移动社交网络中面向隐私保护的精确好友匹配

彭滔[1]，钟文韬[1]，王国军[1]，罗恩韬[2]，熊金波[3]，刘忆宁[4]，Hao Wang[5]

（1. 广州大学计算机科学与网络工程学院，广东 广州 510006；2. 湖南科技学院信息工程学院，湖南 永州 425000；
3. 福建师范大学计算机与网络空间安全学院，福建 福州 350117；4. 桂林电子科技大学计算机与信息安全学院，广西 桂林 541000；
5. 挪威科技大学计算机科学学院，挪威 约维克 2815）

摘 要：好友匹配通过比较用户间属性相似度向交友请求者推荐好友，是移动社交网络应用中的核心功能。然而，在好友匹配的过程中，用户个人信息很可能会被服务器或者其他恶意用户窃取导致隐私泄露，且现有方案存在匹配结果不精确或者无法满足用户多维度隐私保护需求等挑战。基于此，提出面向隐私保护的精确好友匹配（P3M）方案，查询者可以根据自身需求灵活设定特征属性和距离的匹配范围。利用可比较内积编码（CIPE）和 Paillier 加密算法对用户属性和查询范围进行编码和加密，并设计安全点积协议实现用户属性和查询范围的安全比较。相较于现有方案，P3M 方案支持查询者自定义查询范围以获得精确的查询结果，综合考虑用户特征属性及位置属性等多维度的隐私保护。最后，对 P3M 的正确性和安全性进行详细分析和证明，并通过实验验证 P3M 方案的有效性和高效性。

关键词：隐私保护；移动社交网络；好友匹配；保序加密

中图分类号：TN92

文献标志码：A

DOI: 10.11959/j.issn.1000−436x.2022208

Privacy-preserving precise profile matching in mobile social network

PENG Tao[1], ZHONG Wentao[1], WANG Guojun[1], LUO Entao[2], XIONG Jinbo[3], LIU Yining[4], Hao Wang[5]

1. School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China
2. School of Information Engineering, Hunan University of Science and Engineering, Yongzhou 425000, China
3. School of Computer and Cyber Security, Fujian Normal University, Fuzhou 350117, China
4. School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541000, China
5. Department of Computer Science, Norwegian University of Science and Technology, Gjøvik 2815, Norway

Abstract: Profile matching is a key feature in mobile social networking applications, where friends are recommended to requesters by comparing the similarity of attributes between them. However, users' personal information is exposed to the risk of privacy disclosure in the process of profile matching. The existing solutions exist some issues such as inaccurate matching results or inability to meet users' requirements for multi-dimensional privacy protection. Baesed on this, a privacy-preserving precise profile matching (P3M) scheme was proposed, which allowed users to flexibly set the matching range of attributes and distances according to their requirements. The Paillier encryption was utilized to ensure data security of users, and a secure dot product protocol was designed to achieve secure ciphertext comparison of user attributes and query ranges. The P3M realized multi-dimensional privacy-preserving of users including user feature attributes and location attributes. Finally, the correctness and security of P3M scheme were analyzed and proved in detail, and extensive experimental results verified the effectiveness and efficiency of P3M scheme.

Keywords: privacy-preserving, mobile social network, profile matching, order-preserving encryption

收稿日期：2022−06−06；**修回日期**：2022−10−18

通信作者：熊金波，jinbo810@163.com

基金项目：国家自然科学基金资助项目（No.U1905211, No.62272102, No.61872088, No.62072133, No.62172159）；国家重点研发计划基金资助项目（No.2020YFB1005804）；湖南省自然科学基金资助项目（No.2021JJ30294）

Foundation Items: The National Natural Science Foundation of China (No.U1905211, No.62272102, No.61872088, No.62072133, No.62172159), The National Key Research and Development Program of China (No.2020YFB1005804), Hunan Provincial Natural Science Foundation (No.2021JJ30294)

0　引言

智能手机和智能穿戴设备极大地方便了人们的生活，成为人们生活中不可或缺的工具。同时，各种移动社交应用成为人与人之间沟通交流的主要渠道。用户可以通过移动社交网络（MSN, mobile social network）分享照片、视频、位置等信息，并通过好友发现功能拓展自己的社交圈。然而，人们日常使用的手机应用软件中记录了大量的个人隐私信息，越来越多的用户对应用软件使用过程中个人隐私的安全问题产生了担忧。很多网络交友软件基于用户属性的相似度匹配方式来实现用户间配对，这种方式将用户的属性值（如年龄、性别、爱好等）进行比较，并将相似度高的用户建立联系，从而达到交友的目的。如何在匹配过程中寻找一种既高效又能保护隐私的匹配方案成为亟须解决的问题。一些研究提出了社交网络中好友匹配隐私保护的解决方案，如基于隐私集合交集（PSI, private set intersection）的方法[1]旨在通过求取用户属性集合的交集或交集中元素个数来判断 2 个用户的相似度。为了满足用户细粒度的需求，基于用户之间属性向量点积的相似度计算方法[2]使查询者能够对每个属性设定不同的权重。随着研究的推进，基于密码学和秘密共享[3-4]等新方案不断被提出。然而，现有的细粒度好友匹配隐私保护方案仍然存在如下问题。

1) 现有方案采用点积相似度计算方法实现的细粒度好友匹配往往不够精确，因为用户不同属性的取值范围可能是不同的，例如，用户对某事物的喜好程度的取值范围是[0,10]，而年龄的取值范围是[0,100]，假设 2 个用户之间的喜好程度和年龄的差值都是 1，虽然差值相等，但是对 2 个属性来说相差的程度是不同的，这就会导致匹配结果不精确。

2) 现有方案只考虑了用户特征属性的相似性，没有考虑位置属性，无法满足查询者基于位置设定范围的要求。在现实生活中，查询者往往不希望匹配到的用户离他过于遥远，因此对于查询者来说，限定与目标用户间的距离也是一个非常重要的因素。

针对上面所提到的 2 个问题，本文提出 P3M（privacy-preserving precise profile matching）方案，查询者可以为每个属性设定查询范围和权重，并且在比较用户属性相似度的基础上加入对用户位置范围的考虑，能够实现对查询者指定范围内的目标用户进行匹配，并在此基础上使用加密方法实现用户多维度的隐私保护。本文主要贡献如下。

1) 提出 P3M 方案，允许用户根据不同查询需求对目标用户的属性值以及所在位置设定查询范围，在保护用户隐私的同时满足用户更精确的匹配需求。

2) 采用轻量级的可比较内积编码（CIPE, comparable inner product encoding）算法[5]，利用编码前后不会改变明文之间大小关系的特性，实现对属性值和位置设定范围的匹配。

3) 采用 Paillier 加密算法设计一个安全点积协议（SDPP, secure dot product protocol），避免匹配过程中的隐私泄露并且能够防止用户和服务器的合谋。

1　相关工作

早期对于好友匹配隐私保护的研究中，文献[6]提出了一种 PSI 方案，这种方案可以实现求取双方所拥有的集合的交集而不会泄露双方集合中的信息；在文献[7]中，此方法被用于在社交网络中求用户间的相似度；为了避免复杂的密码学计算，基于布隆过滤器的好友匹配方案[8]被提出，由于采用 Hash 函数等伪随机函数进行映射，该方案能够更加快速地求出 2 个集合的交集，提高了匹配的效率。基于 PSI 的方法仅考虑了用户之间相同的属性个数，而在现实生活中查询者对不同属性的偏好程度是不同的，例如，如果 Bob 想要寻找一个一起打篮球的朋友，可能会更关注这个人的身高是否符合需求，而对于这个人的年龄、性别等属性关注度较低，基于 PSI 的方法并不能满足此类需求。

随着研究的推进，基于用户属性向量点积的相似度计算方法[2]被提出，通过求 2 个用户属性向量的点积从而计算出 2 个向量的相似度，可以满足细粒度的查询需求；基于秘密共享[4]等属性相似度匹配方案可以实现多个用户共同计算相似度，并保证整个过程中每个用户的隐私不被泄露。

随着移动设备性能的增强，基于密码学的方案[3,9,10-18]也被提出。在基于属性基加密（ABE, attribute based encryption）的方案[17]中，用户根据自己的属性生成访问控制策略，将其嵌入密钥对信

息进行加密得到密文，其余用户只有当属性个数达到加密者所设定的阈值时才可以对密文进行解密。文献[19]使用同态加密、属性基加密和代理重加密实现了在外包云服务器上的一种细粒度的数据共享和访问控制方案，数据拥有者将加密的数据和设定的访问控制策略一并上传至服务器，当数据使用者请求访问数据时，服务器会根据访问控制策略判断该用户是否为授权用户以及该用户能够访问的部分，并使用代理重加密技术对相应密文数据进行重加密，使数据使用者能够用其私钥解密，整个过程中服务器无法得到数据中的任何信息，确保了隐私的安全；文献[2]使用同态加密和代理重加密技术提出了一种在双云服务器辅助下进行属性文件相似度匹配的方案，采用属性向量点积的相似度计算方法计算用户间的相似度并且将向量点积计算工作分发给2个服务器共同完成计算，利用同态加密实现对密文的运算并且在服务器通信过程中使用盲化操作来保证2个服务器能够共同完成计算而不泄露用户的任何隐私信息，实现了隐私保护并且能够抵御用户和服务器的同谋。

现有的基于属性相似度的匹配方案可大致分为中心化方案和分布式方案。中心化方案在用户之间引入了一个可信第三方（TTP, third trusted party）服务器，形成了“用户–服务器–用户”的架构。在分布式方案中，服务器具有超强的计算和存储能力，用户只需要将属性信息上传至服务器，剩余的属性相似度计算工作均由服务器完成，能够有效解决用户端计算资源不足、存储能力有限等问题，对用户端性能要求较低。文献[2,4,8-9,20]均采用中心化方案。然而中心化方案对服务器可信程度的要求非常高，在日常生活中很难保证服务器的完全可信，如何在服务器不完全可信的情况下保护用户的隐私仍需要进一步研究。与之对应的分布式方案则能够实现用户之间点对点的匹配而不需要中心服务器的参与。分布式方案[21-24]解决了中心化方案的服务器可信问题，但是也引入了新的问题。由于分布式方案的所有匹配操作均是在移动终端设备中进行的，因此移动设备计算和存储能力有限导致分布式方案效率受到限制的问题也凸显出来，并且分布式方案更容易受到恶意用户的攻击，从而导致用户隐私的泄露。由于中心化方案和分布式方案优缺点各异，文献[19]结合了中心化方案和分布式方案的优点，提出了一种混合的方案。

总之，现有方案从一定程度上解决了好友匹配过程中的隐私泄露，但是仍然难以做到更加精确的匹配，并且在属性相似度匹配过程中没有将用户特征属性和位置信息相结合实现更加细粒度的匹配。基于这两点，P3M方案给出了解决方案，采用双服务器的中心化架构保证了在中心服务器不完全可信的情况下用户隐私的安全，使用 CIPE 算法和 Paillier 加密算法实现了查询者对用户属性值设定范围的精确查询，并且设计了安全点积协议来避免匹配过程中的隐私泄露。

2 预备知识

2.1 Paillier 密码系统

1) 密钥生成(1^k)

给定安全参数k，随机选择2个长度为k的大素数p,q，且满足$\gcd\big(pq,(p-1)(q-1)\big)=1$；计算$n=pq$，$\lambda=(p-1)(q-1)$；随机选择整数$g\in Z_{n^2}^*$，计算$\mu=\left(L(g^\lambda \bmod n^2)\right)^{-1}$，其中$L(x)=\dfrac{x-1}{n}$；生成公钥$\mathrm{pk}=(n,g)$和私钥$\mathrm{sk}=(\lambda,\mu)$。

2) 加密

给定公钥(n,g)和明文$m(0\leqslant m<n)$，所对应的密文为$c=g^m r^n(\bmod n^2)$，其中r为随机选取的整数，且$0<r<n, r\in Z_n^*$。

3) 解密

给定私钥(λ,μ)和密文$c,\ c\in Z_{n^2}^*$，所对应的明文为$m=L(c^\lambda \bmod n^2)\mu \bmod n$。

4) 加法同态性

Paillier加密算法具有加法同态性，具体如下。

① $E_{\mathrm{pk}}(m_1)E_{\mathrm{pk}}(m_2)=E_{\mathrm{pk}}(m_1+m_2)$

② $E_{\mathrm{pk}}(m_1)^{m_2}=E_{\mathrm{pk}}(m_1 m_2)$

2.2 CIPE 算法

本文在前期的研究中提出了 CIPE 算法[5]，其主要思想是将2个数值a,b映射到2个向量$\boldsymbol{p}$和$\boldsymbol{q}$，使$\boldsymbol{p}$和$\boldsymbol{q}$点积值的符号和$(b-a)$值的符号一致，即$\boldsymbol{p}\cdot\boldsymbol{q}>0$则$(b-a)>0$，$\boldsymbol{p}\cdot\boldsymbol{q}<0$则$(b-a)<0$。本文对原始的CIPE算法进行了改进，使经过 CIPE 算法加密所得的向量中的每个元素均来自Z_p（p为素数），算法细节如算法1和算法2所示。

算法 1　属性向量编码

输入　属性值 $a_{i_k} \in Z_p$（p 为素数），长度为 $\frac{d}{2}$ 的比特串 L_1, L_2

输出　长度为 d 的索引向量 $\boldsymbol{p}_{i_k}$

1) for $\iota \in \left[1:\frac{d}{2}\right]$ do
2)　if $L_1[\iota]==1$ then
3)　　$\boldsymbol{p}[\iota] := a$
4)　else
5)　　$\boldsymbol{p}[\iota] := 1$
6)　end if
7) end for
8) 从 $Z^+ \cup \{0\}$ 生成一个 $\frac{d}{4}$ 维向量 $\boldsymbol{C} = \left(c_1, c_2, \cdots, c_{\frac{d}{4}}\right)$，其中 $\sum_{\iota=1}^{\frac{d}{4}} c_\iota > 0$
9) 令 $x := y := 1$
10) for $\iota \in \left[\frac{d}{2}+1:d\right]$ do
11)　if $L_2\left[\iota - \frac{d}{2}\right]=1$ then
12)　　$\boldsymbol{p}[\iota] := ac_x$
13)　　$x = x+1$
14)　else
15)　　$\boldsymbol{p}[\iota] := c_y$
16)　　$y = y+1$
17)　end if
18) end for

算法 2　查询向量编码

输入　查询值 $b_{j_{k,\phi}} \in Z_p$（p 为素数），长度为 $\frac{d}{2}$ 的比特串 L_1, L_2

输出　长度为 d 的索引向量 $\boldsymbol{q}_{j_{k,\phi}}$

1) 从 $Z^+ \cup \{0\}$ 生成一个 $\frac{d}{4}$ 维向量 $\overline{\boldsymbol{C}} = \left(\overline{c}_1, \overline{c}_2, \cdots, \overline{c}_{\frac{d}{4}}\right)$，其中 $\sum_{\iota=1}^{\frac{d}{4}} \overline{c}_\iota > 0$
2) 令 $x := y := 1$
3) for $\iota \in \left[1:\frac{d}{2}\right]$ do
4)　if $L_1[\iota]=1$ then
5)　　$\boldsymbol{q}[\iota] := (p-1)\overline{c_x}$
6)　　$x = x+1$
7)　else
8)　　$\boldsymbol{q}[\iota] := b\overline{c_y}$
9)　　$y = y+1$
10)　end if
11) end for
12) for $\iota \in \left[\frac{d}{2}+1:d\right]$ do
13)　if $L_2\left[\iota - \frac{d}{2}\right]=1$ then
14)　　$\boldsymbol{q}[\iota] := p-1$
15)　else
16)　　$\boldsymbol{q}[\iota] := b$
17)　end if
18) end for

3　详细设计

3.1　系统模型

本节通过一个实际的应用场景来阐述 P3M 方案。考虑一个移动社交网络中的交友场景，即 Alice 想要查询 5 km 范围内具有相似特征属性（如年龄、性别、爱好等）的好友，每一个属性对应一个相应的数值，如年龄为 25 岁，性别为 0（女性），对购物的喜好程度为 8。考虑细粒度查询需求，Alice 可以设定一个查询范围，如年龄取值范围为 20～30 岁，性别为女性，对购物的喜好程度取值范围为 8～10 等。Alice 将这些查询信息发送给服务器，服务器进行匹配，并将符合条件的结果返回给 Alice。以上场景中，服务器是诚实且好奇的实体，即能够诚实地执行每一条指令，但会试图获取用户的隐私信息。在整个匹配过程中，Alice 和其他参与交友的用户并不想将任何隐私信息（包括用户上传的属性值信息以及查询信息）透露给除自己之外的任何第三方实体。

在现有的细粒度好友匹配方案中，用户属性间取值范围的不同会使用户间属性值的差异程度存在不同。如上述场景中，用户年龄取值范围是 0～100，而用户对某一事物的喜好程度取值范围则是 0～10。这使采用属性向量点积相似度计算方法得到的结果不够精确。此外，在传统的依赖

可信第三方服务器进行相似度匹配的系统中，收集用户上传属性和进行相似度计算的工作均由同一服务器完成，因此服务器是否可信对方案的安全性至关重要，对用户则存在隐私泄露的风险。针对上述问题，查询者可以针对用户属性值和位置设定查询范围，并向服务器提交加密后的数据，服务器在不需要解密密文的情况下对用户属性值与查询范围进行比较，从而获得更精确的结果；其次，为了保护用户的隐私，P3M 方案在云端部署了 2 个服务器，将原本由一个服务器处理的工作分配给 2 个互不合谋的服务器来完成，利用密码学等技术保证整个过程中用户隐私的安全。这样做的好处在于每个服务器均只参与部分计算任务，无法根据已有数据还原出完整的用户数据，从而保护了用户的隐私。具体的安全性证明将在 4.2 节给出。

P3M 方案框架如图 1 所示，该框架由查询者 Alice、用户 Bob、服务器 A（SA, server A）和服务器 B（SB, server B）组成。用户和查询者分别采用 CIPE 算法对上传的属性值（特征属性值和位置属性值）和查询范围的上下界进行编码，编码后生成固定长度的向量。服务器能够通过向量之间点积的结果判断 2 个编码向量所对应属性值的大小关系，从而在不需要解码的情况下判断属性值与所给范围上下界的大小关系，进一步判断属性值是否在范围内。为了实现查询者和用户之间的编码向量相乘，P3M 方案中所有的用户和查询者均采用同一个 CIPE 比特串进行编码，SA 负责这一比特串的分发。然而公开的比特串无法保证用户隐私的安全，因此用户和查询者还需执行 Paillier 加密算法，使用 SB 的公钥对 CIPE 向量中的每个元素进行加密，得到密文向量并发送给 SA 存储。SA 和 SB 共同计算密文向量的点积并得出用户之间的相似度。SA 利用 Paillier 加密算法的同态性实现基于密文的计算，在此过程中无法获取任何用户的隐私信息；SB 拥有强大的计算和存储能力，负责执行 P3M 方案中的大部分计算任务；SA 和 SB 之间的通信数据均需进行盲化处理，很好地保证了服务器交互计算期间用户隐私的安全。基于 P3M 方案框架的匹配方案分为 3 个过程。

1) Alice 和 Bob 分别生成安全查询向量和安全索引向量并发送给 SA。

2) SA 和 SB 共同计算相似度。

3) SA 将最终的匹配结果发送给 Alice。

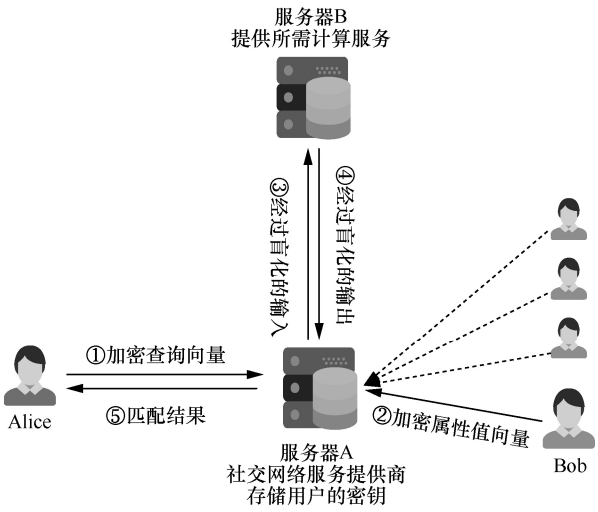

图 1 P3M 方案框架

在整个过程中，SA 和 SB 均无法获取所有用户上传的明文信息，用户之间也无法获取对方任何隐私信息，从而保护了用户数据安全。本文方案的符号描述如表 1 所示。

表 1 符号描述

参数	含义
$\mathcal{D}$	所有用户的数据
$\mathcal{A}_i$	第 i 位用户的属性集合
Q_j	查询者为用户 j 设定的查询范围的集合
$\boldsymbol{T}_{Q_{ji}}$	用户 j 的第 i 个属性对应的安全查询向量
T_{Q_j}	用户 j 所有属性对应的安全查询向量的集合
$\boldsymbol{I}_{i_k}$	用户 i 的第 k 个属性对应的安全索引向量
I_i	用户 i 所有属性对应的安全索引向量的集合
k	Paillier 加密算法安全参数
L	用于 CIPE 的随机比特串
d	比特串 L 的长度（d 为偶数）
$\boldsymbol{p}_{i_k}$	用户 i 的第 k 个属性由 CIPE 所得向量
$\hat{\boldsymbol{p}}_{i_k}$	使用 Paillier 对 $\boldsymbol{p}_{i_k}$ 中每个元素加密后所得向量
$\boldsymbol{q}_{j_{k,\phi}}$	用户 j 的第 k 个属性所对应查询范围的 CIPE 向量
$\hat{\boldsymbol{q}}_{j_{k,\phi}}$	使用 Paillier 对 $\boldsymbol{q}_{j_{k,\phi}}$ 中每个元素加密后所得向量

3.2 威胁模型

在 P3M 方案中，假设所有的实体均是诚实且好奇的实体[2]，即能够诚实地执行方案流程，但也

可能会试图收集和窃取隐私信息；方案中的用户均不相信其他任何实体，但会诚实地执行所有指令，且不会通过更改属性向量来探测明密文的对应关系；与文献[25-26]类似，P3M 方案假设 2 个服务器不会合谋，但是用户可能会和其中一个服务器合谋；假设 P3M 方案中的通信是完全可信的，因为现有的安全方案和协议如安全套接层（SSL, secure socket layer）和安全外壳（SSH, secure shell）已经可以保证信道中通信数据的安全，因此，类似于文献[27-28]的安全假设，本文威胁模型主要关注来自实体的攻击，例如，不完全可信服务器企图获取或者推导用户隐私数据，恶意用户企图窃取其他特定用户的敏感信息，服务器和用户的合谋攻击等。

3.3　基于 CIPE 的属性值范围判断

CIPE 算法能够保证编码前后明文之间的大小不会改变，进而实现对数据大小的比较。因此，利用 CIPE 算法能够判断属性值是否在查询范围内。以位置范围的判断为例（如图 2 所示），假设查询者 Alice 当前的位置为 $\left(E_{\mathrm{CIPE}}(\mathrm{lon}), E_{\mathrm{CIPE}}(\mathrm{lat})\right)$，其中，$E_{\mathrm{CIPE}}(m)$ 表示采用 CIPE 算法对明文 m 进行编码，lon 和 lat 分别表示查询者 Alice 所处的经度和纬度，且数值均已处理为整数。根据查询者设定的半径为 2 km 的查询范围，生成对应的经纬度范围（用圆形范围外接矩形代替）分别表示为 $\left[E_{\mathrm{CIPE}}(\mathrm{lon_l}), E_{\mathrm{CIPE}}(\mathrm{lon_u})\right]$ 和 $\left[E_{\mathrm{CIPE}}(\mathrm{lat_l}), E_{\mathrm{CIPE}}(\mathrm{lat_u})\right]$，其中下标 l,u 代表查询范围的下界和上界。如果用户 Bob 的位置 $\left(E_{\mathrm{CIPE}}(\mathrm{lon}'), E_{\mathrm{CIPE}}(\mathrm{lat}')\right)$ 满足条件 $E_{\mathrm{CIPE}}(\mathrm{lon}')E_{\mathrm{CIPE}}(\mathrm{lon_l})$ 且 $E_{\mathrm{CIPE}}(\mathrm{lon}')E_{\mathrm{CIPE}}(\mathrm{lon_u})>0$（经度范围的判断方法类似），则说明用户 Bob 位置处于 Alice 设定的范围内。

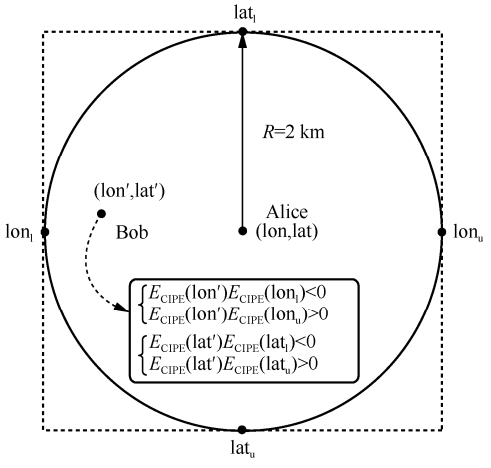

图 2　位置范围示意

在 P3M 方案中，用户和查询者分别采用 CIPE 算法编码属性值和查询值，得到对应的向量 $\boldsymbol{p}$ 和 $\boldsymbol{q}$，随后使用公钥加密算法对 $\boldsymbol{p}$ 和 $\boldsymbol{q}$ 中的每一个元素进行加密，加密后的向量记作 $\hat{\boldsymbol{p}}$ 和 $\hat{\boldsymbol{q}}$，为了计算它们的点积并保证安全性，本文设计了一种安全点积协议（协议 1），使用 Paillier 加密算法和服务器之间的交互求得 2 个向量的点积并保证安全性。

3.4　安全点积协议

Alice 和 Bob 将 $\hat{\boldsymbol{q}}$ 和 $\hat{\boldsymbol{p}}$ 上传至服务器后，SA 和 SB 通过执行安全点积协议（见协议 1）来计算它们的点积，以便通过该值判断属性值是否在查询范围内。假设 $E_{\mathrm{pk_{SB}}}(q_i)$ 和 $E_{\mathrm{pk_{SB}}}(p_i)$ 分别为来自 $\hat{\boldsymbol{q}}$ 和 $\hat{\boldsymbol{p}}$ 中的元素，它们均采用 SB 的公钥加密。根据 Paillier 加密算法的加法同态性，SA 在不需要解密密文的情况下即可计算 q_i 和 p_i 间差值所对应的密文 $E_{\mathrm{pk_{SB}}}(q_i-p_i)$，即执行 $E_{\mathrm{pk_{SB}}}(q_i)\left(E_{\mathrm{pk_{SB}}}(p_i)\right)^{-1}$。

假设 $\hat{\boldsymbol{q}}$ 和 $\hat{\boldsymbol{p}}$ 的长度为 d，SA 则需要一次性计算 d 组差值 $E_{\mathrm{pk_{SB}}}(q_i-p_i), i\in[1,d]$ 并发送给 SB。SB 收到后使用自己的私钥进行解密得到 d 组差值 q_i-p_i。值得注意的是，在 SA 和 SB 互不合谋的假设下，即使 SB 获得了这些明文数据也无法得知这些数据与特定用户之间的关联关系，保证了用户隐私的安全。随后，SB 计算这些数据的平方和 $\sum_{i=1}^{d}(q_i-p_i)^2$ 并使用自己的公钥加密发送给 SA。

对于向量 $\boldsymbol{p}$ 和 $\boldsymbol{q}$，它们的点积可以写为

$$\boldsymbol{p}\cdot\boldsymbol{q}=\sum_{i=1}^{n}p_iq_i=\frac{1}{2}\sum_{i=1}^{n}p_i^{\,2}+\sum_{i=1}^{n}q_i^{\,2}-\sum_{i=1}^{n}(q_i-p_i)^2 \quad (1)$$

因此，SA 在收到 SB 发送的密文之后即可利用 Paillier 的同态性并根据式(1)计算 $E_{\mathrm{pk_{SB}}}(2\boldsymbol{p}\cdot\boldsymbol{q})$，进而计算 $E_{\mathrm{pk_{SB}}}(\boldsymbol{p}\cdot\boldsymbol{q})$。最终 SB 将点积结果映射到 $g'\in\{1,0,-1\}$（g' 的符号和 $\boldsymbol{p}\cdot\boldsymbol{q}$ 的符号相同）并发送给 SA。协议 1 具体细节如下。

协议 1　安全点积协议

输入　SA 的密钥对 $(\mathrm{pk_{SA}},\mathrm{sk_{SA}})$；SB 的密钥对 $(\mathrm{pk_{SB}},\mathrm{sk_{SB}})$；向量 $\hat{\boldsymbol{p}}$ 和 $\hat{\boldsymbol{q}}$ 中的元素 $E_{\mathrm{pk_{SB}}}(p_i)$、$E_{\mathrm{pk_{SB}}}(q_i)$ 以及 $E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{d}p_i^2\right)$、$E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{d}q_i^2\right)$；Alice 执行 CIPE 所使用的随机数组元素之和的密文

$E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{\frac{d}{4}} c_i\right)$；Bob 执行 CIPE 算法所使用的随机数组元素之和的密文 $E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{\frac{d}{4}} \overline{c_i}\right)$

输出 $g' \in \{1, 0, -1\}$

1) SA 计算 $E_{\mathrm{pk_{SB}}}(q_i)\left(E_{\mathrm{pk_{SB}}}(p_i)\right)^{-1} = E_{\mathrm{pk_{SB}}}(q_i - p_i)$ 并将 $E_{\mathrm{pk_{SB}}}(q_i - p_i)$ 发送给 SB。

2) SB 使用私钥 $\mathrm{sk_{SB}}$ 解密得到 $q_i - p_i$ 并计算 $\sum_{i=1}^{d}(q_i - p_i)^2$，使用公钥 $\mathrm{pk_{SB}}$ 执行加密操作得到 $E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{d}(q_i - p_i)^2\right)$ 并将密文发送给 SA。

3) SA 执行如下操作。

① 根据式(1)计算

$$E_{\mathrm{pk_{SB}}}(2\boldsymbol{p}\cdot\boldsymbol{q}) = E_{\mathrm{pk_{SB}}}\left(\sum_{k=1}^{d} p_i^2\right) E_{\mathrm{pk_{SB}}}\left(\sum_{k=1}^{d} q_i^2\right) E_{\mathrm{pk_{SB}}}\left(\sum_{k=1}^{d}(q_i - p_i)^2\right)^{-1} \tag{2}$$

② 计算 $E_{\mathrm{pk_{SB}}}(\boldsymbol{p}\cdot\boldsymbol{q}) = E_{\mathrm{pk_{SB}}}(2\boldsymbol{p}\cdot\boldsymbol{q})^{2^{-1}}$ 得到向量点积的密文。

③ 计算随机数组元素之和

$$E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{\frac{d}{4}} c_i + \overline{c_i}\right) = E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{\frac{d}{4}} c_i\right) E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{\frac{d}{4}} \overline{c_i}\right) \tag{3}$$

④ 将 $\left\langle E_{\mathrm{pk_{SB}}}(\boldsymbol{p}\cdot\boldsymbol{q}) \| E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{\frac{d}{4}} c_i + \overline{c_i}\right)\right\rangle$ 发送给 SB。

4) SB 执行如下操作。

① 解密得到 $\boldsymbol{p}\cdot\boldsymbol{q}$ 并计算

$$g = (\boldsymbol{p}\cdot\boldsymbol{q})\left(\sum_{i=1}^{\frac{d}{4}} c_i + \sum_{i=1}^{\frac{d}{4}} \overline{c_i}\right)^{-1} \tag{4}$$

② 令 $g' = \begin{cases} -1, & g \in \left(0, \dfrac{(p+1)}{2}\right) \\ 0, & g = 0 \\ 1, & g \in \left(\dfrac{(p-1)}{2}, p\right) \end{cases}$，并将 g' 发送至 SA。

5) SA 输出 g'。

3.5 方案流程

在 P3M 方案中，第 i 位用户的属性集合记作 $\mathcal{A}_i = \{\mathrm{lon}, \mathrm{lat}, a_{i_1}, a_{i_2}, \cdots, a_{i_m}\}$，其中，前 2 个属性分别代表用户位置的经度和纬度信息，经度范围为 73.550 00°～135.083 33°，纬度范围为 3.850 00°～53.550 00°，为方便后续方案的执行，需对经纬度的数值乘以 10^5 转换为整数值，因此用户属性向量中经纬度属性值的范围为 $\mathrm{lon} \in [7\,355\,000°, 13\,508\,333°]$，$\mathrm{lat} \in [385\,000°, 5\,355\,000°]$；其余属性值 $a_{i_k}, k \in (1, \cdots, m)$ 为用户个人属性（如年龄、性别等）的值，所有用户的属性集合构成用户数据总集合 $\mathcal{D} = \{\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n\}$。

P3M 方案主要分为 4 个阶段，分别为密钥初始化（Step1）、生成安全索引向量（Step2）、生成安全查询向量（Step3）、相似度比较（Step4），方案流程如图 3 所示。用户通过执行 Step2，将每一个属性值加密成安全索引向量 $\boldsymbol{I}_{i_k}$，所有属性对应的安全索引向量的集合记作 $\boldsymbol{I}_i = \{\boldsymbol{I}_{i_{\mathrm{lon}}}, \boldsymbol{I}_{i_{\mathrm{lat}}}, \boldsymbol{I}_{i_1}, \boldsymbol{I}_{i_2}, \cdots, \boldsymbol{I}_{i_m}\}$。查询者生成对第 j 位用户的第 i 个属性的查询范围为 $Q_{j_i} = [b_{j_{i,1}}, b_{j_{i,\mathrm{u}}}]$，其中 $b_{j_{i,\phi}}, \phi \in \{\mathrm{l}, \mathrm{u}\}$ 分别代表查询范围的下界和上界。用户 j 所有的属性生成的查询范围的集合记作 $Q_j = \{Q_{j_{\mathrm{lon}}}, Q_{j_{\mathrm{lat}}}, Q_{j_1}, Q_{j_2}, \cdots, Q_{j_m}\}$。查询者执行 Step3 生成安全查询向量 $\boldsymbol{T}_{Q_{j_i}}$，Q_j 对应的安全查询向量集合为 $T_{Q_j} = \{\boldsymbol{T}_{Q_{j_{\mathrm{lon}}}}, \boldsymbol{T}_{Q_{j_{\mathrm{lat}}}}, \boldsymbol{T}_{Q_{j_1}}, \boldsymbol{T}_{Q_{j_2}}, \cdots, \boldsymbol{T}_{Q_{j_m}}\}$。最终服务器之间执行 Step4 实现用户间属性相似度的计算。

Step1 密钥初始化

1) 所有用户和服务器生成自己的公私钥对 $(\mathrm{pk}, \mathrm{sk})$。将 Alice 和 Bob 的密钥对分别记为 $(\mathrm{pk_A}, \mathrm{sk_A})$ 和 $(\mathrm{pk_B}, \mathrm{sk_B})$，SA 和 SB 的密钥对分别为 $(\mathrm{pk_{SA}}, \mathrm{sk_{SA}})$ 和 $(\mathrm{pk_{SB}}, \mathrm{sk_{SB}})$。

2) SA 随机选取 2 个长度为 $\frac{d}{2}$ 的比特串 L_1, L_2，将它们首尾相连成长度为 d 的串 L。使用用户公钥加密后发送给用户。

3) 每个用户生成 CIPE 所需的随机数组，计算数组元素之和并加密。分别将加密所得的

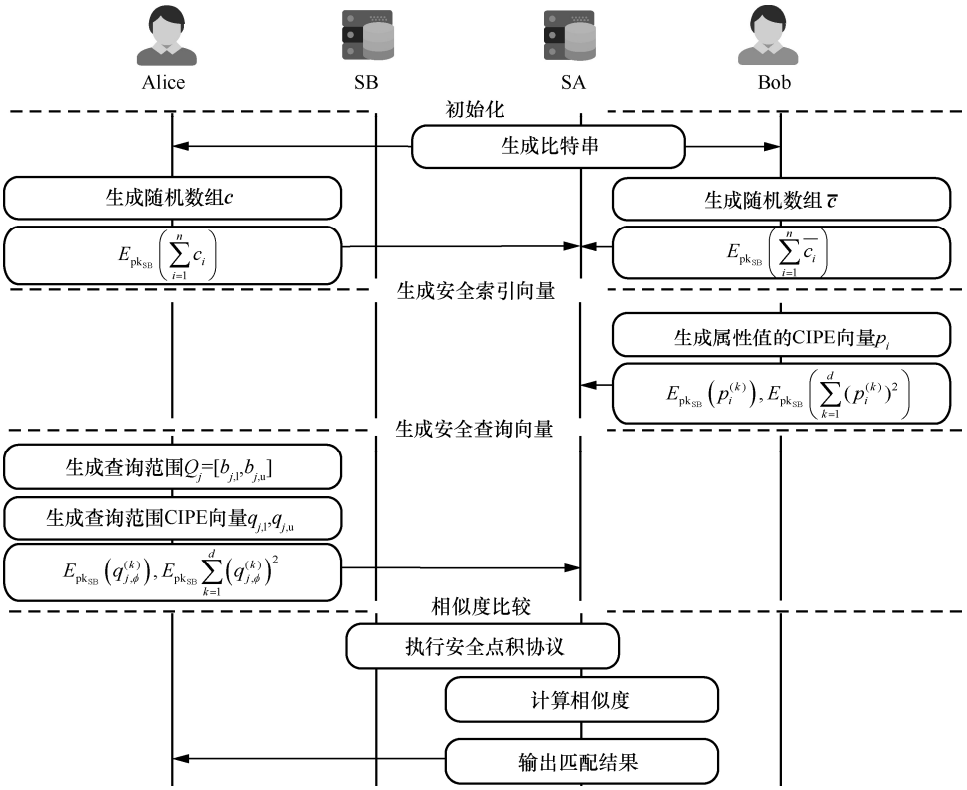

图 3　P3M 方案流程

$E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{\frac{d}{4}} c_i\right)$ 和 $E_{\mathrm{pk_{SB}}}\left(\sum_{i=1}^{\frac{d}{4}} \bar{c}_i\right)$ 发送至 SA。

Step2　生成安全索引向量 $\mathrm{SecIndex}(\mathcal{A}_{\mathrm{Bob}}, L)$

1) 对于每个 $a_i \in \mathcal{A}_{\mathrm{Bob}}$，Bob 执行算法 1，将其转换为 d 维向量 $\boldsymbol{p}_i$。

2) Bob 使用服务器 B 的公钥加密 $p_i^{(k)} \in \boldsymbol{p}_i$，记作 $\hat{\boldsymbol{p}}_i = \left\{k \in [1,d] \mid E_{\mathrm{pk_{SB}}}\left(p_i^{(k)}\right)\right\}$，并计算 $\sum_{k=1}^{d}\left(p_i^{(k)}\right)^2$。

3) 构造安全索引向量集合 $I_{\mathrm{Bob}} = \{\boldsymbol{I}_{\mathrm{lon}}, \boldsymbol{I}_{\mathrm{lat}}, \boldsymbol{I}_1, \boldsymbol{I}_2, \cdots, \boldsymbol{I}_m\}$，其中 $\boldsymbol{I}_i = \left\langle \hat{\boldsymbol{p}}_i \| E_{\mathrm{pk_{SB}}}\left(\sum_{k=1}^{d}\left(p_i^{(k)}\right)^2\right)\right\rangle$。

4) 将 I_{Bob} 发送给 SA。

Step3　生成安全查询向量 $\mathrm{SecQuery}\left(Q_j, L\right)$

1) Alice 构造对 Bob 属性集 $\mathcal{A}_{\mathrm{Bob}}$ 的查询范围的集合 $Q_{\mathrm{Alice}} = \{Q_{\mathrm{lon}}, Q_{\mathrm{lat}}, Q_1, Q_2, \cdots, Q_m\}$，其中 $Q_j = \left[b_{j,1}, b_{j,\mathrm{u}}\right]$（$b_{j,1}$ 和 $b_{j,\mathrm{u}}$ 分别为查询范围的下界和上界）。对于每一个 $Q_j \in Q_{\mathrm{Alice}}$，$j \in \{\mathrm{lon}, \mathrm{lat}, 1, \cdots, m\}$，Alice 运行算法 2，将 $b_{j,\phi}$ 转换为 d 维向量 $\boldsymbol{q}_{j,\phi}$，$\phi \in \{\mathrm{l}, \mathrm{u}\}$。

2) 对于 $q_{j,\phi}^{(k)} \in \boldsymbol{q}_{j,\phi}$，$\phi \in \{\mathrm{l}, \mathrm{u}\}$，Alice 使用 SB 的公钥加密 $q_{j,\phi}^{(k)}$，记作 $\hat{\boldsymbol{q}}_{j,\phi} = \left\{k \in [1,d] \mid E_{\mathrm{pk_{SB}}}\left(q_{j,\phi}^{(k)}\right)\right\}$ 并计算 $\sum_{k=1}^{d}\left(q_{j,\phi}^{(k)}\right)^2$。

3) 构造安全查询向量集合 $T_{Q_{\mathrm{Alice}}} = \{\boldsymbol{T}_{Q_{\mathrm{lon}}}, \boldsymbol{T}_{Q_{\mathrm{lat}}}, \boldsymbol{T}_{Q_1}, \boldsymbol{T}_{Q_2}, \cdots, \boldsymbol{T}_{Q_m}\}$，其中 $\boldsymbol{T}_{Q_j} = \left\langle \hat{\boldsymbol{q}}_{j,\mathrm{l}} \| \hat{\boldsymbol{q}}_{j,\mathrm{u}} \| E_{\mathrm{pk_{SB}}}\left(\sum_{k=1}^{d}\left(q_{j,\mathrm{l}}^{(k)}\right)^2\right) \| E_{\mathrm{pk_{SB}}}\left(\sum_{k=1}^{d}\left(q_{j,\mathrm{u}}^{(k)}\right)^2\right)\right\rangle$，$j \in \{\mathrm{lon}, \mathrm{lat}, 1, \cdots, m\}$。

4) Alice 生成权重向量 $\boldsymbol{w} = (w_1, w_2, \cdots, w_m)$（位置信息不分配权值）和相似度阈值 δ（相似度高于 δ 视为匹配成功）。

5) 将 $E_{\mathrm{pk_{SA}}}\left(T_{Q_{\mathrm{Alice}}} \| \boldsymbol{w} \| \delta\right)$ 发送给 SA。

Step4　相似度比较 $\mathrm{Compare}\left(I_{\mathrm{Bob}}, T_{Q_{\mathrm{Alice}}}, w, \delta\right)$

在此阶段，SA 先对用户的地理位置进行筛选，若用户的地理位置不在查询者给定的范围内，则跳过此用户，不再对其进行后续的属性相似度比较。具体步骤如下。

1) 对于当前用户位置的安全索引向量 $\boldsymbol{I}_\phi$ 和对应的安全查询向量 $\boldsymbol{T}_{Q_\phi}$，SA 和 SB 执行安全点积协议计算 $v_{\phi,\mathrm{l}} = \boldsymbol{p} \cdot \boldsymbol{q}_{\phi,\mathrm{l}}$ 和 $v_{\phi,\mathrm{u}} = \boldsymbol{p} \cdot \boldsymbol{q}_{\phi,\mathrm{u}}$，其中 $\phi \in \{\mathrm{lon}, \mathrm{lat}\}$。

2) 若 $v_{\phi,\mathrm{l}} \leqslant 0$ 且 $v_{\phi,\mathrm{u}} \geqslant 0, \phi \in \{\mathrm{lon}, \mathrm{lat}\}$，继续执

行后续步骤，否则重新选择一个新的用户并回到步骤 1)。

3) 对于 $\boldsymbol{I}_k \in I_{\text{Bob}}, \boldsymbol{T}_{Q_k} \in T_{Q_{\text{Alice}}}, k \in \{1,\cdots,m\}$，执行安全点积协议计算 $v_{k,1} = \boldsymbol{p} \cdot \boldsymbol{q}_{k,1}$ 和 $v_{k,\text{u}} = \boldsymbol{p} \cdot \boldsymbol{q}_{k,\text{u}}$。

4) 生成 m 维的比较结果向量 $\boldsymbol{r} = (r_1, r_2, \cdots, r_m)$，其中 $r_k = \begin{cases} 1, & v_{k,1} \leqslant 0 \text{且} v_{k,\text{u}} \geqslant 0 \\ 0, & \text{其他} \end{cases}$，$k \in [m]$。

5) 计算相似度 $s = \sum_{k=1}^{m} w_k r_k$。

6) 输出匹配结果 $\begin{cases} 1(\text{成功}), & s \geqslant \delta \\ 0(\text{失败}), & \text{其他} \end{cases}$。

4 正确性和安全性

4.1 正确性

在 P3M 方案中，服务器接收到的密文有 $E_{\text{pk}_{\text{SB}}}(p_i)$、$E_{\text{pk}_{\text{SB}}}(q_i)$、$E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} p_i^2\right)$ 和 $E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} q_i^2\right)$，服务器基于 Paillier 加密的同态性可以计算出 $E_{\text{pk}_{\text{SB}}}(q_i - p_i)^2$，服务器 B 可以使用自己的私钥将其解密并计算生成 $E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} (q_i - p_i)^2\right)$，服务器 A 可以通过式(5)得到 $E_{\text{pk}_{\text{SB}}}(\boldsymbol{p} \cdot \boldsymbol{q})$。

$$\left(E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} p_i^2\right) E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} q_i^2\right)\left(E_{\text{pk}_{\text{SB}}}(q_i - p_i)^2\right)^{-1}\right)^{2^{-1}} =$$

$$\left(E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d}\left(p_i^2 + q_i^2 - (q_i - p_i)^2\right)\right)\right)^{2^{-1}} =$$

$$\left(E_{\text{pk}_{\text{SB}}}\left(2\sum_{i=1}^{d} p_i \cdot q_i\right)\right)^{2^{-1}} =$$

$$\left(E_{\text{pk}_{\text{SB}}}(2\boldsymbol{p} \cdot \boldsymbol{q})\right)^{2^{-1}} = E_{\text{pk}_{\text{SB}}}(\boldsymbol{p} \cdot \boldsymbol{q}) \tag{5}$$

4.2 安全性

本文采用与文献[2]类似的方法，即“理想/现实模式”[29]来证明 P3M 方案的安全性。“理想”和“现实”分别指方案在理想环境（服务器完全可信）和现实环境（服务器不可信）中执行。如果在现实环境中，攻击者对协议进行攻击所获得的信息不会多于在理想环境中攻击者对协议进行攻击所获得的信息，那么协议就是安全的。因此，对于任意一个攻击者在现实环境中对协议发动的攻击，总存在一个攻击者在理想环境中对协议发起的攻击，它们所能够获得的信息是相同的。

证明 P3M 方案包含 4 个实体，分别是 Alice、Bob、服务器 A、服务器 B，构造分别代表 4 个实体的模拟器 $\text{Sim} = (\text{Sim}_{\text{A}}, \text{Sim}_{\text{B}}, \text{Sim}_{\text{SA}}, \text{Sim}_{\text{SB}})$，构造与每个实体合谋的攻击者 $(\mathcal{A}_{\text{A}}, \mathcal{A}_{\text{B}}, \mathcal{A}_{\text{SA}}, \mathcal{A}_{\text{SB}})$。

讨论 Sim_{A} 和 $\mathcal{A}_{\text{A}}$ 的情况，对于输入 $\boldsymbol{q} = \{q_1, q_2, \cdots, q_n\}$，$\text{Sim}_{\text{A}}$ 使用服务器 B 的公钥对 q_i 进行数据加密得到 $E_{\text{pk}_{\text{SB}}}(q_i)$，计算 $\sum_{i=1}^{d} q_i^2$ 并加密 $E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} q_i^2\right)$。接下来，随机选择向量 $\boldsymbol{p} = \{p_1, p_2, \cdots, p_n\}$，计算 $\boldsymbol{p} \cdot \boldsymbol{q}$ 并将它的值转换为 $g \in \{-1, 0, 1\}$，g 的 3 个值分别代表 $\boldsymbol{p} \cdot \boldsymbol{q}$ 的值小于 0、等于 0 和大于 0。最终将 g 发送给 $\mathcal{A}_{\text{A}}$，在整个过程中 $\mathcal{A}_{\text{A}}$ 能够获得的信息包括输入 $\boldsymbol{q}$ 中的每个值 q_i、所有的密文 $E_{\text{pk}_{\text{SB}}}(q_i)$、$E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} q_i^2\right)$ 以及结果 g。由于 g 只能表示出 $\boldsymbol{p} \cdot \boldsymbol{q}$ 值的大小，因此无法通过计算还原出 $\boldsymbol{p}$。根据 Paillier 加密算法的安全性可知，在 $\mathcal{A}_{\text{A}}$ 视角下的密文均是不可分辨的。

讨论 Sim_{B} 和 $\mathcal{A}_{\text{B}}$ 的情况，对于输入 $\boldsymbol{p} = \{p_1, p_2, \cdots, p_n\}$，$\text{Sim}_{\text{B}}$ 使用服务器 B 的公钥对 p_i 进行数据加密得到 $E_{\text{pk}_{\text{SB}}}(p_i)$，计算 $\sum_{i=1}^{d} p_i^2$ 并加密 $E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} p_i^2\right)$。根据 Paillier 加密算法的安全性可知，在 $\mathcal{A}_{\text{B}}$ 的视角下的密文均是不可分辨的。

讨论 Sim_{SA} 和 $\mathcal{A}_{\text{SA}}$ 的情况，在 P3M 方案工作的整个过程中，Sim_{SA} 接收到的信息有 $E_{\text{pk}_{\text{SB}}}(p_i)$、$E_{\text{pk}_{\text{SB}}}(q_i)$、$E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} p_i^2\right)$、$E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} q_i^2\right)$ 以及服务器 B 计算出的 $E_{\text{pk}_{\text{SB}}}\left(\sum_{i=1}^{d} (q_i - p_i)^2\right)$，$\text{Sim}_{\text{SA}}$ 计算出 $E_{\text{pk}_{\text{SB}}}(\boldsymbol{p} \cdot \boldsymbol{q})$，最终服务器 B 将匹配的结果发送给 Sim_{SA}。整个过程中的密文均是使用服务器 B 的公钥加密的，在假设服务器之间不能合谋的前提下，根据 Paillier 加密算法的安全性可知，在 $\mathcal{A}_{\text{SA}}$ 的视角下的密文是不可分辨的，并且在方案执行结束之后 $\mathcal{A}_{\text{SA}}$ 只能够得知 2 个用户是否匹配，并不知道具体的 $\boldsymbol{p} \cdot \boldsymbol{q}$ 点积结果。

讨论 Sim_{SB} 和 $\mathcal{A}_{\text{SB}}$ 的情况，在 P3M 方案工作的

整个过程中，$\mathrm{Sim}_{\mathrm{SB}}$ 接收到的所有信息为 $q_i - p_i$ 和 $\boldsymbol{p}\cdot\boldsymbol{q}$，在假设服务器之间不能合谋的前提下，$\mathcal{A}_{\mathrm{SB}}$ 无法将解密得到的明文与具体用户相对应。

5　实验分析

本文采用苹果 Macbook Pro 搭建实验环境并进行仿真测试，利用 Java 作为编程语言进行代码的编写，使用 Intel Core i7 六核处理器，主频为 2.6 GHz，使用 16 GB 2 667 MHz DDR4 内存和 AMD Radeon Pro 5300M 显卡。

为了模拟 P3M 方案在实际应用中的计算开销，实验测试了属性个数从 20 递增到 100 的情况下，P3M 方案在 Paillier 密钥长度 k= 256 bit 和 k=512 bit 以及 CIPE 比特串长度 d=64 bit 和 d=128 bit 条件下的执行时间。实验中的用户位置坐标采用随机坐标生成器从中国地图选择 20 km ×20 km 的范围生成 10 000 个用户坐标，用户属性值以及属性查询范围均为随机生成，并且根据不同的属性设置不同的取值范围，例如，年龄范围为 0～100 岁，身高范围为 0～200 cm，对某一事物的喜好程度范围均设为 0～10。

P3M 在不同用户属性个数和密钥长度下执行时间的对比如图 4 所示。由图 4 可知，相似度比较时间随着属性个数的增加而不断增加，密钥初始化的时间趋势则比较平稳，说明 P3M 方案初始化几乎不受属性个数增多或减少的影响。由于每个属性所对应的查询向量都需要加密其上界和下界的值，因此理论上生成安全查询向量所需时间为生成安全索引向量的 2 倍，这在图 4 中得以验证。对于用户而言，大部分属性值的加密在用户注册时就可以完成，后续对属性值的增减删改仅作用于部分属性，因此用户在日常使用过程中所能感知到的属性加密的时间远小于实验数据，且好友

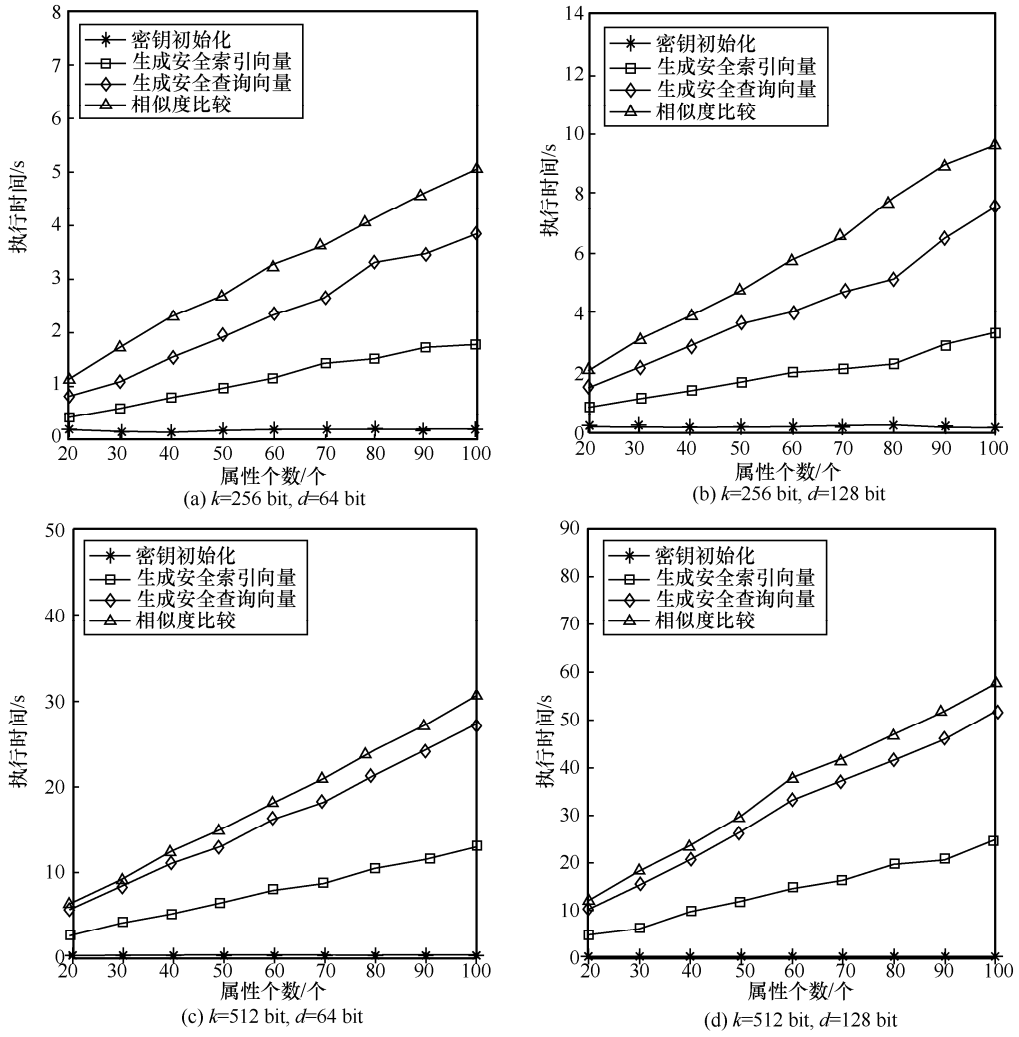

图 4　P3M 在不同用户属性个数和密钥长度下执行时间的对比

匹配过程是在服务器中进行的，鉴于服务器具有较强的计算能力，好友匹配时间相较于实验数据还会进一步缩减。

在 P3M 方案中，网络通信量对方案执行的性能影响也是不可忽视的。P3M 在不同用户属性个数和密钥长度下通信开销的对比如图 5 所示。由图 5 可知，随着密钥长度和属性个数的增加，网络通信量也随之增加，但是即使是如图 5(d)所示的情况，用户所上传到服务器的数据也不超过 8 MB，在 4G 网络下传输仅需 3～6 s 即可上传至服务器，由于用户在注册以及日常使用过程中会提前上传部分属性值，因此用户感知到的文件上传时间将远小于 3～6 s，并且随着 5G 的普及，传输效率将更加高效。在实际应用场景中，Paillier 密钥长度为 512 bit 时能保证较高的安全性，但对应的匹配时间会随之增加；密钥长度为 256 bit 时能够在保证安全性的同时拥有更高的匹配效率。图 6 给出了属性个数固定为 20，不同 CIPE 比特串长度下执行时间的对比。从图 6 中可以看出，初始化时间主要受 Paillier 密钥长度的影响并随着 CIPE 比特串长度不断增加，密钥初始化、生成安全索引向量、生成安全查询向量和相似度比较的执行时间差距逐渐增大。例如，图 6(a)中，生成安全索引向量、生成安全查询向量 2 个过程的执行时间差由 d=16 bit 时的 0.115 s 增加到 d=64 bit 时的 0.379 s。

在 P3M 中还实现了对查询范围的匹配，查询者在递交查询之后，服务器先对目标用户的位置进行筛查，仅当用户的位置符合查询范围时才会继续对用户的属性进行匹配。图 7 和图 8 给出了筛查过程的执行时间，整个测试在 20 km×20 km 范围的地图中随机生成用户的坐标，用户数量为

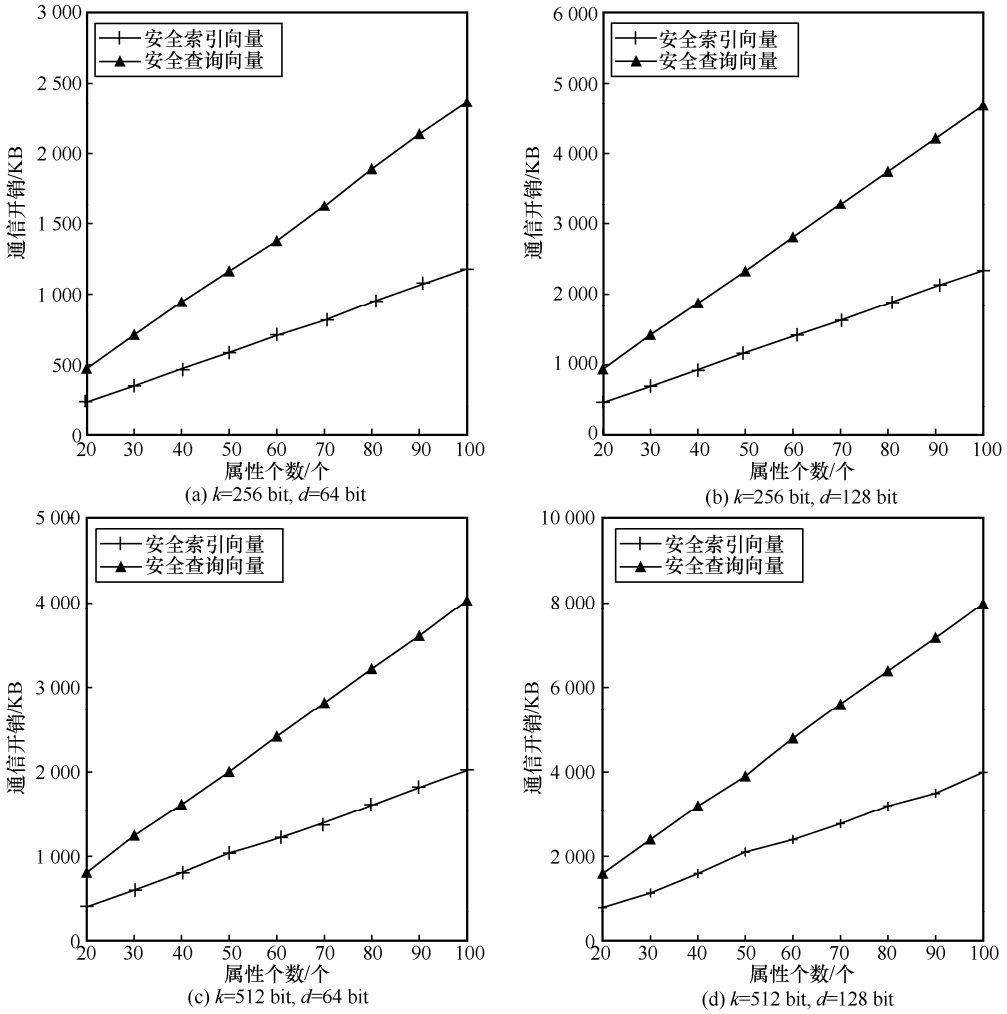

图 5　P3M 在不同用户属性个数和密钥长度下通信开销的对比

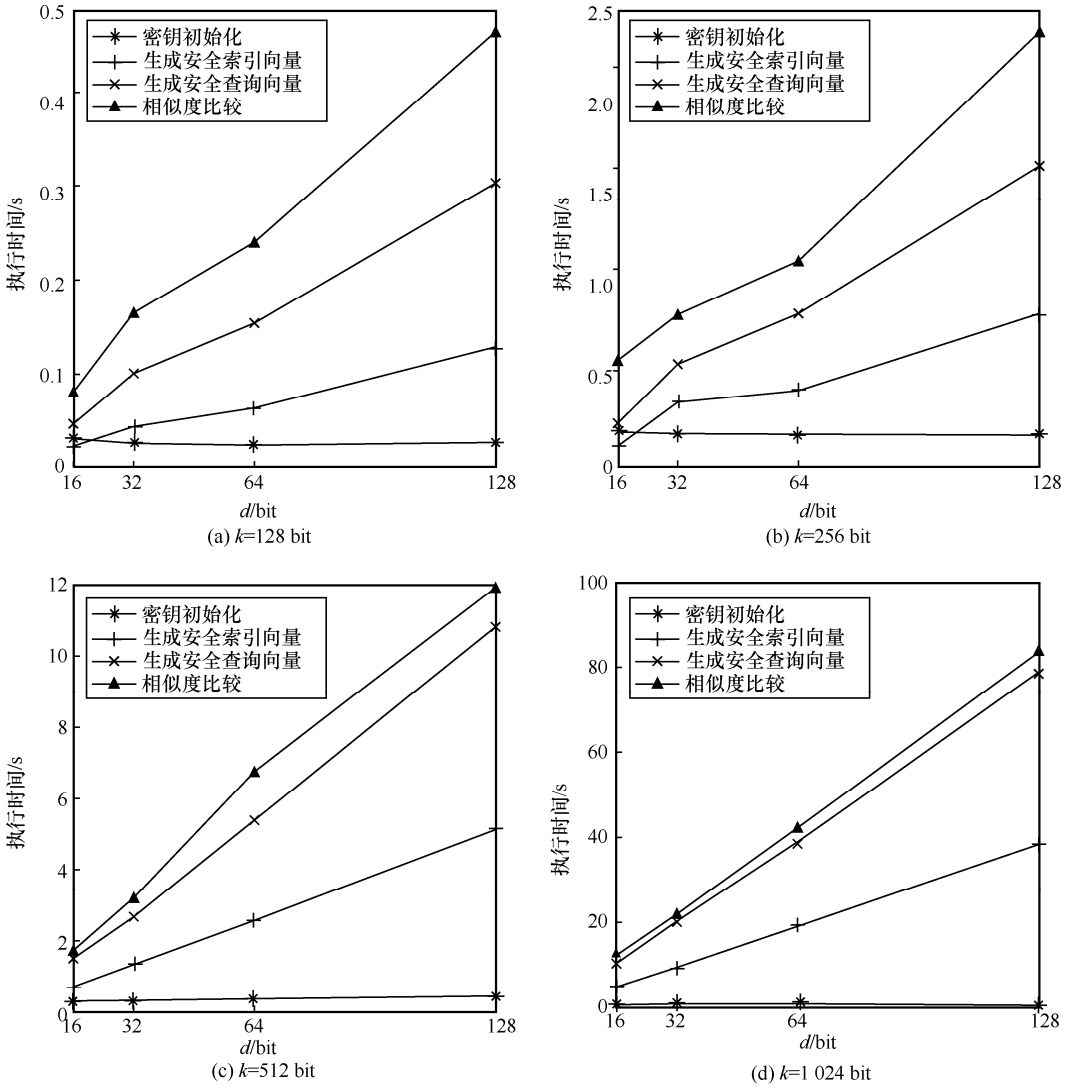

图 6　P3M 在不同 CIPE 比特串长度下执行时间的对比（属性个数=20）

2 000～10 000 个，查询范围为 2～16 km，d=64 bit，k=128 bit。测试中所计算的方案执行时间为 SA、SB 执行方案中的相似度比较（Step4）中步骤 1)、步骤 2)所用时间，即服务器会不断从整个地图区域随机选择用户，判断其位置是否满足查询范围，直到找到一个符合位置范围的用户。从图 7 中可以看出，位置匹配时间受用户数量大小的影响不大，而与查询者设定的查询范围有关。由于服务器选择用户是随机的，查询者设定的查询范围越大，位于查询者设定范围内的用户就越多，进而这些用户被服务器选中的概率就越大，这也就使服务器更容易选择到一个符合查询范围的用户。从图 8 中可以看出，随着查询范围的扩大，服务器找到符合位置范围用户所需时间也逐渐缩短，实验结果也与预期相符合。

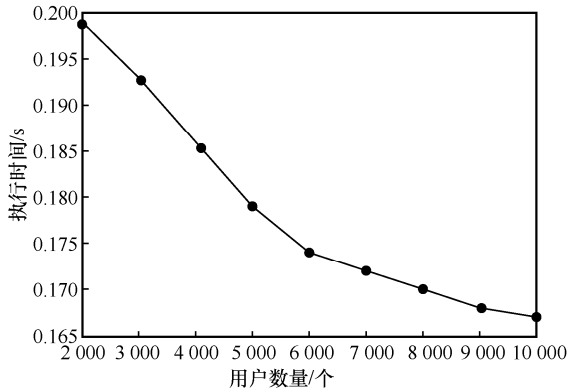

图 7　P3M 在不同用户数量下位置匹配过程执行时间

表 2 展示了 P3M 与现有采用中心化架构的方案的性能对比。从表 2 中可以看出，P3M 在避免用户参与计算和提供细粒度好友匹配的基础上实现了查询者对

表 2 P3M 与现有采用中心化架构的方案的性能对比

方案	中心化架构	匹配过程不需要用户参与	细粒度匹配	设定匹配范围	考虑位置范围
文献[2]方案	√	√	√	×	×
文献[4]方案	√	√	×	×	×
文献[9]方案	√	×	×	×	√
文献[10]方案	√	×	×	×	×
文献[11]方案	√	√	×	×	×
P3M 方案	√	√	√	√	√

用户属性值和位置范围的个性化匹配，相较于其他方案，P3M 方案实现了特征属性及位置属性等多种属性的精确匹配，并综合考虑了用户多维度的隐私保护。

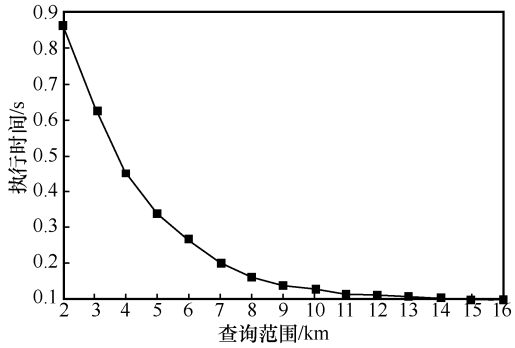

图 8 P3M 在不同位置查询范围下位置匹配过程执行时间

6 结束语

本文提出了面向隐私保护的精确好友匹配方案，相较于现有方案，P3M 方案在保证用户隐私的前提下同时考虑了用户属性和位置的相似度，并且能够支持查询者定义查询的范围以获得更精确的结果。在 P3M 中，利用 CIPE 算法编码前后数据大小关系不变的特性，实现了对用户加密的属性值所在范围的判断，解决了基于用户属性向量点积的相似度计算方法存在的匹配结果不准确的问题。同时，该方案增加了对用户和查询者之间距离的考虑，使查询者能够灵活设定在指定位置范围匹配相似的用户，实现了更细粒度、更精确的好友匹配；同时，P3M 采用了双服务器的系统框架，解决了在半可信服务器下安全计算用户间的相似度的问题，并且能够抵御用户和其中一个服务器之间的合谋攻击；最后，对 P3M 方案执行效率以及通信量进行评估，实验结果表明，在实际情况中 P3M 具有较高的执行效率和较低通信量。但是，P3M 方案在实现更精确、更隐私的好友匹配的同时仍然存在部分问题未解决，用户隐私的保护仍需建立在 2 个服务器不合谋的假设前提下等，这些问题将会在后续的研究中进一步解决。

参考文献：

[1] FREEDMAN M J, HAZAY C, NISSIM K, et al. Efficient set intersection with simulation-based security[J]. Journal of Cryptology, 2016, 29(1): 115-155.

[2] GAO C Z, CHENG Q, LI X, et al. Cloud-assisted privacy-preserving profile-matching scheme under multiple keys in mobile social network[J]. Cluster Computing, 2019, 22(1): 1655-1663.

[3] ATENIESE G, FRANCATI D, NUÑEZ D, et al. Match me if you can: matchmaking encryption and its applications[J]. Journal of Cryptology, 2021, 34(3): 1-50.

[4] YI X, BERTINO E, RAO F Y, et al. Privacy-preserving user profile matching in social networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 32(8): 1572-1585.

[5] LIU Q, WU S X, PEI S Y, et al. Secure and efficient multi-attribute range queries based on comparable inner product encoding[C]// Proceedings of 2018 IEEE Conference on Communications and Network Security. Piscataway: IEEE Press, 2018: 1-9.

[6] AGRAWAL R, EVFIMIEVSKI A, SRIKANT R. Information sharing across private databases[C]//Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2003: 86-97.

[7] VON ARB M, BADER M, KUHN M, et al. VENETA: serverless friend-of-friend detection in mobile social networking[C]//Proceedings of 2008 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications. Piscataway: IEEE Press, 2008: 184-189.

[8] MAHMOUD M, RABIEH K, SHERIF A, et al. Privacy-preserving fine-grained data retrieval schemes for mobile social networks[J]. IEEE Transactions on Dependable and Secure Computing, 2019, 16(5): 871-884.

[9] LUO E T, GUO K H, TANG Y Y, et al. Hidden the true identity and dating characteristics based on quick private matching in mobile social networks[J]. Future Generation Computer Systems, 2020, 109: 633-641.

[10] SHEN H J, ZHOU J, CAO Z F, et al. Blockchain-based lightweight certificate authority for efficient privacy-preserving location-based service in vehicular social networks[J]. IEEE Internet of Things Journal, 2020, 7(7): 6610-6622.

[11] LI J, ZENG F Z, XIAO Z, et al. Drive2friends: inferring social relationships from individual vehicle mobility data[J]. IEEE Internet of Things Journal, 2020, 7(6): 5116-5127.

[12] PENG T, LIU Q, WANG G J, et al. Multidimensional privacy preservation in location-based services[J]. Future Generation Computer Systems, 2019, 93: 312-326.
[13] LIU Q, HOU P L, WANG G J, et al. Intelligent route planning on large road networks with efficiency and privacy[J]. Journal of Parallel and Distributed Computing, 2019, 133: 93-106.
[14] LIU Q, PENG Y, WU J, et al. Secure multi-keyword fuzzy searches with enhanced service quality in cloud computing[J]. IEEE Transactions on Network and Service Management, 2021, 18(2): 2046-2062.
[15] LIU Q, PENG Y, PEI S Y, et al. Prime inner product encoding for effective wildcard-based multi-keyword fuzzy search[J]. IEEE Transactions on Services Computing, 2022, 15(4): 1799-1812.
[16] 李宇溪, 周福才, 徐紫枫. 支持 K-近邻搜索的移动社交网络隐私保护方案[J]. 计算机学报, 2021, 44(7): 1481-1500.
LI Y X, ZHOU F C, XU Z F. Privacy preserving K-nearest-neighbor search over mobile social network[J]. Chinese Journal of Computers, 2021, 44(7): 1481-1500.
[17] 崔炜荣, 杜承烈. 社交网络中基于 CP-ABE 的隐私保护属性匹配方法[J]. 计算机应用, 2018, 38(4): 1051-1057.
CUI W R, DU C L. Privacy preserving attribute matching method based on CP-ABE in social networks[J]. Journal of Computer Applications, 2018,38(4):1051-1057.
[18] LUO E T, LIU Q, ABAWAJY J H, et al. Privacy-preserving multi-hop profile-matching protocol for proximity mobile social networks[J]. Future Generation Computer Systems, 2017, 68: 222-233.
[19] SAMANTHULA B K, ELMEHDWI Y, HOWSER G, et al. A secure data sharing and query processing framework via federation of cloud computing[J]. Information Systems, 2015, 48: 196-212.
[20] XU G Q, LIU B Y, JIAO L T, et al. Trust2Privacy: a novel fuzzy trust-to-privacy mechanism for mobile social networks[J]. IEEE Wireless Communications, 2020, 27(3): 72-78.
[21] LI F H, HE Y Y, NIU B, et al. Small-world: secure friend matching over physical world and social networks[J]. Information Sciences, 2017, 387: 205-220.
[22] ZHANG L, LI X Y, LIU K B, et al. Message in a sealed bottle: privacy preserving friending in mobile social networks[J]. IEEE Transactions on Mobile Computing, 2014, 14(9): 1888-1902.
[23] PENG T, GUAN K, LIU J, et al. A blockchain-based mobile crowdsensing scheme with enhanced privacy[J]. Concurrency and Computation: Practice and Experience, 2021:doi.org/10.1002/cpe.6664.
[24] PENG T, GUAN K, LIU J. A privacy-preserving mobile crowdsensing scheme based on blockchain and trusted execution environment[J]. The IEICE Transactions on Information and Systems, 2021:doi.org/10.1587/transinf.2021BCP0001.
[25] DING X F, WANG Z, ZHOU P, et al. Efficient and privacy-preserving multi-party skyline queries over encrypted data[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 4589-4604.
[26] LIU L, CHEN R M, LIU X M, et al. Towards practical privacy-preserving decision tree training and evaluation in the cloud[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 2914-2929.
[27] PENG T, LIU Q, MENG D, et al. Collaborative trajectory privacy preserving scheme in location-based services[J]. Information Sciences, 2017, 387: 165-179.
[28] LIU Q, TIAN Y, WU J, et al. Enabling verifiable and dynamic ranked search over outsourced data[J]. IEEE Transactions on Services Computing, 2022, 15(1): 69-82.
[29] WANG J. Encyclopedia of data warehousing and mining[M]. Hershey: IGI Global, 2005.

[作者简介]

彭滔（1982− ），女，湖南长沙人，博士，广州大学副教授、硕士生导师，主要研究方向为移动群智感知网、社交网络隐私保护、云计算安全、区块链技术。

钟文韬（1998− ），男，云南曲靖人，广州大学硕士生，主要研究方向为移动群智感知网、社交网络隐私保护。

王国军（1970− ），男，湖南长沙人，博士，广州大学教授、博士生导师，主要研究方向为人工智能、区块链、网络安全、隐私保护。

罗恩韬（1978− ），男，湖南永州人，博士，湖南科技学院教授、硕士生导师，主要研究方向为移动社交网络、机器学习、边缘计算的安全与隐私保护等。

熊金波（1981− ），男，湖南益阳人，博士，福建师范大学教授、博士生导师，主要研究方向为云数据安全与隐私保护、移动数据安全、大数据安全。

刘忆宁（1973− ），男，河南巩义人，博士，桂林电子科技大学教授、博士生导师，主要研究方向为轻量级安全协议。

Hao Wang（1978− ），男，挪威科技大学教授、博士生导师，主要研究方向为大数据、知识管理、工业物联网、高性能计算、可信系统。