软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2019,30(12):3798−3814 [doi: 10.13328/j.cnki.jos.005601]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

移动社交网络中矩阵混淆加密交友隐私保护策略*

罗恩韬[1], 王国军[2], 刘 琴[3], 孟大程[2], 唐雅媛[1]

[1](湖南科技学院 电子与信息工程学院,湖南 永州 425199)

[2](中南大学 信息科学与工程学院,湖南 长沙 410083)

[3](湖南大学 信息科学与工程学院,湖南 长沙 410082)

通讯作者: 唐雅媛, E-mail: tangyayuan_huse@126.com

摘 要: 随着移动设备和在线社交网络的快速发展,通过用户的个人属性配置文件匹配,能够帮助用户在邻近的社交网络中迅速找到和自己共同特征的朋友.然而,交友匹配很有可能泄漏用户的敏感信息,因此用户隐私得不到保障.提出一种移动社交网络中交友匹配过程中的隐私保护协议,用户利用混淆矩阵变换算法和内积计算实现交友过程中的隐私安全和高效的匹配;用户可以细粒度定义自己特征属性的特征权重,从而使匹配结果更精确.此外,利用机会分析模型模拟真实交友场景来保证交友的有效性.安全性分析表明,提出的方法更具有隐私性、可用性和更低的通信和计算开销.通过结合真实的社会网络数据进行测试和评估,对比结果显示,比现有解决方案更有效.

关键词: 移动社交网络;隐私匹配;混淆矩阵;隐私保护;机会计算

中图法分类号: TP393

中文引用格式: 罗恩韬,王国军,刘琴,孟大程,唐雅媛.移动社交网络中矩阵混淆加密交友隐私保护策略.软件学报,2019,30(12):3798−3814. http://www.jos.org.cn/1000-9825/5601.htm

英文引用格式: Luo ET, Wang GJ, Liu Q, Meng DC, Tang YY. Privacy preserving friend discovery of matrix confusion encryption in mobile social networks. Ruan Jian Xue Bao/Journal of Software, 2019,30(12):3798−3814 (in Chinese). http://www.jos.org.cn/1000-9825/5601.htm

Privacy Preserving Friend Discovery of Matrix Confusion Encryption in Mobile Social Networks

LUO En-Tao[1], WANG Guo-Jun[2], LIU Qin[3], MENG Da-Cheng[2], TANG Ya-Yuan[1]

[1](School of Electronics and Information Engineering, Hu'nan University of Science and Engineering, Yongzhou 425199, China)

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

[3](College of Computer Science and Electronic Engineering, Hu'nan University, Changsha 410082, China)

Abstract: With the rapid developments of mobile devices and online social networks, users of mobile social networks (MSNs) can easily discover and make new social interactions with others by profiles matching. However, personal profiles usually contain sensitive information of individuals, while the emerging requirement of profile matching in proximity mobile social networks may occasionally leak the sensitive information and hence violate people's privacy. A profile matching protocol in MSNs is proposed, users utilize the confusion matrix transformation algorithm and dot product to achieve secure and efficient matching results; at the same time, users can customize the matching metrics to involve their own matching preference and to make the matching results more precise. In addition,

* 基金项目: 国家自然科学基金(61632009, 61472451, 61402543, 61272151, 61502163); 湖南省自然科学基金(2018JJ2147, 2018JJ3203); 湖南省教育厅项目(2015C0589, 17C0679); 湖南科技学院计算机应用特色学科项目

Foundation item: National Natural Science Foundation of China (61632009, 61472451, 61402543, 61272151, 61502163); Natural Science Foundation of Hu'nan Province of China (2018JJ2147, 2018JJ3203); Hu'nan Provincial Education Department of China (2015C0589, 17C0679); Construct Program of Applied Characteristic Discipline in Hu'nan University of Science and Engineering

收稿时间: 2016-09-04; 修改时间: 2017-05-09, 2018-03-18; 采用时间: 2018-05-08; jos 在线出版时间: 2019-01-21

CNKI 网络优先出版: 2019-01-22 13:49:05, http://kns.cnki.net/kcms/detail/11.2560.TP.20190122.1348.010.html

opportunistic computing is adopted to simulate the real friend making senario to guarantee the effectiveness. Security analysis shows that the proposed scheme possesses higher privacy, serviceability, and lower computation and communication cost. Assessed by real social network data, the results demonstrate that the proposed scheme is superior to the existing works.

Key words: mobile social networks (MSNs); profile matching; confusion matrix; privacy-preserving; opportunity calculation

移动社交网络(mobile social network,简称 MSN)为用户提供了更多的机会与其周围的移动用户进行社会交往,例如,彼此之间可以互相分享照片、视频、游戏以及进行交流等[1].此外,多样化的移动应用 APP 软件也为用户提供了更多的机会去拓展新的社会关系与商业机会[2](例如,微信应用中“附近的人”“微商”等).

利用用户个性化配置文件进行相似度匹配,是当前移动社交网络中发现新朋友的一个有效手段.例如,用户可以通过 APP 应用与智能设备的 Wi-Fi 接口来发现附近具有某种特征属性的朋友,进而发起交友请求.但是在交友过程中,用户之间共享信息也无形中增加了个人隐私泄漏的风险[3].例如:通过观察对方购物的爱好,可以分析用户的消费能力;通过对用户朋友圈的分析,可以确定用户的身份等.而这些隐私信息一旦被非法利用,极有可能导致被恶意用户利用进行商业欺骗或者从事其他非法活动.

因此,如何在用户之间提供良好交友匹配服务的同时,又能够保护用户个人隐私,是当前交友隐私保护中亟待解决的一个热点问题,也是移动应用服务提供商未来的研究方向.

1 相关工作

1.1 研究背景

目前,针对移动社交网络交友隐私保护的研究主要分为依靠可信服务第三方(trusted third party,简称 TTP)参与的方案和不依靠 TTP 参与的方案.

- 在有 TTP 参与的方案中[4−6],用户需要将他们的特征属性配置文件提交给 TTP,由 TTP 作为匹配中心来计算用户之间的相似度.这种方案虽然在一定程度上解放了智能终端的计算能力,提高了用户的匹配效率,但是依然存在以下安全风险:第一,一旦 TTP 攻击者被攻破,攻击者可以很容易地获取 TTP 上用户的信息,从而造成用户隐私泄漏;第二,事实上,真正意义上完全可信的第三方并不存在,因此有可能存在 TTP 因为商业利益驱动或者其他原因,出现 TTP 非法访问或者出售用户隐私数据的风险;第三,所有用户特征匹配计算均在 TTP 服务器上进行,在服务高峰时期,极有可能会造成 TTP 服务器的计算和服务瓶颈;
- 而不依靠 TTP 参与的方案大多采用复杂的密码计算来保证用户隐私安全,这主要包括对称加密运算与非对称加密运算[7−12].在这类方案中,虽然加密计算提高了对用户隐私的保护,但是在计算用户属性之间的私有交集(private set intersection,简称 PSI)时,需要对加密文件先进行解密,再进行匹配.因此在增加智能终端的计算开销的同时,也直接影响了用户体验.

为解决这个问题,后继工作中,文献[13]提出了一种基于 Paillier 加密算法的保密计算协议,可以有效保证用户隐私不被泄露.但是 Paillier 密码体制是一种具有语义安全的同态密码算法,在密钥的生成和加解密上,计算效率不高.文献[14]提出了一种基于同态加密算法的多服务器的用户特征属性的匹配方案,可以有效地保证用户的隐私不被泄露.但是该方案在表示用户特征时使用了一维向量,只考虑用户共同的属性个数,因此无法细粒度地描述用户对某种特征的偏好程度.

1.2 本文贡献

为降低现有方案的性能瓶颈以及对复杂加解密技术的依赖,同时又可以细粒度地描述用户之间特征属性的相似程度,本文在吸取了以往研究者的经验后,提出一种不依赖可信中心与复杂的加密算法,而是利用矩阵混淆变换与安全内积计算来保证交友用户的隐私安全.

1) 利用轻量级的混淆矩阵变换和向量拆分方法代替复杂的加密运算,不仅可保证用户特征属性隐私,而且能提高匹配过程的效率;

2) 利用安全内积计算用户特征属性的相似度,不需要交友用户频繁解密特征匹配文件,降低用户隐私泄露风险;

3) 利用多跳代理朋友发现机制,可以更精确地找到相同或相似特征用户,更具有可用性.

1.3 本文组织结构

本文第 2 节为方案的预备知识.第 3 节给出方案的系统模型与安全模型.第 4 节对方案进行详细设计.第 5 节讨论安全性证明与交友机会分析.性能分析和详细的实验验证在第 6 节中进行体现.

2 预备知识

2.1 大整数分解困难问题

在数论中,对于给定大于 1 的一个足够大的正整数 N,存在正整数 p,q,计算乘积 $N=p\times q$ 是非常容易的.相反地,求出 p,q,使得 $p\times q=N$,也就是求出 N 的分解式 $N=\sigma_1^{\alpha_1}\sigma_2^{\alpha_2}\sigma_3^{\alpha_3}...\sigma_k^{\alpha_k}$ 是非常困难的(其中,α_i 为正整数,σ_i 为素数,$i=1,2,\ldots,k$).这是因为大数分解与素数的判别紧密相关,而素数在正整数中的分布无任何规律.因此在密码学中,充分地利用了这一数学知识来提高密码破解难度.

为了提高分解 N 的难度,安全素数 p,q 的选择应满足以下条件.

1) p,q 的差值很大,但是位数相差不大;

2) $p-1,q-1,p+1,q+1$ 均有大素数因子;

3) $(p-1,q-1)$的最大公约数很小.

2.2 基于Paillier加密算法的基本内积计算

内积加密计算作为安全多方计算基础协议之一,主要应用于保密计算中,具体描述如下.

1) 假设发起者持有私有向量 $X=(x_1,x_2,\ldots,x_n)$,应答者持有私有向量 $Y=(y_1,y_2,\ldots,y_n)$,令 H_A^+ 为发起者的同态公钥, H_B^+ 为发起者的同态公钥,那么发起者和应答者之间的点积计算可以如下表示.

2) 发起者为向量 X 的每一个元素生成一个随机数 r_i,并利用 H_A^+ 加密向量 X,生成 $Encrypt_{H_A^+}(x_i,r_i)$ 发给应答者;

3) 应答者接受到发起者的消息后,利用自身向量元素 y_i 与 $Encrypt_{H_A^+}(x_i,r_i)$ 计算得到向量内积对应的密文 $w=\prod_{i=1}^{d}E_{H_A^+}(x_i,r_i)^{u_i}$;

4) 应答者生成新的随机数 r',利用 H_B^+ 计算 $w'=w\cdot E_{H_B^+}(0,r')$, 将计算结果 w'发送给发起者;

5) 发起者利用自己的私钥解密 w'计算得到两者的交集.

2.3 安全内积加密计算

1) 假设发起者持有私有向量 $X=(x_1,x_2,\ldots,x_n)$,应答者持有私有向量 $Y=(y_1,y_2,\ldots,y_n)$,发起者如果想知道与应答者之间的相似度,那么发起者需要计算:

$$X\bullet Y=\sum_{i=1}^{n}x_iy_i \tag{1}$$

其中,“•”表示内积.

2) 计算不可区分性:对于任意两个随机变量 X,Y,存在 $X=\{X_\omega\}_{\omega\in S},Y=\{Y_\omega\}_{\omega\in S}$ 称为计算不可区分,记为 $X\overset{c}{\equiv}Y$,如果对于任意多项式 $\{C_n\}_{n\in N}$,存在多项式 $P(\cdot),\omega\in S\cap\{0,1\}^n$,有下式成立,则满足计算不可区分:

$$|\Pr[C_n(X_w)=1]-\Pr[C_n(Y_w)=1]|<\frac{1}{P(n)} \tag{2}$$

3) 如果应答者愿意参与计算与发起者的相似度,却又不希望泄漏个人私有信息(可描述为私有向量),则可

以进行以下计算.

① 发起者 Alice 和应答者 Bob 协商 n 阶矩阵 M 和常数 k,其中,$k<n$;

② Alice 计算 $X'_A = X_A * M^{-1} = [X'_{A_1},...,X'_{A_n}]$,并将 $[X'_{A_1},...,X'_{A_K}]$ 发送给 Bob;

③ Bob 计算 $X'_B = M * X_B = [X'_{B_1},...,X'_{B_n}]$,并将 $[X'_{B_{k+1}},...,X'_{B_n}]$ 发送给 Alice;

④ Alice 计算 $Y_A = \sum_{i=k+1}^{n} x_{A_i} \bullet x'_{B_i}$;

⑤ Bob 计算 $Y_B = \sum_{i=1}^{k} x'_{A_i} \bullet x'_{B_i}$ (这一步为可选步骤).

安全性:Alice 从 $[X'_{B_{K+1}},...,X'_{B_n}]$ 中得到 $n–k$ 个有关 X_B 中分量的线性关系,同理,Bob 从 $[X'_{A_1},...,X'_{A_{K+1}}]$ 中得到 k 个有关 X_A 中分量的线性关系,但是从 $n–k$ 或者 $k(0<k<n)$方程是获取不到对方所有私有向量的分量信息的,因此可以保证私有向量的安全.经过对 Paillier 加密算法的基本内积计算与安全内积计算的比较,安全内积计算因为不需要生成同态公钥进行加密计算,显而易见地,在计算效率上更有优势.

3 系统模型与安全模型

在移动社交网络中,用户通过分享彼此个性化的特征属性文件(购物爱好、投资兴趣、地理位置、个人健康信息等,见表 1)有利于找到与自己相同或者相近特征属性的潜在朋友,从而为进一步交流和交友提供便利.但是在交友过程中,陌生用户之间互相拥有对方的特征属性信息,极有可能造成隐私泄漏,从而增加安全风险.

Table 1 User profile in mobile social networks

表 1 移动社交网络用户个人属性文档信息

姓名	Alice	Bob	Cindy
性别	女	男	女
年龄	18	52	35
特征 1	网购	医疗咨询	网购
特征 2	健康运动	投资理财	唱歌
特征 3	旅游	网购	投资理财
特征…	…	…	…
位置	**高校	**机关单位	**银行

3.1 系统模型

在以往的研究模型中,用户特征属性的隐私安全主要包括以下两个方面.

1) 特征敏感属性的隐私安全:所有参与匹配过程的发起者、应答者,都不能随意暴露自己和他人的隐私.任何一方无意或者恶意暴露用户隐私都是非法行为[15−18];
2) 通信信道安全:发起者和应答者之间信息交互时,应当保证通信信道安全,防止攻击者窃听或者截获交互信息,造成用户隐私泄漏[19−21].

因此在本方案中,为保护用户的隐私,系统模型设计如下.

假设 Alice 为发起者,Bob 和 Cindy 为应答者,发起者和应答者的角色可以进行互换,交友匹配过程如图 1 所示.经过第 1 轮一跳范围内(通信距离)交友匹配过程结束后,发起人知道与所有应答用户(通信范围内)匹配交集的大小,而应答者不知道任何发起者的隐私信息.如果发起者有意愿寻找下一跳更匹配的用户,发起者将第 1 轮匹配得到最大交集结果(阈值)和发起者经过混淆的个人特征属性配置文件交给代理用户(应答者)进行转发,由应答者作为代理寻找更远距离的匹配用户,一旦出现交集大于第 1 轮的匹配结果,那么将由代理用户通知发起者,由代理用户帮助发起者和第 2 轮的应答用户建立起联系,如果没有更佳的匹配用户存在,发起者仍然选择第 1 轮匹配交集排名最高的应答者进行交友匹配.

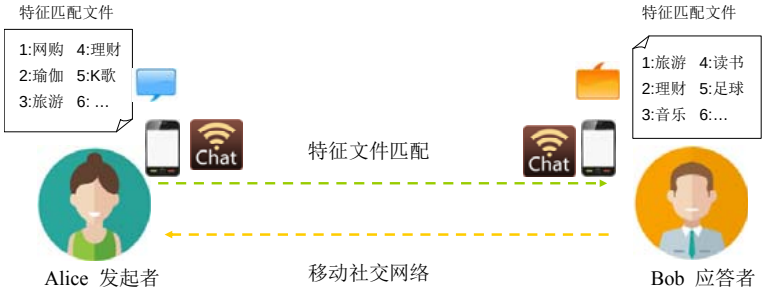


Fig.1 Profile matching model in mobile social networks

图 1 移动社交网络匹配过程模型图

3.2 攻击模型

目前,国内外研究交友匹配过程中的隐私保护,均假设存在两种攻击者.

1) 内部攻击者,也称为诚实而好奇的攻击者(honest-but-curious)[22]

在匹配过程中,内部攻击者诚实地遵守双方协议,通常不破坏协议流程,但是试图从获取的信息中通过用户行为分析[23]来获得用户更多的隐私信息(例如:通过用户每天的消费习惯来推测用户的信用额度,或者通过用户关注的医疗健康信息来了解用户的身体状况).

2) 外部攻击者,也称为恶意攻击者(malicious model)攻击模型[24,25]

外部攻击者通常不遵守协议流程,采用暴力或非法窃取合法交友用户的信息来访问未获授权的信息;监听合法交友用户通信信道并进行破解;截获合法交友用户的通信信息,进行伪装和篡改后再重传给接收者,从而阻止资源的合法管理等.在本文中,通过窃听、暴力攻击等手段的非法授权用户都属于外部攻击者.

为了进一步明确模型中的工作角色,本方案假设发起者 Alice 是完全可信的,应答者 Bob,Cindy 是诚实而好奇的,即 Bob,Cindy 等应答者会按照既定协议工作,但是不排除他们试图从获取的信息中采用用户行为分析等技术手段去窥视用户更多的隐私信息.而网络中存在的恶意攻击者是完全不可信的,即恶意攻击者有可能通过膨胀攻击、暴力推测[26−28]等方法来非法访问未经授权的数据.因此用户在网络信道上传输隐私数据之前,需要对数据进行利用大素数混淆处理.

3.3 安全模型

一般来说,要获得更高的隐私安全,那么在通信效率和计算效率上就要付出更高的计算代价.因此,针对移动社交网络的真实需求,本文的安全目标拟达到定义 1、定义 2 来保证用户之间的隐私.

定义 1(抵御内部攻击者). 匹配完成时,交友匹配双方仅仅知道彼此之间是否存在交集(共同属性),以及如果存在交集,发起者还应知道应答者与自身具体匹配的属性.除此之外,匹配双方均不知道对方与共同属性无关的其他任何信息.

定义 2(抵御外部攻击者). 匹配完成时,假设外部攻击者拦截到交互过程中的消息,外部攻击者也无法将这些消息进行解密恢复消息明文.如果外部攻击者存在身份伪装欺骗等恶意行为,那么用户能够快速识别.

基于定义 1、定义 2,本文的安全目标应能够确定信息是否来源于合法交友用户,应答者能够确定所获得的信息在传输过程中是否被篡改,用户的隐私信息在整个匹配过程中能够保证其隐私性、完整原子性、可验证性和不可抵赖性.

4 方案设计

方案包括以下 4 个阶段:系统初始化阶段,矩阵混淆和权重变换阶段,用户属性匹配阶段,分布式计算代理寻找最优匹配阶段.本文的详细匹配过程如图 2、图 3 所示.

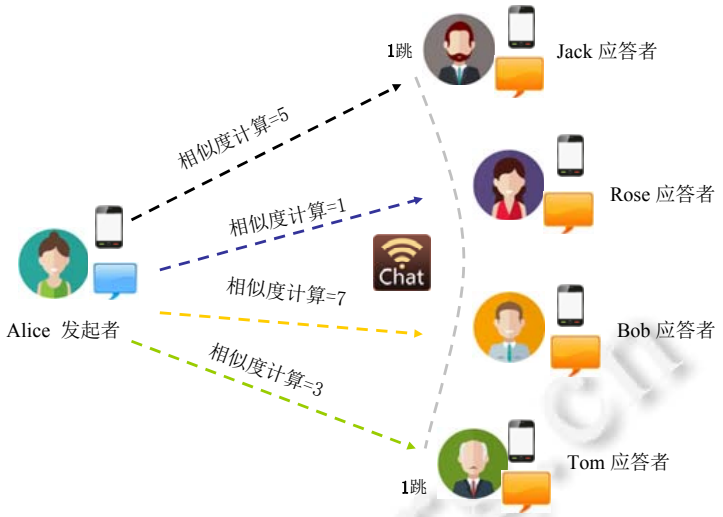


Fig.2 Multi-hops profile matching model in mobile social networks (I)

图 2 移动社交网络多跳匹配过程模型图(I)

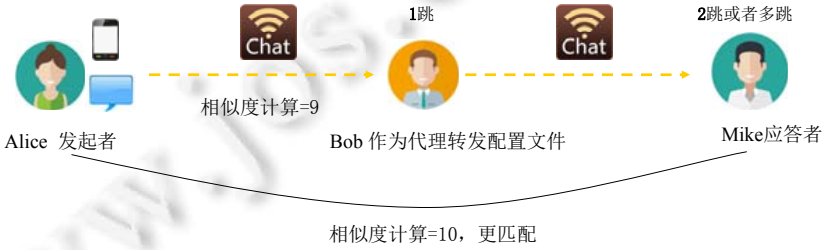


Fig.3 Multi-hops profile matching model in mobile social networks (II)

图 3 移动社交网络多跳匹配过程模型图(II)

4.1 系统初始化阶段

假设应用开发者在交友 APP 中定义了一系列个人属性可供用户选择,例如 q 个属性可分别对应 q 个用户特征向量$\{I_1,I_2,\ldots,I_q\}$,用户可选择其中自身感兴趣的 m 个属性 $m\in q$,以及对某种属性的偏好程度(个人属性兴趣权重).属性权重可以由整数 i 进行表示,$i\in[1,n]$,n 可根据实际应用场景对某种属性的偏好程度进行细粒度的设置.

当一个发起者 Alice 有意寻找他/她邻近范围内的潜在交友用户时,Alice 首先选择一定数目的属性 $Attr_{\text{Alice}}=\{Attr_{a_1},Attr_{a_2},\ldots,Attr_{a_n}\}$ 以及属性对应的权重组成矩阵 $MA_{m\times n}(m\neq n)$,矩阵的元素由 a_{ij} 表示,$a_{ij}\in[0,1]$.

假设用户拥有 3 个属性,对应矩阵的第 1 列~第 3 列,分别为看电影、游泳和购物;假设 Alice 对看电影的偏好程度为 4 级,那么需要将矩阵 a_{41} 的元素置为“1”,而该列的其他元素则设为“0”;同理,如果 Alice 对游泳的兴趣爱好为 1,购物的兴趣爱好为 5,那么 Alice 的个人属性配置矩阵可表示为

$$MA_{m\times n}=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}.$$

直观地,如果用户直接将 $MA_{m\times n}$ 发送给周围参与匹配的交友用户.如果这些用户存在攻击者,那么攻击者就可以掌握用户所有的特征属性,从而造成用户隐私泄漏.

因此,为保证用户隐私安全,本文基于大数分解困难以及非方阵求逆复杂问题,利用大素数对矩阵元素进行信息混淆,同时利用单位矩阵对 Alice 的矩阵行向量(权重信息)进行转换.转换规则由用户自身掌握,转换的目的可以有效保证用户配置文件即使被泄漏,攻击者依然不能对矩阵权重元素进行信息对应,从而可以抵制攻击者复制 Alice 配置文件副本对 Alice 进行膨胀攻击.为简化描述计算过程,本文此后的 $MA_{m\times n}$ 矩阵均用 2 行 3 列矩阵进行表示.

4.2 矩阵混淆和权重变换阶段

在本节中,Alice 通过随机产生的大素数 α,β 与两个随机产生的矩阵 $MC_{m\times n},MR_{m\times n}$ 对 $MA_{m\times n}$ 中的元素进行混淆,同时,随机生成经过行变换的单位矩阵 $MI_{m\times m}$,用来对 $MA_{m\times n}$ 矩阵的权重属性进行变换.密钥 $\vec{K}$ 在混淆过程中产生($\vec{K}$ 是用来在后继阶段将密文恢复成明文).

本步骤结束后,发起者将经过混淆的矩阵 $MA^*_{m\times n}$ 和其他信息构成请求消息 Msg_{I2R},通过广播的方式发送给周围有意愿参与交友匹配的应答者.发起者个人特征属性矩阵的混淆算法详细过程见算法 1:

$$Msg_{I2R}=\{MA^*_{m\times n},ID_{\text{Alice}},H(ID_{\text{Alice}}),\Delta t,t_1\} \tag{3}$$

ID_{Alice} 代表发起者身份特征;$H(\cdot)$是一个公开的哈希函数,$H(ID_{\text{Alice}})$代表 Alice 身份特征的哈希值;Δt 代表时间戳,用于抵抗重放攻击.为简单描述计算过程,假设用户 Alice 的属性矩阵 $MA_{m\times n}=\begin{bmatrix}1&0&1\\0&1&0\end{bmatrix}$,根据算法 1,关键计算过程步骤如下.

算法 1. 发起者属性矩阵初始混淆算法.

Input:发起者 Alice 构造属性矩阵 $MA_{m\times n},MA_{m\times n}\neq\varnothing,m\neq n$;

Output:生成混淆矩阵 $MA^*_{m\times n}$.

1. Alice 构造矩阵 $MA_{m\times n}$;
2. Alice 随机选择两个大素数 α,β,其中,$\beta>(n+1)m^2\alpha^2$;
   同时,随机生成两个大素数矩阵 $MC_{m\times n},MR_{m\times n}$,生成元素随机,并且 $\neq\alpha,\beta$;
3. $MA_{m\times n}$ 中的每一个元素分别乘以 $c_{ij},r_{ij},\forall c_{ij}\in MC_{m\times n},\forall r_{ij}\in MR_{m\times n}(1\leqslant i\leqslant m,1\leqslant j\leqslant n)$,同时,$c_{ij}$ 远远小于大素数 α;
4. **for** i=1; $i\leqslant m$; i++ **do**
5. $k_i=0;k_i\in\vec{K}$
6. **for** j=1; $j\leqslant n$; j++ **do**
7. **if** ma_{ij}==1 **then**
8. $ma_{ij}=\alpha+c_{ij}+r_{ij}\times\beta$
9. **else**
10. $ma_{ij}=c_{ij}+r_{ij}\times\beta$
11. **end if**
12. $k_i=k_i+(r_{ij}\beta-c_{ij})$;
13. **end for**
14. **end for**
15. 生成一个单位矩阵 $MI_{m\times m}$,用来对 $MA_{m\times n}$ 矩阵的权重属性进行变换.
16. $MA^*_{m\times n}=MI_{m\times m}\times MA_{m\times n};ma^*_{ij}=\sum_{k=1}^{n}a_{ik}\cdot b_{kj},\forall ma^*_{ij}\in MA^*_{m\times n},\forall a_{ik}\in MI_{m\times m},\forall b_{kj}\in MA_{m\times n}\ (0\leqslant i\leqslant m,0\leqslant j\leqslant n)$
17. **Return** $MA^*_{m\times n}$.

$$MA_{m\times n}=\begin{bmatrix}\alpha+c_{11}+r_{11}\beta & c_{12}+r_{12}\beta & \alpha+c_{13}+r_{13}\beta\\ c_{21}+r_{21}\beta & \alpha+c_{22}+r_{22}\beta & c_{23}+r_{23}\beta\end{bmatrix} \tag{4}$$

那么经过行初等变换的矩阵为

$$MA^*_{m\times n}=\begin{bmatrix} c_{21}+r_{21}\beta & \alpha+c_{22}+r_{22}\beta & c_{23}+r_{23}\beta \\ \alpha+c_{11}+r_{11}\beta & c_{12}+r_{12}\beta & \alpha+c_{13}+r_{13}\beta \end{bmatrix}.$$

$$\vec{K}=\left(\sum_{j=1}^{3}r_{1j}\beta-c_{1j},\sum_{j=1}^{3}r_{2j}\beta-c_{2j}\right) \tag{5}$$

例如:$k_1=(r_{11}\beta-c_{11})+(r_{12}\beta-c_{12})+(r_{13}\beta-c_{13})$.

4.3 用户属性匹配阶段

4.3.1 矩阵相乘

假设应答者 Bob(或者其他应答者)接收到发起者的查询信息 Msg_{I2R},并且有意与发起者 Alice 进行交友,那么 Bob 将对 Alice 的身份信息进行数据完整性验证.

首先,Bob 计算消息在接收时刻 t_2 减去发送时刻 t_1 是否小于Δt 来对抗重放攻击;同时,利用公开的哈希函数 $H(\cdot)$对 ID_{Alice} 进行哈希,并与 Msg_{I2R} 中 $H(ID_{\text{Alice}})$进行比较,如果值相等,说明信息在传递过程中身份信息 ID_{Alice} 没有被攻击者篡改.

其次,Bob 根据自身属性配置文件 $Attr_{\text{Bob}}=\{Attr_{b_1},Attr_{b_2},...,Attr_{b_n}\}$ 构造转置矩阵 $MB^T_{m\times n}$ 与 $MA^*_{m\times n}$ 进行匹配,为方便矩阵计算,应答者对自身矩阵进行了转置处理,计算 $MA^*_{m\times n}\times MB^T_{m\times n}$,算法详细过程见算法 2.

在计算过程中,当两个矩阵拥有交集的时候,算法将对应矩阵元素乘积设置为 1,否则设置为 0.算法结束后,Bob 将获得一个新矩阵 $MD_{m\times m}$:

$$MD_{m\times m}=MA^*_{m\times n}\times MB^T_{m\times n}=(d_{ij})_{m\times m} \tag{6}$$

最后,Bob 将 $MD_{m\times m}$ 组成应答消息 Msg_{I2R} 发送给 Alice,发送的消息为

$$Msg_{I2R}=\{MD_{m\times m},ID_{\text{Bob}},H(ID_{\text{Bob}}),\Delta t\} \tag{7}$$

算法 2. 应答者矩阵相乘算法.

Input: $MA^*_{m\times n}$,$MB_{m\times n}$;

Output:生成新矩阵 $MD_{m\times m}=(d_{ij})_{m\times m}$.

1. Bob 根据自身兴趣爱好生成 $MB_{m\times n}$,并计算 $MD_{m\times m}=MA^*_{m\times n}\times MB^T_{m\times n}$;
2. **for** i=1; $i\leqslant m$; i++ **do**
3. **for** j=1; $j\leqslant n$; j++ **do**
4. $temp$=0;
5. **for** q=1; $q\leqslant n$; q++ **do**
6. **if** mb_{jq}==1 **then**
7. $temp=temp+\alpha\times ma^*_{iq}$
8. **else**
9. $temp=temp+ma^*_{iq}$
10. d_{ij}=$temp$;
11. **end if**
12. **end for**
13. **end for**
14. **end for**
15. **Return** $MD_{m\times m}$

为简单描述计算过程,假设 $MB_{m\times n}=\begin{bmatrix}0&0&1\\1&1&0\end{bmatrix}$,关键计算过程步骤如下:

$$MD_{m\times m}=MA^{*}_{m\times n}\times MB^{T}_{m\times n}=\begin{bmatrix}a^{*}_{11} & a^{*}_{12} & a^{*}_{13}\\ a^{*}_{21} & a^{*}_{22} & a^{*}_{23}\end{bmatrix}\begin{bmatrix}b_{11} & b_{12} & b_{13}\\ b_{21} & b_{22} & b_{23}\end{bmatrix}^{T}=\begin{bmatrix}d_{11} & d_{12}\\ d_{21} & d_{22}\end{bmatrix} \tag{8}$$

推导过程如下:

$$\begin{aligned}d_{11}&=(c_{21}+r_{21}\beta)+(\alpha+c_{22}+r_{22}\beta)+\alpha(c_{23}+r_{23}\beta);\\ d_{12}&=\alpha(c_{21}+r_{21}\beta)+\alpha(\alpha+c_{22}+r_{22}\beta)+(c_{23}+r_{23}\beta);\\ d_{21}&=(\alpha+c_{11}+r_{11}\beta)+(c_{12}+r_{12}\beta)+\alpha(\alpha+c_{13}+r_{13}\beta);\\ d_{22}&=\alpha(\alpha+c_{11}+r_{11}\beta)+\alpha(c_{12}+r_{12}\beta)+(\alpha+c_{13}+r_{13}\beta).\end{aligned}$$

4.3.2 矩阵解密属性交集

Alice 接收到 Bob 的信息 Msg_{I2R} 后,将利用单位矩阵 $MI_{m\times m}$ 的逆矩阵 $MI^{-1}_{m\times m}$ 来恢复原矩阵的权重关系;同时,使用算法 1 中的解密密钥 $\vec{K}$ 和算法 3 的步骤 2、步骤 3 对矩阵 $MD_{m\times m}$ 中的元素进行数据还原工作,具体执行步骤如下:

$$MT_{m\times m}=MI^{-1}_{m\times m}\times MD_{m\times m}=\begin{bmatrix}T_{11} & T_{12}\\ T_{21} & T_{22}\end{bmatrix}=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}d_{11} & d_{12}\\ d_{21} & d_{22}\end{bmatrix}=\begin{bmatrix}d_{21} & d_{22}\\ d_{11} & d_{12}\end{bmatrix} \tag{9}$$

推导过程如下:

$$\begin{aligned}T_{11}&=(d_{21}+k_1)\bmod\beta\\ &=\{(\alpha+c_{11}+r_{11}\beta)+(c_{12}+r_{12}\beta)+\alpha(\alpha+c_{13}+r_{13}\beta)+k_1\}\bmod\beta\\ &=\{(\alpha+c_{11}+r_{11}\beta)+(c_{12}+r_{12}\beta)+\alpha(\alpha+c_{13}+r_{13}\beta)+(r_{11}\beta-c_{11})+(r_{12}\beta-c_{12})+(r_{13}\beta-c_{13})\}\bmod\beta\\ &=\{[(\alpha+c_{11}+r_{11}\beta)+(r_{11}\beta-c_{11})]+[(c_{12}+r_{12}\beta)+(r_{12}\beta-c_{12})]+[\alpha(\alpha+c_{13}+r_{13}\beta)+(r_{13}\beta-c_{13})]\}\bmod\beta\\ &=\{[\alpha+2r_{11}\beta]+2r_{12}\beta+\alpha^2+(\alpha+1)r_{13}\beta+(\alpha-1)c_{13}\}\bmod\beta\\ &=\{[\alpha^2+\alpha+(\alpha-1)c_{13}]+[2r_{11}+2r_{13}+(\alpha+1)r_{13}]\beta\}\bmod\beta\\ &=\alpha^2+\alpha+(\alpha-1)c_{13}.\end{aligned}$$

得到 T_{11} 值之后,将利用模运算对 T_{11} 进行归一化,得到 T^{*}_{11}.

$$T^{*}_{11}=\frac{T_{11}-(T_{11}\bmod\alpha^2)}{\alpha^2}=\frac{T_{11}}{\alpha^2}-\frac{T_{11}\bmod\alpha^2}{\alpha^2}=\frac{\alpha^2+\alpha+(\alpha-1)c_{13}}{\alpha^2}-\frac{[\alpha^2+\alpha+(\alpha-1)c_{13}]\bmod\alpha^2}{\alpha^2}=1-\frac{1}{\alpha^2}\approx 1 \tag{10}$$

因为α,β为一个大素数,同时在算法 1 中,定义 C_{ij} 远远小于大素数α,因此可得到$\alpha+(\alpha-1)c_{13}<\alpha^2$,所以 $T^{*}_{11}=1-\frac{1}{\alpha^2}\approx 1$.同理可以求得 $T_{12}=1,T_{21}=0,T_{22}=1$.

通过以上步骤,可以获得两个矩阵的相似度矩阵为 $MT_{m\times m}=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$.

4.3.3 权重矩阵构造和相似度计算

在 Alice 接收到 Bob 或者其他应答者反馈的消息 $MD_{m\times m}=(d_{ij})_{m\times m}$ 的同时,Alice 将根据 $MD_{m\times m}$ 矩阵元素下标构造一个权重矩阵$(W_{ij})_{m\times m}$ 来恢复原矩阵的权重关系,该权重矩阵用来描述发起者和应答者特征属性之间的关系,也是为进行矩阵内积计算而得到两者之间的相似度.根据权重映射关系(对角线上的元素相似度最大),同时为保证权重之间的差异化,设计具体权重转换公式:

$$(W_{ij})_{m\times m}=\begin{cases}m+i^2, & i=j\\ m+j-i, & i-j>0\\ m-j+i, & i-j<0\end{cases} \tag{11}$$

根据公式(11),可求得矩阵 $MD_{m\times m}$ 权重关系为 $W_{ij}=\begin{bmatrix}1 & 1\\ 1 & 4\end{bmatrix}$.

根据权重关系 W_{ij} 与相似度矩阵 MT 对应元素内积计算,可以精确计算发起者和应答者相似度值 *Similary*,其中,“•”代表内积.

假设存在另外一个应答者 Cindy,其个人属性矩阵为 $MCindy_{m\times n}=\begin{bmatrix}1 & 1 & 1\\ 0 & 0 & 0\end{bmatrix}$,那么她与 Alice 运算得到的相似度矩阵为 $MT_{m\times m}=\begin{bmatrix}2 & 0\\ 1 & 0\end{bmatrix}$,其相似度计算结果为

$$Similary_{Cindy}=MT\bullet W=\begin{bmatrix}2 & 0\\ 1 & 0\end{bmatrix}\bullet\begin{bmatrix}1 & 1\\ 1 & 4\end{bmatrix}=\begin{bmatrix}2\times 1 & 0\times 1\\ 1\times 1 & 0\times 4\end{bmatrix}=2+0+1+0=3.$$

通过比较发现,Bob 与 Alice 的特征属性更相似.类似地,Alice 就可以从所有的应答者中选择相似度值最大的用户作为其自身的匹配用户.

算法 3. 特征属性相似度(交集)计算算法.

Input:$(W_{ij})_{m\times m}$,$MD_{m\times m}=(d_{ij})_{m\times m}$;

Output:通过安全内积计算出用户之间相似度λ.

1. Alice 计算 $MT_{m\times m}=MI_{m\times m}^{-1}\times MD_{m\times m}$,$MI_{m\times m}^{-1}$ 是 $MI_{m\times m}$ 逆矩阵;
2. Alice 计算 $mt_{ij}=(k_i+d_{ij}) \bmod \beta$;$\forall mt_{ij}\in MT_{m\times n}$;
3. Alice 进一步计算 $mt_{ij}=\dfrac{mt_{ij}-(mt_{ij} \bmod \alpha^2)}{\alpha^2}$;
4. Alice 计算 $H_{m\times m}=MT_{m\times m}\bullet(W_{ij})_{m\times m}=\sum_{i=1}^{m}\sum_{j=1}^{m}h_{ij}=\sum_{i=1}^{m}\sum_{j=1}^{m}a_{ij}b_{ij}$;$\forall h_{ij}\in H_{ij}$,“$\bullet$”为内积;
5. **Return** $\lambda=\sum_{i=1}^{m}\sum_{j=1}^{m}h_{ij}$.

4.4 分布式计算代理寻找最优匹配阶段

完成步骤 4.3,Alice 将知道在她一跳范围内最匹配的交友用户.但是在实际的情况中,有可能在应答者的下一跳会出现与 Alice 特征属性更匹配的用户,因此本文假设第 1 轮的应答者 Bob 作为代理转发 Alice 的配置文件和第 1 轮所有应答用户最大相似度的值$\lambda_{\max}$.为减少代理和发起者的通信开销,本方案设计只有相似度值大于$\lambda_{\max}$ 的用户才能够得到返回.同时,为了避免过大的通信开销和计算时间消耗,Alice 可以选择代理转发的跳数.通过这种方法,发起者重新与新的最佳应答者(代理的下一跳的用户)建立起通信会话,从而找到更适合自己的交友用户,具体过程见图 3 和算法 4.

算法 4. 分布式计算代理寻找最优匹配算法.

Input: $MA_{m\times n}^{*}$ 和$\lambda_{\max}$;

Output:选择最匹配的用户.

1. 代理用户(Bob)转发发起者配置文件,由代理用户和其下一跳进行计算,算法过程和算法 1~算法 3 相同
2. 代理得到他们下一跳应答者最大的匹配值 $\lambda'_{\max}$
3. $\lambda_{\max}=Swap(\lambda_{\max},\lambda'_{\max})$
4. **if** $(\lambda_{\max}<\lambda'_{\max})$
5. $temp=\lambda_{\max}$
6. $\lambda_{\max}=\lambda'_{\max}$
7. $\lambda'_{\max}=temp$
8. **end if**
9. **Return** $\lambda_{\max}$

5 安全与机会分析

本节将分别考虑针对恶意攻击和诚实而好奇攻击情况下对本方案的安全性进行证明,着重讨论发起者与应答者之间的隐私保护.为简化描述,假设 Bob 是 Alice 最佳匹配者.

5.1 安全分析

5.1.1 抵御外部攻击者膨胀攻击和暴力攻击

挑战 1. 攻击者可以攻击用户之间的通信信道,如果攻击者可以成功截获用户之间的通信密文,并能够将此密文恢复成明文,那么攻击者将赢得这个挑战.

引理 1. 本方案可以成功抵御外部恶意攻击者的膨胀攻击.

证明:假设外部攻击者可以窃听用户 Alice 和应答者 Bob 之间的通信过程,并且拦截到 Alice 与 Bob 的消息 $MA^*_{m\times n}$,攻击者伪造 $MA^*_{m\times n}$,生成文件的拷贝 $MA'^*_{m\times n}$, 并利用拷贝文件与 $MA^*_{m\times n}$ 相乘,计算出 $MD'_{m\times m}$ 并发送给 Alice,从而欺骗 Alice 为最佳匹配.但是因为伪装拷贝矩阵与原矩阵都是大素数对位相乘,因此矩阵中元素的数值将非常庞大,这将会很快被 Alice 发现异常.

另外,因为攻击者不是合法用户,所以攻击者并不知道 Bob 的计算规则(算法 2),因此伪造的 $MD'_{m\times m}$ 不能被 Alice 进行解密,所以攻击者将很快地被 Alice 识别,攻击失败, □

引理 2. 本方案可以成功做到抵御外部者恶意攻击者的暴力推测.

证明:假设攻击者利用背景知识或者其他攻击手段试图推导用户更多的隐私,攻击者可根据用户的特征属性、兴趣爱好、地理位置等构造攻击字典,并试图利用攻击字典来暴力破解用户的隐私.但是根据腾讯微博调查结果显示:移动社交网络中用户的特征属性具有多样化的特征,普通社交用户至少拥有 11 个特征属性来细粒度地表示自己的兴趣爱好等.在本方案中,如果每个属性有 10 个权重,那么暴力攻击者试图分析出用户的真实属性,将至少会有 $K=\prod_{1\leqslant i\leqslant 10} i>2^{21}$ 种选择.也就是说,只有 $1/K=1/3628800$ 的概率可以分析出用户的真实矩阵情况.因此,攻击者试图通过构造攻击字典来暴力攻击,那么这个计算开销将非常庞大. □

5.1.2 抵抗内部攻击者的用户行为分析

挑战 2. 假设 Bob 是内部攻击者,彼此试图通过接收到的信息 $MA^*_{m\times n}$ 进行用户行为分析,从而推测对方所有真实属性和对某种属性的偏好程度.如果 Bob 可以成功恢复出原始矩阵 $MA_{m\times n}$,那么攻击者将赢得这个挑战.

引理 3. 本协议可以成功保护发起者的隐私.

证明:发起者 Alice 根据自身特征属性,编码混淆矩阵 $MA^*_{m\times n}$ 发送给应答者 Bob.在正常情况下,Bob 只需要按照协议执行,通过矩阵计算得到 $MD_{m\times m}$(内部的元素被混淆,Bob 不知道与 Alice 的真实交集)并将计算结果返回给 Alice.假设 Bob 是诚实而好奇的攻击者,试图通过用户行为分析推测 Alice 的真实属性和兴趣权重.但是由于 $MA_{m\times n}$ 中所有元素都被大素数 α,β 与随机矩阵 MC,MR 进行混淆,同时用户的兴趣权重也通过单位矩阵 $MI_{m\times m}$ 进行转换,而具体的变换规则和大素数只有 Alice 知道,因此 Bob 试图通过 $MA^*_{m\times n}$ 恢复出 $MA_{m\times n}$ 将面临大数分解难题,而这已经被证明是非常困难的,因此发起者隐私将得到保护. □

5.2 交友机会分析

为衡量真实移动社交网络中交友匹配的参与用户,本文模拟在时间 t 内将参与人作为代理,为满足交友计算需求能够提供的有效计算资源数.

引理 4. 在[0,t]时间内,所有用户可以提供预期总计算资源为 $E\{R(t)\}=\frac{\lambda t^2 p\eta}{2}$.

证明:假设所有用户可以提供预期总资源 $E\{R(t)\}$ 为 1,那么:

$$E\{R(t)\}=\sum_0^\infty (P\{N_q(t)=n\}E\{R(t)\mid N_q(t)=n\})=\sum_0^\infty P\{N_q(t)=n\}\cdot\frac{nt\eta}{2}=\frac{t\eta}{2}\cdot E(N_q(t))=\frac{\lambda t^2 p\eta}{2}=1.$$

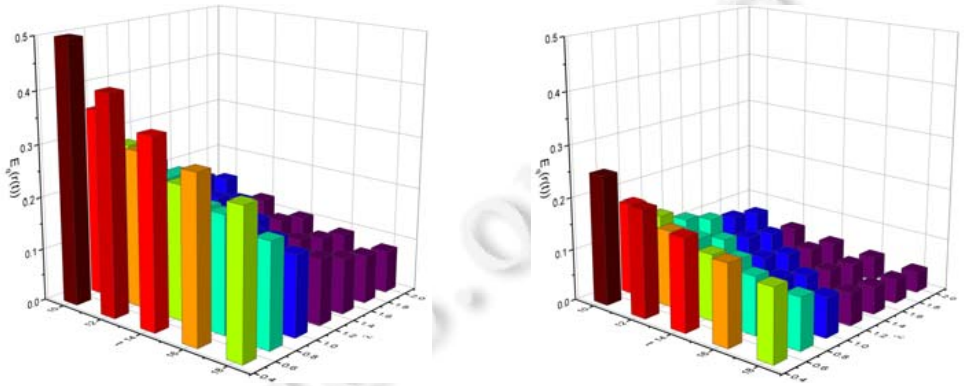
根据 $E(N_q(t))=\lambda tp$,单个用户提供的计算资源为 $E(r_q(t))=E\{R(t)\}/\lambda tp$,并满足 $(R/\lambda tp)<\frac{nt\eta}{2}$.

根据真实应用场景需求,将时间考虑在 60s、120s、180s、240s、300s 的参与人数和单个用户提供的计算资源,计算模型仿真结果如表 2 和图 4 所示.

Table 2 Provide resource expectation

表 2 提供资源预期

p / λ	$E(r_q(t))=R/\lambda tp$									
	p=0.4					p=0.8				
	60	120	180	240	300	60	120	180	240	300
0.5	0.083 333	0.041 667	0.027 778	0.020 833	0.016 667	0.041 667	0.020 833	0.013 889	0.010 417	0.008 333
0.7	0.059 524	0.029 762	0.019 841	0.014 881	0.011 905	0.029 762	0.014 881	0.009 921	0.007 44	0.005 952
0.9	0.046 296	0.023 148	0.015 432	0.011 574	0.009 259	0.023 148	0.011 574	0.007 716	0.005 787	0.004 63
1.1	0.037 879	0.018 939	0.012 626	0.009 47	0.007 576	0.018 939	0.009 47	0.006 313	0.004 735	0.003 788
1.3	0.032 051	0.016 026	0.010 684	0.008 013	0.006 41	0.016 026	0.008 013	0.005 342	0.004 006	0.003 205
1.5	0.027 778	0.016 026	0.009 259	0.006 944	0.005 556	0.016 026	0.006 944	0.004 63	0.003 472	0.002 778
1.7	0.024 51	0.012 255	0.008 17	0.061 728	0.004 902	0.012 255	0.061 728	0.004 085	0.003 064	0.002 451



(a) 设定阈值为 P=0.4 时,所有代理用户预计提供的资源预期 (b) 设定阈值为 P=0.8 时,所有代理用户提供的资源预期

Fig.4 Opportunity calculation model

图 4 机会计算参与人数和提供资源模型图

从图 4(a)、图 4(b)可以看出:总资源一定的情况下,在移动社交活动稀疏时刻,计算资源闲置时,可通过降低参与计算应答者的概率 p,做到增加社交交友的参与人数 $E(N_q(t))$,从而促进社交活动更有效地开展.

6 性能分析

6.1 复杂度分析

6.1.1 计算开销

在本节中,将与现有研究工作进行计算开销的对比分析,对于计算开销,本文主要考虑方案中乘法运算和加法运算的次数.

本文采用 exp1 标识 1 024 位的求幂操作,exp2 标识 2 048 位的求幂操作,*add* 表示模加运算,*mul*1,*mul*2 分别表示 1 024 位和 2 048 位的乘法运算.详细比较结果见表 3.

Table 3 Computation cost

表 3 计算数据开销

方案	离线计算开销		在线计算开销	
	发起者	应答者	发起者	应答者
本文方案	$2m{\cdot}n{\cdot}mul1+3m{\cdot}n{\cdot}add$	–	$2m{\cdot}m{\cdot}mul1+3m{\cdot}m{\cdot}add$	$m{\cdot}m{\cdot}n{\cdot}mul1+m{\cdot}m{\cdot}n{\cdot}add$
WAS[29]	$n{\cdot}$exp1$+n{\cdot}h$	$n{\cdot}$exp1$+n{\cdot}h$	$n{\cdot}$exp1	$n{\cdot}$exp1
Fine-grained[30]	$2m{\cdot}n{\cdot}$exp1$+m{\cdot}n{\cdot}mul2$	–	1·exp2	1·exp1+1·exp2$+n{\cdot}mul2$

表 3 中,因为本文协议采用矩阵运算和大素数混淆运算,发起用户在离线状态时,仅需要将原始矩阵与混淆矩阵相乘,再利用随机行变换矩阵进行混淆,因此需要 $2m{\cdot}n{\cdot}mul1$ 的计算开销.矩阵相乘后对矩阵运算进行相加

计算,需要 $3m\cdot n\cdot add$ 的计算开销,所以发起者离线状态下的总计算开销为 $2m\cdot n\cdot mul1+3m\cdot n\cdot add$,见算法 1.与离线状态下 WAS 与 Fine-grained 方案均采用了较复杂 exp1 指数运算对比,显然更有优势.

同时,在发起用户在线阶段,需要进行 2 次乘法运算和 3 次加法运算,见算法 3.因此,总计算开销为 $2\cdot m\cdot m\cdot mul1+3\cdot m\cdot m\cdot add$;而应答用户则需要 $m\cdot m\cdot n\cdot mul1+m\cdot m\cdot n\cdot add$ 的计算开销,见算法 2.相比 WAS 与 Fine-grained 方案,本方案计算开销更小.

6.1.2 通信开销

在本节中,将与现有研究工作进行通信开销的对比分析.通信开销通常是由协议中的通信次数或者协议中发送的比特位数来决定.

假设每个用户的属性个数和属性权重分别是 n 和 m,接收和发送数量用比特位数进行计算.在用户信息交互过程中,发起者仅仅需要将自身矩阵大小乘以可变密钥长度,所以通信开销为 $m\cdot n\cdot k$;而应答者因为在多跳阶段需要承担交友配置文件的转发任务,因此通信开销为 $2\cdot m\cdot n\cdot k$.

本方案因为采用可变密钥 k,因此可适用不同的安全需求场景.在安全性需求较高的情况下,用户可以选择较长的可变密钥 k;在安全性需求较低的情况下,用户可以选择较短的可变密钥 k.与 WAS 和 Fine-grained 协议使用固定长度 1 024bit 和 2 048bit 的密钥相比,显然更为灵活高效.

6.2 模拟实验和仿真结果

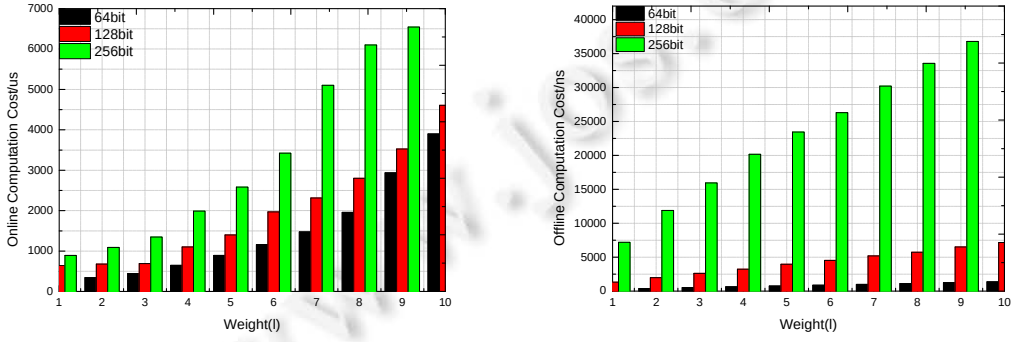
在本文的测试环境中,利用小米手机 NOTE 版进行群组测试.编程环境使用 Eclipse,利用 Java 作为编程语言进行代码开发.硬件条件为:CPU 骁龙™ 8X74AC 801 处理器,主频 2.5GHz,使用 LPDDR3 933MHz 3G 高速内存,支持蓝牙 4.0 和 Wi-Fi 双频.

开发库为(java.math.BigInteger/java.util.Arrays/java.util.Random),用户特征属性(兴趣爱好)利用爬虫代码从社交网站进行抓取并进行处理.

考虑用户实际应用对数据安全的差异性需求,本文分别采用 64 位、128 位、256 位的大素数作为密钥进行实验,同时采用不同数目的权重和不同数目的属性对算法进行了对比分析.

图 5(a)~图 5(d)分别显示密钥长度为 64 位、128 位、256 位在线计算开销和离线计算开销在不同权重和属性影响下的评价结果.离线计算开销表示发起者构造 $MA^*_{m\times n}$ 的计算时间,在线计算开销表示应答者计算 $MD_{m\times m}=MA^*_{m\times n}\times MB^T_{m\times n}$ 的时间.经过比较,在线情况下,属性值对计算开销的影响比权重的影响要大,在线计算开销时间(单位:μs)稍大于离线计算开销时间.这是因为在线处理时 $MA^*_{m\times n}$ 中的元素经过了大素数和随机矩阵的混淆计算,而离线计算情况下,$MA_{m\times n}$ 中的元素还是 0 或者 1.

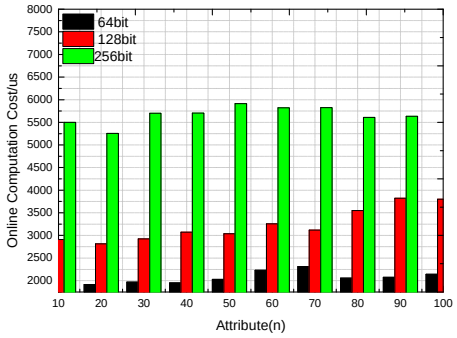
同时,从图 5 中可以看出:在线计算的时间以μs 为单位,而离线计算的计算时间以 ns 为单位,这个时间对于进行移动社交网络交友的用户几乎可以忽略不计,很好地保证了在交友过程中的用户体验.



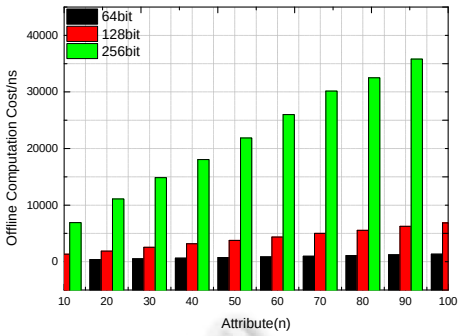
(a) 在线:权值递增,属性值固定为 10 (b) 离线:权值递增,属性值固定为 10

Fig.5 Initiator offline and online computation cost

图 5 发起者离线计算和在线计算开销图



(c) 在线:属性递增,权值固定为 10

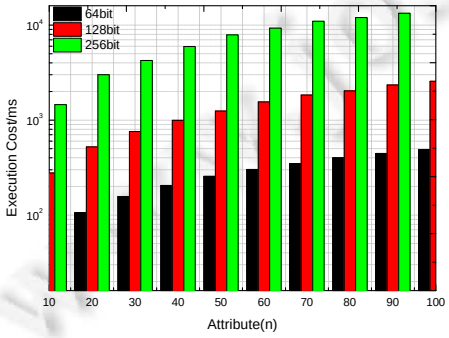


(d) 离线:属性递增,权值固定为 10

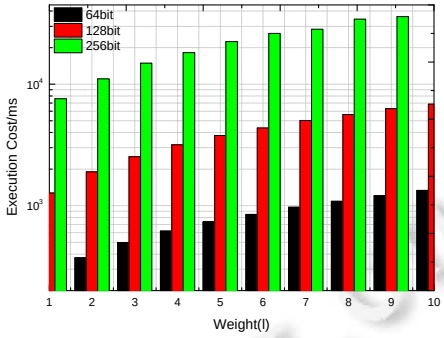
Fig.5 Initiator offline and online computation cost (Continued)

图 5 发起者离线计算和在线计算开销图(续)

图 6(a)、图 6(b)分别表示权值和属性变化时对计算总时间的影响.当权值和属性个数发生改变时,相较于权值的改变,属性个数改变对计算总时间产生更大的影响,这也符合真实移动社交网络交友匹配的情景.因为通常在交友活动中,用户希望提供更细粒度的属性选择,以便更精确地匹配.特别地,在图 7(a)中,当属性值依次从 10~100 依次进行递增,执行总时间相差并不大.在图 7(b)中,属性保持 10 个可选属性,密钥长度使用 256 位的大素数加密时,依然能保持很好的用户体验.



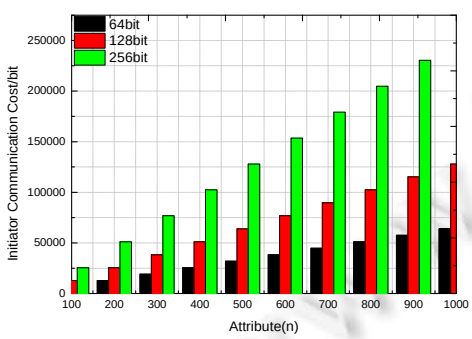
(a) 权值固定为 10,属性值递增



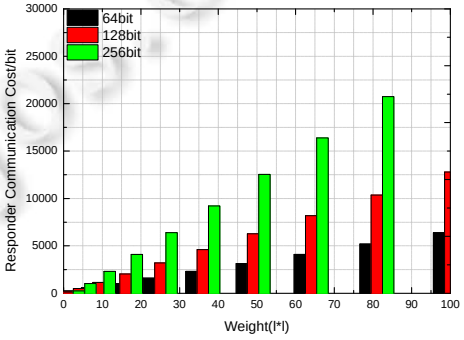
(b) 属性值固定为 10,权值递增

Fig.6 Attributes and weights change execution cost

图 6 属性和权重分别变化执行总时间图



(a) 发起者:权值固定为 10,属性值递增



(b) 应答者:权值递增,属性值不变

Fig.7 Initiator and responder communication cost

图 7 发起者和应答者的通信开销

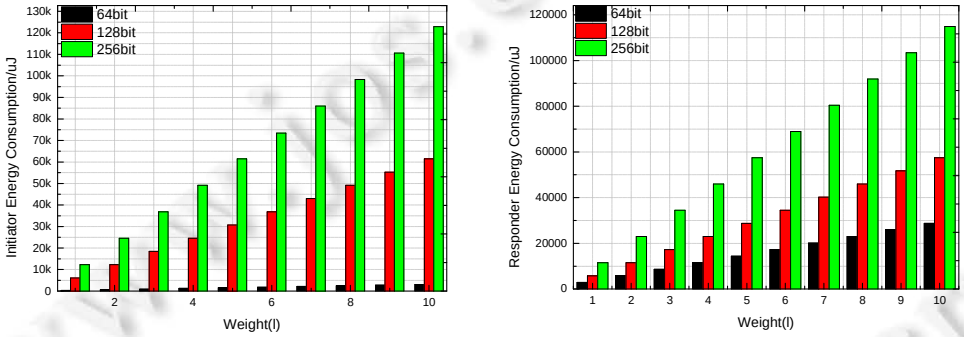
图 7(a)、图 7(b)分别表示发起者和应答者的通信开销.本文假设扩大 1 倍的通信范围去寻找与发起者更匹配的用户,图 7(a)表示权重固定为 10,属性数目依次递增发起者的通信开销,横坐标表示属性数目.图 7(b)表示应答者通信开销,横坐标表示权重,通过与表 4 中其他协议比较发现:即使本文提出的方案扩大了通信范围,因为采用代理承担通信负载,因此通信效率得到了提高.由此可得出结论:本文在扩大交友匹配范围的同时,通信消耗并没有明显的级数增长,依然是线性的.

Table 4 Communication cost

表 4 通信数据开销

方案	通信开销(bits)	
	发起者	应答者
本文方案	$(m{\cdot}n){\cdot}k,k$ 为可变密钥长度	$2(m{\cdot}n)k,k$ 为可变密钥长度
WAS	$2n{\cdot}(m{\cdot}n){\cdot}1024$	$(n+2){\cdot}1024$
Fine-grained	$m{\cdot}n{\cdot}2048$	$m{\cdot}n{\cdot}2048$

同时,在移动社交网络中,运行安装在移动终端上的 APP 的能耗也是一个重要的考虑因素.本文通过参考文献[8],利用能耗计算公式 $E=N_t{\cdot}E_t+N_r{\cdot}E_r$ 进行了计算,其中,N_t,N_r 分别代表传输数据和接受数据.根据每比特的发送能量消耗 $E_t{\approx}4.8\mu J$ 和接收能量消耗 $E_r{\approx}6.7\mu J$,为简单描述,本文仅仅选取权重属性作为能量消耗的参考因素进行对比,得出如下的计算结果,如图 8 所示.



(a) 发起者:能量消耗 (b) 应答者:能量消耗

Fig.8 Initiator and responder energy consumption

图 8 发起者和应答者能量消耗图

通过全面的对比分析,本方案与传统的利用对称加密、非对称加密技术进行社交交友方案相比较,在计算开销、通信开销和能量消耗上均有较明显的优势.最后,本文在方法的适应性上与其他协议进行了比较,可以发现,本方案更具有通用性(见表 5).

Table 5 Adaptability comparison of typical privacy preserving methods in mobile social networks

表 5 移动社交网络典型隐私保护方法适应性比较

方案	核心算法	优点	缺点
Privacy-preserving mathmaking[21]	数据扰乱	能保护基本隐私	算法效率低
Distributed private matching[14]	同态加密	能抵御恶意攻击	计算开销较大
Secure handshake[6]	双线性映射	安全性高,实现简单	难以扩充到多属性匹配运用中
S-match[12]	布隆过滤器迭代	减少存储空间	有一定的误判率
本文方案	矩阵混淆计算	能抵御恶意攻击,安全性高,实现简单	算法效率高,无误判率,可以扩充到多属性,多权重匹配

7 结束语

在移动社交网络中,最大化增强彼此之间的联系和交流,同时又保护用户的个人隐私问题,是当前隐私保护

方向的一个研究热点.本文基于数论基础,提出不依赖 TTP 可信服务器轻量级的矩阵混淆和多跳代理方案,实现了移动社交网络交友匹配的隐私保护.在计算量上没有使用复杂的双线性对和指数运算,只使用了计算开销较小的哈希函数运算、取模运算和内积计算等.该方案提高了移动社交网络中用户的交友效率,使用户能够迅速发现邻近范围内与发起者属性匹配的用户,减少了移动终端计算和通信开销.通过机会分析、安全和性能分析,本文提出的协议可以在终端资源受限的情况下,让用户更有效、更安全地进行移动社交活动.

References:

[1] Fu YY, Zhang M, Feng DG, Chen KQ. Attribute privacy preservation in social networks based on node anatomy. Ruan Jian Xue Bao/Journal of Software,2014, 25(4):768−780 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4565.htm [doi: 10.13328/j.cnki.jos.004565]

[2] Zhang L, Li XY, Liu Y. Message in a sealed bottle: Privacy preserving friending in social networks. In: Proc. of the IEEE Int'l Conf. on Distributed Computing Systems (ICDCS). 2013. 327−336. [doi: 10.1109/ICDCS.2013.38]

[3] Wang Y, Vasilakos AV, Jin Q. Survey on mobile social networking in proximity (MSNP): Approaches, challenges and architecture. Wireless Networks, 2013,20(6):1295−1311. [doi: 10.1007/s11276-013-0677-7]

[4] Guo L, Zhang C, Sun J. A privacy-preserving attribute-based authentication system for mobile health networks. IEEE Trans. on Mobile Computing, 2014,13(9):1927−1941. [doi: 10.1007/s11276-013-0677-7]

[5] Guo L, Zhu X, Zhang C. Privacy-preserving attribute-based friend search in geosocial networks with untrusted servers. In: Proc. of the IEEE Int'l Conf. on Global Communications (GLOBALCOM). 2013. 629−634. [doi: 10.1109/GLOCOM.2013.6831142]

[6] Lu R, Lin X, Liang X, Shen X. A secure handshake scheme with symptoms-matching for mHealthcare social network. Mobile Networks and Applications, 2011,16(6):683−694. [doi: 10.1007/s11036-010-0274-2]

[7] Sarpong S, Xu C. A secure and efficient privacy-preserving attribute matchmaking protocol in proximity-based mobile social networks. In: Proc. of the Advanced Data Mining and Applications. 2014. 305−318. [doi: 10.1007/978-3-319-14717-8_24]

[8] Li M, Cao N, Yu S, Lou W. Findu: Privacy-preserving personal profile matching in mobile social networks. In: Proc. of the IEEE Int'l Conf. on Computer Communications (INFOCOM). 2011. 2435−2443. [doi: 10.1109/INFCOM.2011.5935065]

[9] Yan Z, Ding W, Niemi V. Two schemes of privacy-preserving trust evaluation. Future Generation Computer Systems, 2015,62(C): 175−189. [doi: 10.1016/j.future.2015.11.006]

[10] Kiraz MS, Genc ZA, Kardas S. Security and efficiency analysis of the hamming distance computation protocol based on oblivious transfer. Security & Communication Networks, 2015,8(18):4123−4135. [doi: 10.1002/sec.1329]

[11] Zhang R, Zhang J, Zhang Y, Sun J. Privacy-preserving profile matching for proximity-based mobile social networking. IEEE Journal on Selected Areas in Communications, 2013,31(9):656−668. [doi: 10.1109/JSAC.2013.SUP.0513057]

[12] Liao X, Uluagac S, Beyah RA. S-MATCH: Verifiable privacy-preserving profile matching for mobile social services. In: Proc. of the IEEE Int'l Conf. on Dependable Systems and Networks. 2014. 287−298. [doi: 10.1109/DSN.2014.37]

[13] Yi X, Bertino E, Rao FY. Practical privacy-preserving user profile matching in social networks. In: Proc. of the IEEE Int'l Conf. on Data Engineering. 2016. 373−384. [doi: 10.1109/ICDE.2016.7498255]

[14] Pattuk E, Kantarcioglu M, Ulusoy H. Optimizing secure classification performance with privacy-aware feature selection. In: Proc. of the IEEE Int'l Conf. on Data Engineering. 2016. 217−228. [doi: 10.1109/ICDE.2016.7498242]

[15] Jung T, Mao X, Li XY, Tang SW. Privacy preserving data aggregation without secure channel: multivariate polynomial evaluation. In: Proc. of the IEEE Int'l Conf. on Computer Communications (INFOCOM). 2013. 2634−2642. [doi: 10.1109/INFCOM.2013.6567071]

[16] Jung T, Li XY, Wan Z, Wan M. Control cloud data access privilege and anonymity with fully anonymous attribute based encryption. IEEE Trans. on Information Forensics and Security, 2015,10(1):190−199. [doi: 10.1109/TIFS.2014.2368352]

[17] Jung T, Li XY. Collusion-tolerable privacy-preserving sum and product calculation without secure channel. IEEE Trans. on Dependable and Secure Computation, 2015,12(1):45−57. [doi: 10.1109/TDSC.2014.2309134]

[18] Jung T, Li XY, Wan Z, Wan M. Privacy preserving cloud data access with multi-authorities. In: Proc. of the IEEE Int'l Conf. on Computer Communications (INFOCOM). 2013. 2625−2633. [doi: 10.1109/INFCOM.2013.6567070]

[19] Abawajy J, Ninggal MI, Herawan T. Privacy preserving social network data publication. IEEE Communications Surveys & Tutorials, 2016,18(3):1974−1997. [doi: 10.1109/COMST.2016.2533668]

[20] Yan Z, Ding W, Niemi V. Two schemes of privacy-preserving trust evaluation. Future Generation Computer Systems, 2016,62(C): 175−189. [doi: 10.1016/j.future.2015.11.006]

[21] Qian J, Qiu F, Wu F. Privacy-preserving selective aggregation of online user behavior data. 2016. 1. [doi: 10.1109/TC.2016.2595562]

[22] Kim M, Mohaisen A, Cheon JH. Private over-threshold aggregation protocols over distributed databases. IEEE Trans. on Knowledge & Data Engineering, 2016. 1. [doi: 10.1109/TKDE.2016.2572686]

[23] Tsai CH, Liu HW, Ku T. Personal recommendation engine of user behavior pattern and analysis on social networks. In: Proc. of the IEEE Int'l Conf. on Computational Science and Computational Intelligence (CSCI). 2015. 404−409. [doi: 10.1109/CSCI.2015.46]

[24] Mishra BK, Jha N. SEIQRS model for the transmission of malicious objects in computer network. Applied Mathematical Modelling, 2010,34(3):710−715. [doi: 10.1016/j.apm.2009.06.011]

[25] Sun J, Zhang R, Zhang Y. Privacy-preserving spatiotemporal matching. In: Proc. of the IEEE Int'l Conf. on Computer Communications (INFOCOM). 2013. 800−808. [doi: 10.1109/INFCOM.2013.6566867]

[26] Schweitzer N, Stulman A, Shabtai A. Mitigating denial of service attacks in OLSR protocol using fictitious nodes. IEEE Trans. on Mobile Computing, 2016,15(1):163−172. [doi: 10.1109/TMC.2015.2409877]

[27] Geer DE. Attack surface inflation. IEEE Educational Activities Department, 2011,9(4):85−86. [doi: 10.1109/MSP.2011.78]

[28] Najafabadi MM, Khoshgoftaar TM, Calvert C. Detection of SSH brute force attacks using aggregated netflow data. In: Proc. of the IEEE Int'l Conf. on Machine Learning and Applications (ICMLA). 2016. 283−288. [doi: 10.1109/ICMLA.2015.20]

[29] Niu B, Zhu X, Liu J. Weight-aware private matching scheme for proximity-based mobile social networks. In: Proc. of the IEEE Global Communications Conf. (GLOBECOM). 2013. 3170−3175. [doi: 10.1109/GLOCOM.2013.6831559]

[30] Zhang R, Zhang R, Sun J. Fine-grained private matching for proximity-based mobile social networking. In: Proc. of the IEEE Int'l Conf. on Computer Communications (INFOCOM). 2012. 1969−1977. [doi: 10.1109/INFCOM.2012.6195574]

附中文参考文献:

[1] 付艳艳,张敏,冯登国,陈开渠.基于节点分割的社交网络属性隐私保护.软件学报,2014,25(4):768−780. http://www.jos.org.cn/1000-9825/4565.htm [doi: 10.13328/j.cnki.jos.004565]

罗恩韬(1978－),男,湖南永州人,博士,教授,主要研究领域为移动社交网络隐私保护,云安全,大数据聚类分析.

王国军(1970－),男,博士,教授,博士生导师,主要研究领域为可信计算,净室安全计算,网络空间安全.

刘琴(1982－),女,博士,副教授,主要研究领域为云安全,信息安全,隐私保护.

孟大程(1994－),男,硕士,主要研究领域为移动医疗网络隐私保护,大数据安全.

唐雅媛(1982－),女,博士,副教授,主要研究领域为网络计算,人工智能,信息安全.