of encryption, the server could not know the user's location information, thus improved privacy protection of the user location. The central third party server only need to do simple comparison work, so its processing time overhead would effectively decrease. Security analysis certificate the security of the proposed approach and simulation experimental show the proposed approach can reduce the time cost of the central third party server.
Key words: trusted third party, location based service, location privacy protection, user-defined grid, order preserving encryption

1 引言

随着移动互联网技术和定位技术的不断进步，基于位置的服务（LBS, location-based service）也迅猛发展，并受到越来越多的青睐。用户可以通过带有定位功能的智能移动终端获取自己实时的位置

收稿日期：2016-12-21；**修回日期**：2017-03-30
基金项目：国家重点基础研究发展计划（“973”计划）基金资助项目（No.2013CB834204）；国家自然科学基金资助项目（No.61272423, No.61672300）；天津市自然科学基金资助项目（No.16JCYBJC15500, No.17JCZDJC30500）；中国民航大学信息安全测评中心开放课题基金资助项目（No.CAAC-ISECCA-201702）
Foundation Items: The National Basic Research Program of China (973 Program)(No.2013CB834204), The National Natural Science Foundation of China (No.61272423, No.61672300), The Natural Science Foundation of Tianjin (No.16JCYBJC15500, No.17JCZDJC30500), Open Project Foundation of Information Security Evaluation Center of Civil Aviation, Civil Aviation University of China (No.CAAC-ISECCA-201702)

坐标，并将其上传到相关的服务器，就能够享受到服务商提供的各种 LBS 服务。如用户可获取周围一定范围内的宾馆、餐厅等兴趣点（POI, points of interest）的位置坐标和详细的说明信息。

然而，享受 LBS 便利服务的同时，隐私保护问题也不容忽视，尤其是用户敏感的实时位置信息，存在着严重的泄露风险。当前绝大多数的服务提供商都会不同程度地在服务器上收集用户的信息，把它们存储起来以便进行分析处理。如果这些服务器不可信任或被攻击，就会导致大量用户位置数据泄露，一旦它们被某些恶意攻击者掌握，就能从中分析出用户的隐私信息，如某个特定用户的工作地址、生活习惯等[1]，直接威胁人们的安全。

为了应对 LBS 服务中存在的位置隐私泄露问题，国内外已经展开了不少的研究工作。现有的保护方案主要依赖 2 种基本结构：分布式点对点结构和集中式可信第三方结构[2]。分布式点对点结构由用户端（user）和服务提供商（SP, service provider）2 个部分组成，user 之间通过互相协作的方式隐藏或混淆用户位置信息。集中式可信第三方结构，在 user 和 SP 之间引入一个中心查询服务器（QS, query server），专门集中负责隐私化处理用户的位置信息。该结构降低了 user 的负担，易于大规模部署，便于离线数据挖掘，是目前常用的系统结构。QS 主流的隐私保护机制是 *k*-匿名机制[3]，它通过将真实的用户位置混入 *k*−1 个其他匿名用户位置中，使攻击者无法分辨。但是该结构也有以下缺点：1) 由于用户的位置数据集中存放在 QS 中，就需要 QS 是完全可信第三方（TTP, fully-trusted third party），一旦 QS 被攻破或不再可信，会导致严重的隐私泄露；2) QS 集中对存放的位置信息进行隐私化处理，承担计算任务过重，极易出现性能瓶颈；3) *k*-匿名机制自身也存在很大的安全隐患。

由于 TTP 结构存在的问题，近年来位置信息隐藏技术开始被重视，将位置坐标网格匿名化[4~6]的相关研究成为一个新的研究热点。但在以前的网格化技术研究中，也存在一些缺陷。1) 应用 *k*-匿名技术修正用户的查询范围（user 自动搜索周围 *k*−1 个实际用户的位置和查询范围，将其合并为本次查询的扩大查询范围），在实践中很容易因为单位区域内用户密度不同，导致 2 种极端情况出现，由于用户密度过密，查询范围过小，从而暴露用户的位置；或查询范围过大，增加不必要的开销。此外，user 搜索周边 *k*−1 个实际用户的位置及查询范围，本身也会增加 user 额外的处理开销和安全隐患。2) 对每一个返回的查询结果，中心服务器 QS 都把它们与其自身存储的所有网格坐标进行一一匹配操作，进行结果筛选。这样，在实际场景中，很容易出现在查询范围比较大或为了提高系统的隐私保护程度所划分的网格数目很多的情况下，由于 QS 需要存储和比较的坐标数目过多，使 QS 计算量增长很快，从而影响其性能。

针对相关缺陷，本文在集中式第三方结构的基础上，摒弃不安全的 *k*-匿名机制，综合运用坐标自动网格化处理、保序加密机制、基于身份加密和完整性验证等多种密码学手段，提出了基于保序加密的网格化（OPEG, order preserving encryption of the grid）位置隐私保护方案。

本文方案的核心目标有以下几点。1) 提高集中式第三方位置隐私保护架构（TTP）的安全性，当中心服务器不完全可信时，依然能够保证用户隐私信息不被泄露。2) 有效解决中心服务器易阻塞的性能瓶颈问题。与用户使用直接相关的就是中心服务器，而高峰时段当大量用户同时在线查询时，短时间内查询任务量迅速上升，极易导致中心服务器出现拥堵问题，甚至崩溃。中心服务器的承压能力能否及如何正常工作是整个系统运行的关键。

OPEG 方案由 user 首先对周边区域进行网格化划分，并根据用户设定的隐私度来获得一个确定的查询范围，避免了可能出现的极端情况。与此同时，OPEG 方案将网格化后的查询范围进行保序加密，QS 中只保留能标识查询范围的 2 个加密坐标（只需加密并保存其右上角和左下角这 2 个坐标，即可确定整个查询范围），而不是存储其所有网格化坐标。利用保序加密“有序性”这一特点，QS 进行查询结果比较时，不管划分网格数目有多么庞大，都只需简单比较 2 对密态网格坐标的大小即可。因此，当查询范围扩大或划分网格数目增多时，都不会增加 QS 的计算量。同时，QS 只对密态网格坐标进行存储和比较操作，也有利于 QS 安全性的提高。

此外，先前的方案均未考虑用户服务注册的问题，用户真实身份容易被泄露；本文采取不可逆散列函数的手段，为用户关于身份的敏感信息提供更安全的保护。先前方案也未考虑消息的完整性问题，本文采取“时间戳”方案来验证消息可靠性和

完整性，进一步提高了系统安全性。

总之，在OPEG方案中，由于QS是在全密态下进行筛选，使其完全不能获得任何用户具体的位置信息，这极大地增强了该方案的安全性，使QS可以不完全可信；QS在整个查询过程中只需进行简单的数据比较操作，计算开销明显降低，有效缓解中心服务器在服务高峰期出现的性能瓶颈问题。OPEG方案可以达到期望的目标。

2 OPEG方案的系统模型

2.1 系统架构

OPEG方案的系统架构如图1所示，它只包含3个实体部分：用户端user、中心查询服务器QS和位置服务提供商SP。

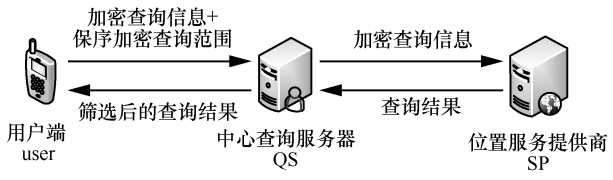

图1 OPEG方案的系统架构

user是拥有定位功能的智能移动终端，可以获得自身实时的位置坐标(x_u, y_u)。OPEG方案允许用户设定查询范围和隐私级别，周边区域的网格化匿名处理和查询范围的保序加密操作也由user来完成。

QS位于user和SP之间，负责把user发送的加密查询请求信息传递给SP。此外，QS也存储经保序加密处理后的网格化查询范围，用于对从SP返回的加密查询结果进行筛选，再返回给user。

SP拥有大型位置服务数据库，为用户提供各种LBS服务。它接收到QS转发的查询请求信息后，在其数据库中搜索到相应的POI信息，并将查询结果进行处理后，返回给QS。

查询方案流程如下。

1) user对周边区域进行网格化处理，并将处理方案、兴趣点类型、产生的密钥和时间戳等查询信息利用SP的公钥加密发送给QS。同时，user还将自己设定的查询范围进行保序加密处理，一并发送给QS。

2) QS存储经保序加密处理后的网格化查询范围，同时，将公钥加密的查询信息转发给SP。

3) SP解密查询信息，并按照查询请求查找到自身数据库内符合条件的兴趣点集，对其相关信息进行对称加密处理。然后，对每个兴趣点的实际坐标同样也进行相应的网格化保序加密处理，附在其加密信息之后，返回给QS。

4) 在保序加密的状态下，QS对从SP返回的加密查询结果进行简单的大小对比筛选。将符合查询范围内的兴趣点予以保留，返回给user；将其他不符合的兴趣点删除。

5) user还需对返回的查询结果解密，进行更进一步的求精操作。

查询的每一阶段都应满足相应的安全需求，具体步骤在第4节详述。

2.2 威胁模式

OPEG方案希望达到的目标是在满足查询要求的前提下，QS不能知道任何用户的实时位置信息及查询信息，SP不能获取用户的真实身份和确切位置。

在OPEG方案中，恶意攻击者可能通过监听不安全的无线链路，试图窃取用户的隐私信息；也可能通过重复攻击某个服务器，再结合已掌握的背景知识，来推断用户的实时位置信息。

OPEG方案设定只有在用户手动操作下，user才能发送含有匿名化用户位置的查询信息；平时不允许user自动发送此类信息，防止因短时间内自动的连续查询泄露位置隐私的可能性。

OPEG方案还设定QS和SP都是“诚实且好奇”的，一方面它们不会破坏协议流程，能按照协议忠实地完成自己的工作；另一方面，它们又都想从自己所掌握的信息中，分析出更多的关于用户的其他敏感信息。同时，更进一步设定QS和SP是不能相互勾结的，即它们不会同时被一个攻击者控制。这个设定是合理的，因为如果QS与SP相互勾结，对于用户将没有任何秘密可言。

3 密码学工具

OPEG方案在对数据加密问题上，使用一个不可逆的散列函数和3种加密模式：公钥加密、对称密钥加密和保序加密。不可逆的散列函数方法很多，本文采用SHA1方法。公钥加密技术使用一对非对称的密钥分别进行加密和解密操作，本文采用RSA加密算法，用于从user往SP间安全的传递查询请求信息。对称密钥加密技术使用同一个密钥进行数据的加解密操作，本文采用AES（advanced encryption standard）加密算法，用于从SP向user安全地返回具体的查询结果。保序加密算法如第3.2

节所示，用于对网格化坐标的密态处理和相应的筛选操作。

3.1　密钥交换协议

在 OPEG 方案中，每次查询前，均需要在 user 与 SP 之间建立起一个安全信道，以防攻击者窃听。安全信道的建立，可使用基于身份加密[7]（IBE, identity-based encryption）的密钥交换协议。

定义 1　基于身份加密对于给定的公钥身份 *ID*、安全参数 *k* 以及由私钥提取算法得到的对应私钥 *d*，使用私钥 *d* 能够从用 *ID* 加密的密文 *c* 中解读出其相对应的明文消息 *m*。即 $Dec(k, d, c)=m$，任取 $m\in D$，其中，$c=Enc(k, ID, m)$，*D* 为明文空间。

3.2　保序加密

保序加密最初由 Agrawal 等[8]提出，能有效查询已加密的数据。Boldyreva 等[9]进一步提出基于折半查找和超几何概率分布的保序对称加密（OPSE, order-preserving symmetric encryption）算法，增强了安全性。

定义 2　保序加密。如果加密函数 *Enc* 和明文空间 *D* 是确定性的，满足若 $x_1>x_2(x_1,x_2\in D)$，那么 $Enc(x_1)>Enc(x_2)$，则确定性加密函数 *Enc* 是保序的。

定义 3　保序对称加密。设 OPSE=(*k*,*Enc*,*Dec*)，其中，*k* 是由输入的安全参数生成的随机密钥，加密函数 *Enc* 和解密函数 *Dec* 均在明文空间 *D* 和密文空间 *C* 上，用 *k* 加密明文 *m* 可得到密文 *c*，同样用 *k* 解密密文 *c* 可得到明文 *m*。

3.2.1　保序加密在 OPEG 方案中的安全性

OPSE 满足不可区分选择明文攻击（IND-CPA）的安全性。假设函数 $LR=(\cdot,\cdot,b)$表示输入 m_0、m_1，得到 m_b，OPSE 是一个对称加密方案，$b\in\{0,1\}$，攻击者 *A* 考虑以下攻击实验[10]。

1) $\mathrm{Exp}_{\mathrm{OPSE}}^{ind-cpa-b}(A)$;

2) $K\xleftarrow{\$}k$;

3) $b'\xleftarrow{\$}A^{Enc(K,LR(\cdot,\cdot,b))}$;

4) 返回 b';

在以上实验中，要求每个查询(m_0,m_1)的消息长度相同，即$|m_0|=|m_1|$。攻击者 *A* 的优势为

$$Adv_{\mathrm{OPSE}}^{ind-cpa}(A)=\Pr\left[\mathrm{Exp}_{\mathrm{OPSE}}^{ind-cpa-1}(A)=1\right]-\Pr\left[\mathrm{Exp}_{\mathrm{OPSE}}^{ind-cpa-0}(A)=1\right] \quad (1)$$

对任意攻击者 *A*，其优势 $Adv_{\mathrm{OPSE}}^{ind-cpa}(k)$ 都是关于安全参数可忽略的，故 OPSE 是 IND-CPA 安全的。

因此，在 OPEG 方案中使用保序加密算法对网格坐标进行加密能够保证其安全性。

3.2.2　保序加密在 OPEG 方案中的可行性

在 OPEG 方案中，user 使用保序加密算法对查询范围的网格坐标进行加密，SP 也使用同样的保序加密算法对查询所得 POI 的网格化坐标进行加密，而 QS 需要根据 user 所提供的密态查询范围，将 SP 返回的查询结果进行筛选，这就需要 QS 既能对网格化坐标进行大小比较，在比较大小的同时又不能将坐标信息泄露给 QS。因此，使用保序加密算法可以在保护网格化坐标不被泄露的同时进行网格坐标的比较。

OPEG 中采用的是保序对称加密 OPSE 算法[9]。该算法主要分为 3 个阶段：建模、平铺和转换。在这 3 个阶段中，对算法开销影响最大的是建模阶段，当明文空间和密文空间都较大时，建模阶段中的桶划分过程计算负载较大。在本文方案中，user 和 SP 需要使用保序加密算法对网格化坐标进行加密，而网格化坐标是在用户实际坐标基础上计算出的相对坐标，这些网格化坐标均为数值较小的正整数，并且数值范围不会很大（一般情况下坐标最大值的数量级不超过 10^3），这也就意味着 OPEG 方案中保序加密算法的明文空间不会很大，因此，并不会造成很大的计算负载。

在 OPEG 方案中使用保序加密算法的最大优点是可以大幅度减轻 QS 的计算负载。此前的网格化方案[6]中，QS 需要存储查询范围内所有网格，并对查询结果的网格进行一一匹配操作；而在本文方案中，QS 只需保存 user 发送来的查询区域的 2 个加密后对角的网格化坐标。在 SP 返回查询结果时，QS 只需将密态查询结果和这 2 个密态坐标进行大小比较，就可以筛选出符合用户需求的结果。假设 SP 返回的网格坐标数量为 *n*，在此前的方案中，QS 筛选出符合用户要求的坐标所需的平均比较次数为 $O(n^2)$；而在 OPEG 方案中，QS 所需的平均比较次数为 $O(n)$。很明显，本文方案中 QS 的比较次数是呈线性增长的，而此前的方案为指数级增长。因此，OPEG 中使用保序加密算法可以大大减少 QS 的比较次数，从而减轻计算负载。

4　OPEG 方案设计

OPEG 方案共有 5 个步骤，分别是 user 的网格化处理、QS 的存储转发、SP 的查找兴趣点、QS

的筛选处理和 user 的结果求精，如图 2 所示，其中，M_{u_qs} 表示从 user 发送至 QS 的消息，其他消息类似表示。

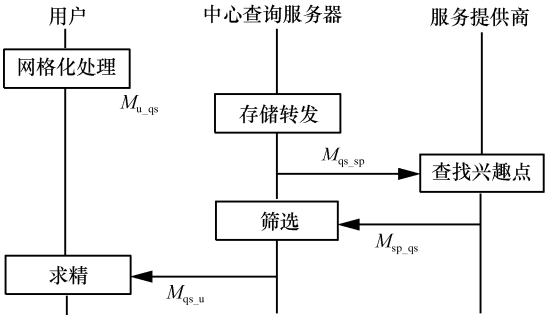

图 2 OPEG 方案的具体步骤

下面将分别描述 OPEG 方案各个步骤的具体实现情况，其中涉及的符号如表 1 所示。

表 1 OPEG 方案涉及符号的说明

符号	类型说明
r	查询半径
p	隐私保护程度
$type$	查询兴趣点类型
(x, y)	位置坐标
S	网格化的周边区域
(c, d)	网格化坐标
R	网格化的查询范围
key_{op}	OPSE 算法的密钥
T	网格化的扩大查询范围
$time$	时间戳
key_{u_sp}	user 生成的对称密钥
Pub_{sp}	SP 生成的公钥
ϕ	用 key_{u_sp} 加密的兴趣点实际信息
ψ	用 key_{op} 加密的兴趣点网格坐标

用户想要获取 LBS 服务，必须先注册。user 将用户的真实身份 *ID*，应用不可逆的散列函数 $H(\bullet)$，转换成与之相对应的“假 *ID*”，即 $FID=H(ID)$，并将其存储在 SP 中。注册结束后，user 和 SP 之间通过 IBE 密钥交换协议，建立起安全通信联系。而存储在 SP 中的 *FID*，也会定期更新。

4.1 user 的网格化处理

1) 用户设定本次查询参数

用户首先要在 user 设定查询范围 r、隐私保护度 p 以及兴趣点类型 *type*；若未设定则按上一次的查询的设定值执行。其中，r 决定查询范围的半径，*type* 是查询 POI 的类型，p 会影响用于隐藏用户实时位置的扩大查询范围的大小以及用户网格化周边区域的网格数目设定。例如，在隐私度为 8 的情况下，用户查询 1 km 范围内的餐馆和电影院，可以设定为：$\{r=1, p=8, type=\{4,6\}\}$。

2) user 形成网格化坐标体系

user 根据用户当前的位置(x_u, y_u)，自动地把用户周边街区划分为一个正方形的网格化周边区域 S。为了使用户此时位置坐标不能位于 S 的中心点，系统自动可将 S 进行随机方向的偏移调整，这样，用户的位置坐标将随机地出现在 S 区域的某一任意位置上。S 生成以后，它的位置可由其左下角的坐标 (x_0, y_0) 和右上角的坐标 (x_1, y_1) 来共同确定。然后，将 S 区域划分为大小相等的 $n\times n$ 个正方形网格，可表示为

$$S \leftarrow \{(x_0, y_0), (x_1, y_1), n\} \tag{2}$$

用户设定的隐私度 p 越高，n 越大，划分的网格越密，相应的安全性越高。在 S 网格结构中，每个小的网格都有其对应的网格化坐标(c, d)，其中，c 表示网格行坐标，d 表示网格列坐标，且 $1\leqslant c$，$d\leqslant n$。例如，图 3 中的用户当前位置(x_u, y_u)，则它的网格化坐标为

$$(c,d) \leftarrow \left(\left\lfloor \frac{x_u - x_0}{\frac{x_1 - x_0}{n}} \right\rfloor, \left\lfloor \frac{y_u - y_0}{\frac{y_1 - y_0}{n}} \right\rfloor \right) \tag{3}$$

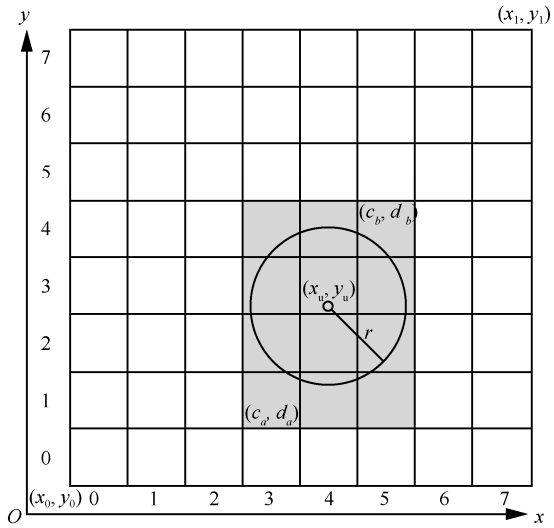

图 3 用户端网格化处理和确定查询范围

由以上定义可知，图 3 中用户所在的周边区域 S 为$\{(x_0, y_0), (x_1, y_1), 8\}$，用户当前位置为(4, 3)。

3) user 确定查询范围

user 在网格化环境中，以(x_u, y_u)为中心，半径为 r 的圆形范围作为本次查询范围，并将与其覆盖及相交的网格作为查询范围 R。如图 3 所示，阴影部分为查询范围 R，同样可以应用式(2)，转换成相应的网格化坐标，以便计算和比较。查询范围可由其左下角的坐标(c_a, d_a)和右上角的坐标(c_b, d_b)所共同确定。例如，图 3 查询范围 R 为{(3,1), (5,4)}。然后，用户使用 OPSE 算法，生成本次查询的密钥 key_{op}，用其将用户确定的查询范围 R 进行保序加密操作，得到加密后的查询范围 R'，可表示为

$$R' \leftarrow \{key_{op}(c_a, d_a), key_{op}(c_b, d_b)\} \tag{4}$$

4) user 生成扩大查询范围

为了使 SP 在查询过程中，不能知道用户确切的位置，本文方案采用了构造扩大查询范围的方法。具体方法是：先用 p 乘以一个在一定区间内的随机系数 k，再以坐标(x_u,y_u)为原点，向一个随机方向移动 $p\times k$ 距离，得到此时坐标为(x_{u1},y_{u1})。然后，获得以(x_{u1},y_{u1})为中心点，以｜$r+p\times k$｜长度为半径的圆形区域相交的网格集。如图 4 所示，阴影部分即为扩大查询范围 T，它同样可以被转换成网格化坐标，由其左下角的坐标(c_{a1},d_{a1})和右上角的坐标(c_{b1},d_{b1})确定。例如，图 4 中扩大查询范围为{(1,1), (6,7)}。这样，扩大查询范围可表示为

$$T \leftarrow \{(c_{a1}, d_{a1}), (c_{b1}, d_{b1})\} \tag{5}$$

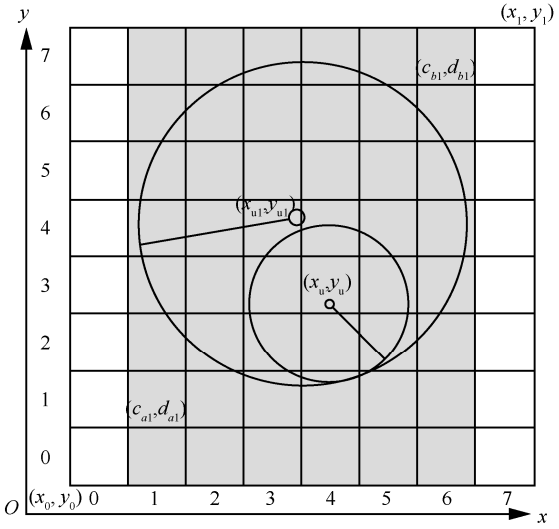

图 4　网格化的扩大查询范围

5) user 密钥的生成

为了数据在链路传输中的安全性，user 应用 AES 加密方法，随机生成本次查询的对称密钥 key_{u_sp}，用它加密从 SP 返回的具体查询结果。同时，为了防止篡改攻击，验证数据的完整性，user 还自动生成时间戳 $time$。

user 和 SP 之间，通过 IBE 密钥交换协议，可以使用户获得 SP 的公钥 Pub_{sp}，用户可使用它对查询集合{S, T, key_{op}, key_{u_sp}, $type$, $time$}进行加密，形成加密集 Pub_{sp}{S, T, key_{op}, key_{u_sp}, $type$, $time$}。

最后，用户将以上消息合并成查询请求消息 M_{u_qs}，发送给 QS，即

$$M_{u\_qs} \leftarrow \{R', Pub_{sp}\{S, T, Key_{op}, Key_{u\_sp}, type, time\}\} \tag{6}$$

4.2　QS 的存储转发

QS 接收到查询消息 M_{u_qs} 之后，留下并存储其中的经过 OPSE 加密后的查询范围 R'，用于对查询结果的比较筛选。将其他查询消息整理成新的查询请求消息 M_{qs_sp}，继续转发给 SP，即

$$M_{qs\_sp} \leftarrow \{Pub_{sp}\{S, T, Key_{op}, Key_{u\_sp}, type, time\}\} \tag{7}$$

4.3　SP 的查找兴趣点操作

SP 接收到查询消息 M_{qs_sp} 之后，首先使用自身的私钥 Pri_{sp} 来解密式(7)中的加密信息，然后再根据 S 提供的数据，在 SP 上恢复用户设定的“网格化周边区域”，并同时获得其中的扩大查询范围 T、保序加密密钥 key_{op}、对称密钥 key_{u_sp} 以及待查询的 POI 的类型 $type$ 和时间戳 $time$。

将扩大查询范围 T 建立在已经网格化的 S 上之后，SP 查找自己的数据库中坐标在 T 范围内所有符合 $type$ 类型的兴趣点，得到 t 个 POI 点集。如图 5 所示，阴影部分为扩大查询范围 T，图中三角代表 t 个 POI 在 T 中的分布，每个兴趣点都可以从 SP 的数据库中，找到对应的精确位置坐标和详细说明信息。例如，第 i 个兴趣点 POI_i，它的精确坐标为(x_i,y_i)，说明信息 $text_i$ 为{“万达电影院”}。它也同样可以转换成相应的网格化坐标(c_i,d_i)。这样，把所有查找到的兴趣点集中，每个点都可以分为 2 个部分{ϕ,ψ}，分别进行不同的加密操作。兴趣点具体坐标、说明信息和时间戳使用对称密钥 key_{u_sp} 加密，用于信息传递；兴趣点网格化坐标使用 OPSE 密钥 key_{op} 加密，用于返回 QS 进行筛选。

$$\phi_i \leftarrow key_{u\_sp}\{(x_i, y_i), text_i, time\} \tag{8}$$

$$\psi_i \leftarrow key_{op}(c_i, d_i) \tag{9}$$

$$POI_i \leftarrow (\phi_i, \psi_i)\ (1 \leqslant i \leqslant t) \tag{10}$$

多个兴趣点的网格化坐标只需加密操作一次即可。在图 5 中，网格坐标(4,5)，含有 2 个兴趣点，但它的保序加密只需加密一次，另一个点自动获得相同的密态数据，不需重复。将 t 个兴趣点信息对处理完毕后，整理成返回消息 $M_{\text{sp\_qs}}$，返回给 QS，即

$$M_{\text{sp\_qs}} \leftarrow \{POI_i\}, 1 \leqslant i \leqslant t \tag{11}$$

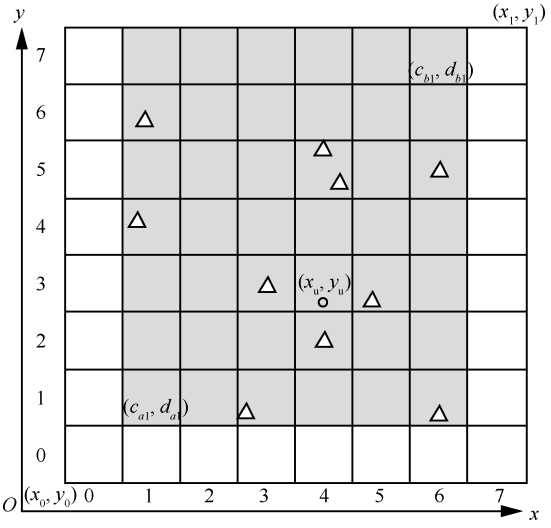

图 5 扩大查询范围内的兴趣点查找结果

4.4 QS 的筛选处理

QS 接收返回消息 $M_{\text{sp\_qs}}$ 之后，通过比较自身存储加密后的查询范围 R' 与 $M_{\text{sp\_qs}}$ 中每个 POI 的 ψ_i $(1 \leqslant i \leqslant t)$进行筛选，即属于范围$\{(c_a,d_a), (c_b,d_b)\}$内的兴趣点保留，其他兴趣点删除。由于保序对称加密操作的有序性，可以在加密状态下，判断被加密数据的相对大小。因此，很容易通过简单比较操作，筛选出所有满足条件的 POI 集，并组成返回信息 $M_{\text{qs\_u}}$ 转发给 user，假设共有 t' 个 POI 集，即

$$M_{\text{qs\_u}} \leftarrow \{POI_i\}, \ 1 \leqslant \text{i} \leqslant t' \tag{12}$$

如图 6 所示，筛选后 POI 集为$\{P_1, P_2, P_3, P_4\}$。

本次查询结束后，QS 自动删除整个查询期间保存的加密查询范围 R' 和所有收到的 SP 返回信息。

4.5 user 的结果求精

user 接收到返回消息 $M_{\text{qs\_u}}$ 后，用对称密钥 $key_{\text{u\_sp}}$ 解密每个 POI 中的 ϕ，得到其精确位置、详细说明信息和时间戳 *time*。首先，对每条信息都验证其时间戳 *time*，防止在信息转发过程中被篡改攻击或被添加假的兴趣点信息。验证每条信息的有效性后，再对查询结果进行进一步求精操作。由于实际查询范围是一个圆，而本文方案中的查询范围 R 是一个方形网格，因此，有可能存在少量的兴趣点恰好不在实际查询范围之内，故有必要继续求精。计算返回消息中每一个 POI 点到用户当前位置的距离，如果其距离超过查询半径 r，则舍弃。例如，图 6 中的 POI 点 P_2 就恰好不在查询范围之内，故舍弃。$\{P_1, P_3, P_4\}$匹配成功，这 3 个点即是本次查询操作最终的结果，再附上它们的相应说明信息，便可以显示到用户的智能移动终端的屏幕上。

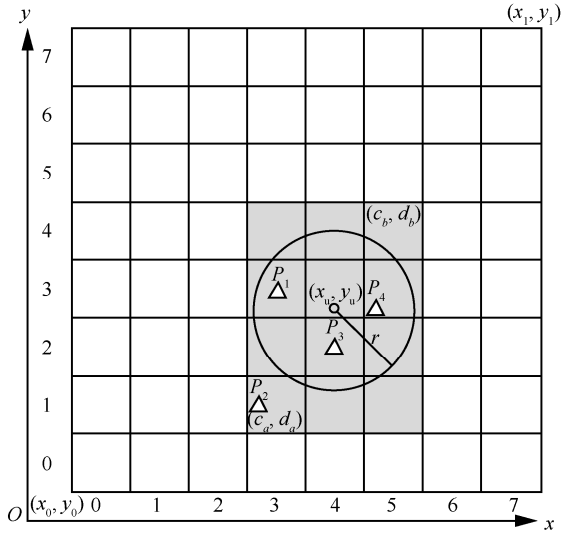

图 6 兴趣点查找结果筛选和求精操作

5 安全性分析

OPEG 方案设定了 QS 和 SP 都是诚实且好奇的，而且它们不能被同一个攻击者同时控制。本文方案希望达到的目标是在满足用户查询要求的前提下，QS 不能知道任何用户实时的位置信息及查询信息；SP 不能获取用户的真实身份和确切位置。

在以上设定的前提下，对几个定理进行证明，以说明 OPEG 方案的安全性。

引理 1 在已知任意网格化区域 S 的情况下，获得用户确切的位置信息的概率是可忽略的。

证明 若已知区域为 S=$\{(x_0,y_0), (x_1,y_1),n\}$，$(x_0, y_0)$和$(x_1, y_1)$为区域对角的坐标，$n$ 表示 S 内的网格数。那么用户的具体位置可能在任意一个网格内。因此，用户的具体位置在 S 中某一个网格内的概率为 $\dfrac{\dfrac{x_1 - x_0}{y_1 - y_0}}{(x_1 - x_0)(y_1 - y_0)} = \dfrac{1}{n}$，也就是攻击者能够成功获得用户位置所在网格的概率 $p_1 = \dfrac{1}{n}$。而在一个网格

内确定用户位置是可能发生的事件，但存在可忽略函数 $negl_1(n)$ 使在一个网格内确定用户位置的概率 p_2 满足 $p_2 \leqslant negl_1(n)$。那么在已知 S 的情况下获得确切的用户位置信息的概率 $p = p_1 p_2 \leqslant \frac{1}{n} negl_1(n) \leqslant negl(n)$。于是引理 1 得证。

定理 1　当攻击者在控制 QS 时，能够成功获得用户位置信息的概率是可忽略的。

证明　当攻击者在控制 QS 时，能够获得的有关用户位置信息的数据为 $M_{u\_qs} \leftarrow \{R', Pub_{sp}\{S, T, key_{op}, key_{u\_sp}, type, time\}\}$，其中，$R' \leftarrow \{key_{op}(c_a,d_a), key_{op}(c_b,d_b)\}$。从数据的组成来看，攻击者能获得的数据都是密文，分别是以 RSA 和 OPSE 的方式进行加密得到的。攻击者在不知道 RSA 的私钥时，成功获得 S 和 T 的概率等同于攻破 RSA 的概率，由 RSA 的安全性可知，攻击者成功攻破 RSA 的概率满足 $p \leqslant negl_1(n)$，那就意味着攻击者成功获得 S 和 T 的概率也应当满足 $p_1 \leqslant negl_1(n)$。同理也可知，攻击者成功获得 (c_a,d_a) 和 (c_b,d_b) 的概率满足 $p_2 \leqslant negl_2(n)$。而由引理 1 可知，在获得 S 和 T 或 (c_a,d_a) 和 (c_b,d_b) 之后，攻击者成功获得用户位置信息的概率满足 $p_3 \leqslant negl_3(n)$。因此，攻击者控制 QS 时能够成功获得用户敏感位置信息的概率满足 $p = (p_1 + p_2)p_3 \leqslant (negl_1(n) + negl_2(n))\, negl_3(n) \leqslant negl(n)$。于是定理 1 得证。

定理 2　当攻击者在控制 SP 时，能够成功获得用户确切位置信息的概率是可忽略的。

证明　当攻击者在控制 SP 时，能够获得的有关用户位置信息的数据为 S 和 T，由引理可知，攻击者能够获得用户确切位置信息的概率满足 $p \leqslant negl(n)$。于是定理 2 得证。

定理 3　当攻击者在控制 QS 时，能够成功获得查询结果信息的概率是可忽略的。

证明　当攻击者在控制 SP 时，能够获得的有关查询结果信息的数据为 $\phi_i \leftarrow key_{u\_sp}\{(x_i, y_i), text_i, time\}$ 以及 $\psi_i \leftarrow key_{op}(c_i,d_i)$。这 2 个部分数据分别是以 AES 和 OPSE 的方式加密后得到的，那么攻击者成功获得查询结果 (x_i,y_i) 或 (c_i,d_i) 的概率等同于攻破 AES 和攻破 OPSE 这 2 个事件中至少发生一个的概率。由已知的 AES 和 OPSE 的安全性可知，攻击者攻破 AES 和 OPSE 的概率分别满足 $p_1 \leqslant negl_1(n)$ 和 $p_2 \leqslant negl_2(n)$，于是攻击者成功获得查询结果的概率 $p = p_1 + p_2 \leqslant negl_1(n) + negl_2(n) \leqslant negl(n)$。于是定理 3 得证。

定理 4　当攻击者在控制 SP 时，能够成功获得用户身份与位置信息之间关联的概率是可忽略的。

证明　SP 中存储的是用户身份信息的不可逆散列函数值。由不可逆散列函数的特点可知，当攻击者在控制 SP 时，通过得知用户的 *FID* 来获取用户真实身份信息的概率不小于求 *FID* 在不可逆散列函数下的原象的概率。而求不可逆散列函数的原象的概率满足 $p \leqslant negl_1(n)$，那么攻击者获得用户真实身份信息的概率满足 $p_1 \leqslant negl_1(n)$。于是攻击者将当前查询中的用户位置信息与其真实身份对应起来的概率是可忽略的。定理 4 得证。

目前，主流使用的 *k*-匿名 TTP 方案，必须保证 *k*-匿名中心服务器是可信的。由于该方案传递的 *k* 条匿名信息中，本身就含有真实的信息，如果中心服务器不可信，非法收集匿名数据并发起重放攻击，再采用去匿名技术处理，会造成真实信息的泄露[11]。而 OPEG 方案中，将位置信息完全转换成相对网格坐标形式，并且对网格化方案采取了加密处理。由于在信息传递和处理过程中，攻击者无法解密获得网格化信息，传递的信息中又完全不包含任何真实位置信息，使攻击者无法分辨。所以，采用精确坐标网格匿名化及密钥加密等隐私保护手段的 OPEG 方案，不但能够达成所设定的安全目标，而且要比目前主流使用的 *k*-匿名方案更加安全。

6　性能测试与评估

实验硬件平台为 1 台 Win8 操作系统的 64 位 PC 机，CPU 为 2.4 GHz，内存为 4 GB。模拟实验在 Java 环境下，应用 MyEclipse 平台实现；利用 Brinkhoff 提供的基于交通网络的移动对象生成器，生成实验数据集，具体描述如下：采用美国 Hennepin 郡真实的交通路网图来作为输入，自动生成 5 000 个实时注册的用户及 10 000 个不同类型的兴趣点数据集。为了便于在同样的环境下比较，设定用户和兴趣点都是均匀分布的，且设定在 OPEG 方案中，扩大查询范围的半径为查询范围半径 r 的 2 倍。

实验的目的：1) 为验证本文方案的有效性，实验将 OPEG 方案分别与 *k*-匿名的 TTP 方案[12]、DGS 方案[6]里的中心服务器 QS 的性能进行比较；2) 为说明本文方案的整体性能，将测试 OPEG 方案中相关实验参数的变化对系统整体时间开销和通信开

销的影响。

6.1 中心查询服务器的性能对比

中心查询服务器 QS 位于集中式第三方结构的核心位置，高峰时很容易造成数据阻塞，使系统运行不畅。因此，QS 的性能是本文关注的焦点。

目前，常用的 TTP 方案[12]中，QS 会负责生成 $k-1$ 条假的查询点信息，来保证查询的安全性。匿名度 k 的增加可以使安全性提高，但同时也会导致 QS 性能的下降。为了便于比较，设定在 OPEG 方案中，扩大查询范围刚好可以完全覆盖 TTP 方案中 k 个查询点。下面通过实验，改变 k 的大小，比较这 2 种方案中 QS 的时间开销和空间开销。

图 7 为上述统一的实验环境中，OPEG 方案与 TTP 方案中 QS 的时间开销对比。通过对图 7 的分析可以看出，随着 k 的不断增大，查询范围也不断增大，OPEG 方案的优势越来越明显。这是因为在 OPEG 方案中，QS 只承担简单数据的比较工作；而在 TTP 方案中，QS 不但要负责 k-匿名的构造，还要承担去除大量 $k-1$ 个假匿名点的查询结果的工作。因此，OPEG 方案相对于 TTP 方案，在这方面优势非常明显，能有效解决 QS 计算量大的问题。

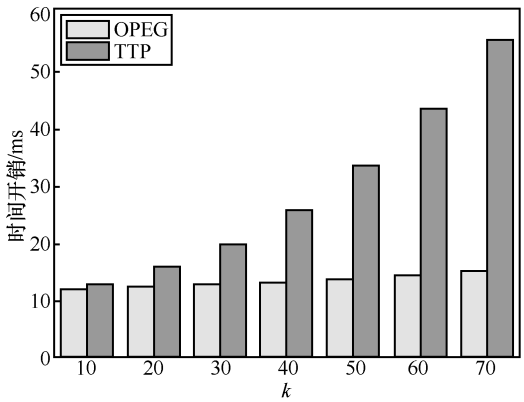

图 7 OPEG 与 TTP 中心查询服务器时间开销的对比

图 8 为上述统一的实验环境中，OPEG 方案与 TTP 方案中 QS 通信开销的对比。通过对图 8 的分析可以看出，TTP 方案略优于 OPEG 方案，这 2 种方案差距不大。这是因为 OPEG 方案中，QS 还要发送网格构造、密钥、时间戳等新增信息。

图 9 为 OPEG 方案与 DGS 方案中 QS 的时间开销的对比，n 为查询范围内的网格个数。通过对图 9 的分析可以看出，随着需要比较的划分网格数目不断增大，OPEG 方案的优势比较明显。这是因为在 DGS 方案要对所有存储的网格化坐标一一匹配，而在 OPEG 方案中，QS 对每个查询结果只需比较 2 个密态坐标，即可确定是否保留该查询结果。因此，OPEG 方案中 QS 时间开销增长幅度很小。

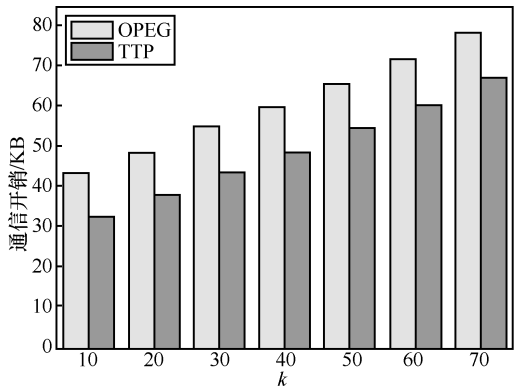

图 8 OPEG 与 TTP 中心 QS 通信开销的对比

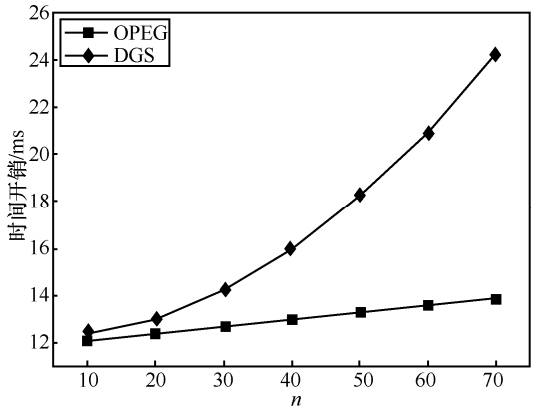

图 9 OPEG 与 DGS 中心查询服务器时间开销的对比

OPEG 方案与 DGS 方案通信开销差别很小，因为 OPEG 方案只需发送密钥、时间戳等少量额外信息。同时，由于在 OPEG 方案中，每次用户查询时，QS 只需存储 2 个密态坐标用于查询结果筛选，而 DGS 方案要存储全部网格化坐标。因此 OPEG 方案中，QS 所需的存储空间也相对要小。

综上，OPEG 方案只需付出少量通信开销为代价，就可以明显降低中心 QS 对于用户每次查询所需的时间开销，能够有效缓解当大量用户同时查询时，中心服务器所面临严重的计算压力。更重要的是，本文方案相对于原 TTP 方案和 DGS 方案，当用户查询范围扩大时，QS 时间开销增长非常小，这主要是因为在本文方案中，无论用户查询的查询范围多大，中心服务器只需要固定比较 2 个密态坐标即可。因此，OPEG 方案比原 TTP 方案和 DGS 方案更有优势。

6.2　查询半径和网格划分对 OPEG 整体性能的影响

实验观察查询半径 r 与网格划分个数 n 的变化对 OPEG 方案整体性能的影响。由图 10 和图 11 可知，时间开销和通信开销均随 r 和 n 的增大而增大，其中，时间开销随 n 增大的增长幅度更大。

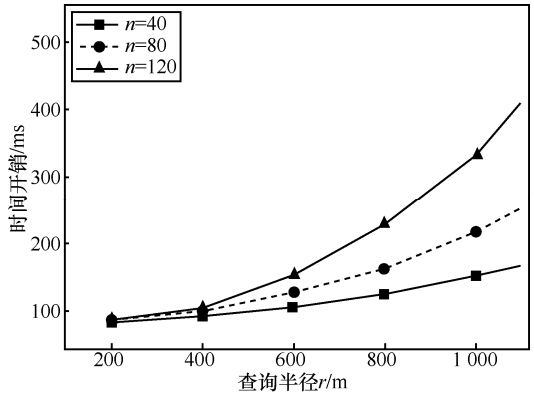

图 10　查询半径和网格划分的变化对系统整体时间开销的影响

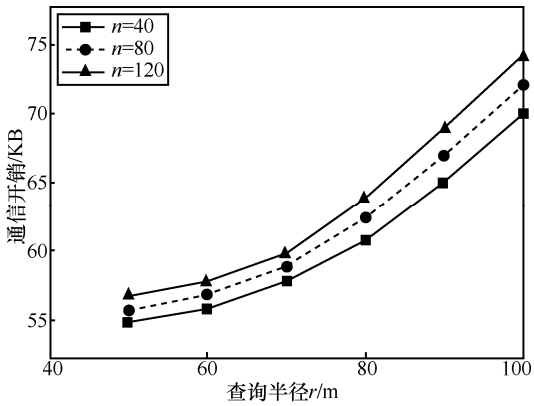

图 11　查询半径和网格划分的变化对系统整体通信开销的影响

这是因为查询半径 r 越大，形成的查询范围就越大，其所需的时间和通信开销也就越大。而 n 越大，单位网格面积就越小，网格个数增多，用户在相同查询范围中需要处理的网格数目也越多，系统所需处理的时间和通信开销也随着相应增加。

图 10 和图 11 显示在 OPEG 方案中，随着 n 的变大，时间开销的增幅要明显高于空间开销。因为随着查询范围的变大或网格被划分得更细，需要处理的网格坐标总数会增加。对于通信开销而言，需处理的网格区域增多，体现在其左上角和右下角这 2 个网格坐标值的变大，对存储空间并无影响；而查找到的兴趣点 POI 增加，这些 POI 的处理、发送操作，才是其通信开销增加的主要因素。这就是图 11 中空间开销增长趋势不明显的原因。

图 10 显示系统整体时间开销增长幅度大，这是因为系统整体时间开销主要受 user、QS 和 SP 这 3 个方面计算开销的影响。随着相关参数的变大，虽然 user 和 QS 计算量增长很微小，但由于需要处理的网格坐标和查找到的兴趣点都在增多，故 SP 端需要进行的保序对称加密操作也会相应增多，这会使 SP 的计算量明显增大，导致系统整体的时间开销增长幅度相对较大。

7　结束语

基于位置服务中的隐私保护问题，是当前热门的研究问题。本文在集中式第三方结构的基础上，针对以前方案存在的缺陷，舍弃不安全的 k-匿名机制，综合运用坐标自动网格化处理技术及保序加密机制，提出了一种基于保序加密的网格化位置隐私保护方案。OPEG 方案将具体的位置坐标转换为网格化形式，整个查询过程全密态进行，避免了扩展查询范围时可能出现的极端情况，提高了隐私保护程度；同时，由于中心服务器只需比较 2 个密态坐标的大小，有效减轻了高峰期中心服务器的计算压力。安全分析表明，OPEG 方案是安全的；通过实验与原 TTP 方案及 DGS 方案进行比较，OPEG 方案能有效解决中心服务器的性能瓶颈问题。但该方案也有不足的地方，由于 LBS 服务提供商 SP 需要给所有查找到的兴趣点网格化坐标进行保序加密，增加了其计算开销。因此，下一步工作计划尝试将 LBS 服务数据库分别部署在多个“云端”的服务提供商上，把需要进行的查找范围及兴趣点操作，也进行相应分割，分别在多个“云端”同步进行，使 OPEG 方案更加高效。

参考文献：

[1] 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, 26(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, 26(9): 2373-2395.

[2] GAO S, MA J, SHI W, et al. TrPF: a trajectory privacy-preserving framework for participatory sensing[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(6): 874-887.

[3] KIDO H, YANAGISAWA Y, SATOH T. Protection of location privacy using dummies for location-based services[C]//The 21st International Conference on Data Engineering Workshops. 2005: 1248-1248.

[4] DEBNATH R, VELAN G S. Semi trusted third party using dynamic grid system for locationbased services[J]. Networking and Communication Engineering, 2016, 8(5): 195-199.

[5] YI X, PAULET R, BERTINO E, et al. Practical approximate k nearest neighbor queries with location and query privacy[J]. IEEE Transac-

tions on Knowledge and Data Engineering, 2016, 28(6): 1546-1559.
[6] SCHLEGEL R, CHOW C Y, HUANG Q, et al. User-defined privacy grid system for continuous location-based services[J]. IEEE Transactions on Mobile Computing, 2015, 14(10): 2158-2172.
[7] GENTRY C, HALEVI S. Hierarchical identity based encryption with polynomially many levels[C]//Theory of Cryptography Conference. 2009: 437-456.
[8] AGRAWAL R, KIERNAN J, SRIKANT R, et al. Order preserving encryption for numeric data[C]//The 2004 ACM SIGMOD International Conference on Management of Data. 2004: 563-574.
[9] BOLDYREVA A, CHENETTE N, LEE Y, et al. Order-preserving symmetric encryption[C]//Annual International Conference on the Theory and Applications of Cryptographic Techniques. 2009: 224-241.
[10] AHMADIAN M, PAYA A, MARINESCU D C. Security of applications involving multiple organizations and order preserving encryption in hybrid cloud environments[C]// 2014 IEEE International Conference on Parallel & Distributed Processing Symposium Workshops (IPDPSW). 2014: 894-903.
[11] NARAYANAN A, SHMATIKOV V. De-anonymizing social networks[C]//2009 30th IEEE Symposium on Security and Privacy. 2009: 173-187.
[12] PENG T, LIU Q, WANG G. Enhanced location privacy preserving scheme in location-based services[J]. IEEE Systems Journal, 2014, 11(1): 219-230.

作者简介：

沈楠（1980-），男，天津人，南开大学博士生、讲师，主要研究方向为信息安全、可信计算、密码学及应用等。

贾春福（1967-），男，河北文安人，博士，南开大学教授、博士生导师，主要研究方向为信息安全与可信计算、恶意代码发现与分析。

梁爽（1992-），女，天津人，南开大学硕士生，主要研究方向为信息安全、隐私保护、应用密码学等。

李瑞琪（1993-），男，吉林德惠人，南开大学博士生，主要研究方向为密码学理论、密码技术应用等。

刘哲理（1978-），男，山东潍坊人，博士，南开大学副教授，主要研究方向为密码学及应用、智能卡操作系统。