



移动社会网络中用户相似性计算的隐私保护研究^{*}

李金双, 李凤云

(东北大学计算中心 沈阳 110819)

摘 要: 移动社会网络(MSN)增强了虚拟世界与现实世界的联系,它允许移动用户发现其现实世界中附近的用户并与之交流。在 MSN 应用中,为提供更好的服务,通常会比较两个用户的相似性,但在计算用户相似性的过程中很有可能泄露用户的隐私信息。针对这一问题,设计了两种新的相似性计算协议,这些协议保证了在进行相似性计算时不会泄露用户的任何隐私信息。

关键词: 移动社会网络;相似性计算;同态加密;隐私

doi: 10.3969/j.issn.1000-0801.2013.10.021

Protected Private Information for Proximity Computing in Mobile Social Network

Li Jinshuang, Li Fengyun

(Department of Computing Center, Northeastern University, Shenyang 110819, China)

Abstract: Mobile social network (MSN) extend social network in the cyberspace into the real world. It allows mobile users to discover and interact with a stranger in their physical vicinity. Proximity computing means that two users can compare their personal profiles in MSN application. It is, however, possible disclosing user's privacy information. To address this concerns, two novel proximity computing protocols were designed. The protocols enable two users to perform proximity computing without disclosing any information about their profiles.

Key words: mobile social network, proximity computing, homomorphic encryption, private

1 引言

移动终端的随身、随时、便捷、不受环境约束等特点,给人们的工作、学习、娱乐生活带来了丰富多彩的资讯内容,基于移动终端的移动社会网络(mobile social network, MSN)能够为用户提供随时随地的服务。相对于普通的社会网络用户,MSN 用户提供的个人信息、日程活动、生活工作圈等信息更为真实,也更为丰富。在 MSN 中建立的朋友关系能够更真实地反映用户的实际交往信息。

基于位置的服务是移动网络的特色服务,它是通过移

动运营商的无线电通信网络(如 GSM 网、CDMA 网)或使用外部定位方式(如 GPS)获取移动终端用户的位置信息(地理坐标)。在应用软件的支持下,为用户提供个性化服务的一种应用业务。

MSN 所提供的一个重要的位置服务是允许移动用户发现在自己周围的其他用户(例如陌陌^[1]、微信^[2]的摇一摇)。为了提供更好的、更个性化的服务,在附近的众多用户当中,应用应尽可能优先选择与当前用户兴趣、爱好接近的用户推荐给当前用户,这类应用可通过计算用户的相似性的方式对用户的相似程度进行评价。用户相似性计算中的隐私

^{*} 教育部英特尔信息技术专项科研基金资助项目(No. MOE-INTEL-2012-06),东北大学基本科研业务费资助项目(No.N110216001)



保护问题在在线社会网络应用和 MSN 中都获得了广泛的关注,但在 MSN 中此问题的解决尤其紧迫,因为不可信用户直接参与到计算过程中,而在计算用户相似性的过程中,用户间必须传递用户的属性信息,如何在计算用户相似性的过程中避免泄露用户的属性信息是本文研究的重点。

2 相关工作

在基于位置服务的 MSN 应用中,移动用户与周围的其他陌生用户在实际面对面交流之前,通常会根据用户间的相似性进行筛选^[3]。而在没有第三方安全服务器参与的情况下,在相似性计算过程中必然要涉及的个人兴趣、爱好等隐私信息是否会泄露问题越来越受到研究人员的关注^[4-7]。参考文献^[8-11]均采用互斥的数据来描述用户的兴趣和爱好,因而无法区分用户爱好的程度,这显然不利于准确地反映用户的相似性。例如两个用户都设置了爱看电影的爱好,第一个用户每周看一次电影,第二个用户每月看一次电影,在前述的研究中都仅标注为用户爱看电影,对于爱好的程度是无法区分的。针对这一问题,本文设计了针对用户属性的细粒度描述(如属性值设为 $[0,10]$ 中的任意数值,0代表没有兴趣,10代表兴趣最大)的安全的相似性计算协议。

3 问题定义与加密方法

3.1 问题定义

寻求移动用户安全相似性计算方法如下。

在某 MSN 应用中,计算的发起者用户 A 拥有 n 个属性的属性集合 $U=\{u_1, u_2, \dots, u_n\}$, 计算的参与者用户 B 也拥有 n 个属性的属性集合 $V=\{v_1, v_2, \dots, v_n\}$, 提供一种方法,使得用户 A 和用户 B 在不向对方泄露相应的属性信息 u_i 和 v_i 的情况下,能够计算出其相似性值。

设 $F(U_S, V_S)$ 是一个由两个集合 U_S 和 V_S 到实数 R 的映射组成的映射集合,其中 $U_S \subseteq U, V_S \subseteq V$, 是参与相似性计算的属性集合。映射 $f \in F$, 定义下面两种隐私保护的用户相似性计算协议。

定义 1 (Level-1 级隐私保护用户相似性计算) 在用户相似性计算完成后,计算的发起者 A 知道映射 f 、用户 A 和用户 B 的相似性,计算的参与者 B 只知道映射 f , 双方均不知道对方参与计算的属性的具体值。

定义 2 (Level-2 级隐私保护用户相似性计算) 在用户相似性计算完成后,计算的发起者 A 知道映射 f 、用户 A 和用户 B 的相似性,计算的参与者 B 不知道任何信息,双

方均不知道对方参与计算的属性的具体值。

3.2 相似性计算方法

在基于位置的 MSN 应用中,由于当前用户网络结构信息并不完整,因此基于结构的相似性度量方法(如共同邻居法、最短路径法)不再适用。如前所述,在这类应用中用户的属性用集合表示,按约定的次序排列属性可得到属性的向量表示,因此可以采用向量的相似性度量方法计算用户间的属性相似性。计算两个向量之间的相似性主要有计算向量之间的夹角(如余弦相似性)和向量顶点间的距离(如欧几里得距离)两种方法。两种方式各有其弱点,如图 1(a)所示,以夹角为核心的相似性计算,由于向量 p 与向量 a 的夹角小于向量 p 与向量 b 之间的夹角,因此会认为向量 p 与 a 更为相似,但显然,向量 p 与向量 b 相似性更好;而如图 1(b)所示,以距离为核心进行相似性比较会认为向量 c 与向量 p 更为相似,因为它们之间的距离更近,但向量 p 和向量 c 在 x 轴的投影相反,表明它们之间差异更大,因此向量 p 与向量 d 更为相似。

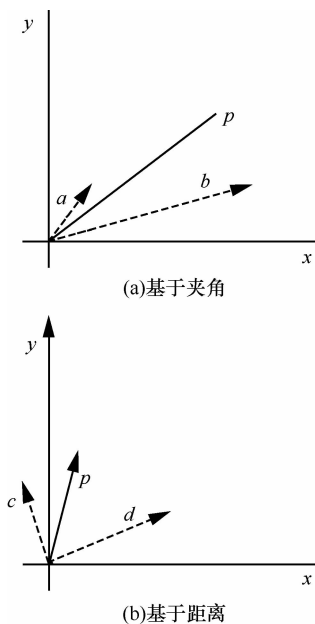


图 1 相似性计算方法

在 MSN 用户相似性度量中,因为属性值都设置为 $[0,d]$ 的数($d \in R^+$),因此所有用户的属性向量都位于第一象限,因此采用距离相似性计算效果更好,本文采用欧几里得距离的平方作为相似性评价的方式。

3.3 加密算法

本文采用同态加密(homomorphic encryption)技术,同态加密是一种加密形式,它允许人们对密文进行特定的代

数运算,得到的新的加密值与对明文进行同样的运算再将结果加密效果相同。同态加密技术是解决多方参与计算时隐私保护的一种有效的安全加密技术。

Paillier 加密系统^[12]是一种可靠的、性能优良的半同态加密技术,其要点如下。

密钥生成: 随机选择两个大素数 p 和 q , 使其满足 $\gcd(pg, (p-1)(q-1))=1$, 计算 $n=pq, \lambda=\text{lcm}(p-1, q-1)$, 随机选择整数 $g, g \in \mathbb{Z}_n^*$, 保证其满足 $\gcd(L(g^\lambda \bmod n^2), n)=1$, 其中 $L(x)=(x-1)/n$, 则该加密系统的公钥为 (g, n) , 私钥为 λ 。

加密方法: m 是要加密的明文, $m \in \mathbb{Z}_n$, 选择随机数 $r, r \in \mathbb{Z}_n^*$, 对应的密文为 $c=e(x)=g^m \cdot r^n \bmod n^2$ 。

解密方法: 对于密文 c , 解密方法为 $m=\frac{L(c^\lambda \bmod n^2)}{L(g^\lambda \bmod n^2)} \bmod n$ 。

Paillier 加密系统有以下两种同态特点。

(1) 加法同态

解密两个密文的乘积可以获得这两个密文中明文的和。

$$D(E(m_1, r_1) \cdot E(m_2, r_2)) \bmod n^2 = m_1 + m_2 \quad (1)$$

(2) 乘法同态

解密某个密文的 k 次幂可以得到密文中明文与 k 的积。

$$D(E(m, r)^k) \bmod n^2 = km \bmod n \quad (2)$$

4 隐私保护的用户相似性计算协议

针对两个层次的隐私保护定义提出了两个隐私保护的用户相似性计算协议。协议 1 以较小的计算量完成 Level-1 级隐私保护用户相似性计算, 协议 2 以较高的计算代价完成 Level-2 级隐私保护用户相似性计算。

设计算的发起者 A 与计算的参与者 B 参与相似性计算的属性个数为 m , 用户相似性计算采用欧几里得距离的平方表示, 计算式如式(3)所示:

$$D_{\text{sim}} = \sum_{i=1}^m (u_i - v_i)^2 \quad (3)$$

协议 1 对应于 Level-1 级隐私保护用户相似性计算

(1) 用户 A 取属性集 U_s 中的每一个元素 u_i , 采用 Paillier 加密系统加密, 得到加密后的属性集 $e(U_s)$ 。

(2) 用户 A 将加密后的属性集 $e(U_s)$ 、映射函数 f 、公钥信息 (g, n) 发送给 B。

(3) 用户 B 根据映射函数 f 的要求, 首先利用式(1)和式(2)计算:

$$e(u_i - v_i)^2 = (e(u_i) \cdot e(-v_i))^2$$

其中, $e(u_i) \in e(U_s), v_i \in V_s$ 。

然后利用式(1)和式(3)计算加密后的用户的相似性 $e(D_{\text{sim}})$:

$$e(D_{\text{sim}}) = \sum_{i=1}^m e(u_i - v_i)^2$$

(4) 用户 B 将 $e(D_{\text{sim}})$ 发送给 A。

(5) 用户 A 用其私钥 λ 解密 $e(D_{\text{sim}})$, 得到用户 A 和用户 B 的相似性度量值 D_{sim} 。

协议 2 对应于 Level-2 级隐私保护用户相似性计算

(1) 用户 A 根据映射函数 f , 将属性集 U 中不参与计算的属性值 u_i 设置为零, 得到计算属性集 U_c , 并产生计算屏蔽集 $U_t; U_t = \{u_i^f | u_i^f = 1, u_i \in U_s\}$ 。

(2) 用户 A 对计算属性集 U_c 、计算屏蔽集 U_t 的每一个元素采用 Paillier 加密系统加密, 得到相应的加密集 $e(U_c)$ 和 $e(U_t)$ 。

(3) 用户 A 将加密后的计算属性集 $e(U_c)$ 、计算屏蔽集 $e(U_t)$ 和公钥信息 (g, n) 发送给用户 B。

(4) 用户 B 利用式(2)计算得到计算属性集 V_c :

$$V_c = \{v_i^c | v_i^c = e(u_i^f v_i^f), u_i^f \in U_t, v_i^f \in V\}$$

(5) 用户 B 利用式(1)和式(2)计算:

$$e(u_i^c - v_i^c)^2 = (e(u_i^c) \cdot e(-v_i^c))^2$$

然后利用式(1)和式(3)计算加密后的用户的相似性 $e(D_{\text{sim}})$:

$$e(D_{\text{sim}}) = \sum_{i=1}^n e(u_i^c - v_i^c)$$

(6) 用户 B 将 $e(D_{\text{sim}})$ 发送给 A。

(7) 用户 A 用其私钥 λ 解密 $e(D_{\text{sim}})$, 得到用户 A 和用户 B 的相似性度量值 D_{sim} 。

在具体的 MSN 应用中, 系统采用上述两个协议之一计算所有邻近用户的相似性, 应用按相似性高低或者设定某一阈值来控制并显示推荐列表, 供用户选择。

5 安全性证明

定理 1 在 Paillier 加密系统是安全的情况下, 协议 1 能保证 Level-1 级隐私保护。

证明 用户 B 收到的是用 Paillier 加密系统加密后的用户 A 的属性集 $e(U_s)$ 以及用户 A 的公钥 (g, n) 和映射函数 f , 由于 Paillier 加密系统是安全的, 用户 B 在不知道用户 A 私钥 λ 的情况下无法通过 $e(u_i)$ 得出 u_i , 因此用户 B 仅知道映射规则 f 而不知用户 A 的属性信息; 对于用户 A 来说, 由于用户 B 返回的是 $e(D_{\text{sim}})$, 用户 A 无法猜出用户



B 的某一个具体的 v_i 值,因而用户 A 虽然可以知道双方的相似性,却不知道用户 B 的属性信息。

证毕。

定理 2 在 Paillier 加密系统是安全的情况下,协议 2 能保证 Level-2 级隐私保护。

证明 用户 B 收到的是用 Paillier 加密系统加密后的用户 A 的属性集 $e(U_e)$ 、加密后的屏蔽属性集 $e(U_f)$ 以及用户 A 的公钥 (g, n) , 由于 Paillier 加密系统是安全的, 用户 B 在不知道用户 A 私钥 λ 的情况下无法通过 $e(u_i^e)$ 和 $e(u_i^f)$ 得出 u_i^e 和 u_i^f , 用户 B 在进行求 V_e 的计算中, 由于 Paillier 加密系统的乘法同态特性, 不需要参与计算的属性值将被屏蔽(与零相乘), 然后再利用 Paillier 加密系统的加法同态性和乘法同态性计算出 $e(D_{sim})$, 用户 B 无法获知映射函数 f 的信息, 因此用户 B 既不知道映射函数 f 也不知道用户 A 的属性信息; 对于用户 A 来说, 由于用户 B 返回的是 $e(D_{sim})$, 用户 A 无法猜出某一个具体的 v_i 值, 因而用户 A 虽然可以知道双方的相似性, 却不知道用户 B 的属性信息。

证毕。

需要加以说明的是, 用户 A 可以通过只设定一个属性值来发起相似性计算, 并通过不断调整该属性值从而达到准确知道用户 B 相同属性的设定值, 只要允许用户 A 不断发起这样的计算, 用户 A 就可以知道用户 B 的所有属性信息。但这一问题并不会影响协议的应用, 只要在 MSN 应用中简单设置一个可以接受同一个用户发起相似性计算的时间阈值(如 10 min)即可, 从计算资源利用的角度来说, 这样的设置显然是合理的, 也是必要的。

6 结束语

随着移动社会网络服务应用的迅猛增长, 设计安全的用户间信息交流协议的需要日益紧迫。本文针对移动社会网络中基于位置的用户相似性计算中的属性信息泄露问题进行了研究, 提出了两种不同强度的隐私保护级别, 基于同态加密技术分别设计了两种级别的隐私保护相似性计算协议, 并给出了协议的安全性证明。

下一步的研究工作是在移动终端中开发实现上述协议的应用。

参考文献

- 1 陌陌. <http://www.immomo.com/>
- 2 微信. <http://weixin.qq.com/>

- 3 Li M, Yu S, Cao N, *et al.* Privacy-preserving distributed profile matching in proximity-based mobile social networks. *IEEE Transactions on Wireless Communications*, 2013, 12 (5): 2024~2033
- 4 Arb M V, Bader M, Kuhn M, *et al.* VENETA: serverless friend-of-friend detection in mobile social networking. *Proceedings of the WIMOB*, Avignon, France, 2008
- 5 Lu R, Lin X, Liang X, *et al.* A secure handshake scheme with symptoms-matching for mhealthcare social network. *Mobile Networks and Applications*, 2010, 16(6): 683~694
- 6 Dong W, Dave V, Qiu L, *et al.* Secure friend discovery in mobile social networks. *Proceedings of the INFOCOM*, Shanghai, China, 2011
- 7 Li M, Cao N, Yu S, *et al.* FindU: privacy-preserving personal profile matching in mobile social networks. *Proceedings of the INFOCOM*, Shanghai, China, 2011
- 8 Jahanbakhsh K, King V, Shoja G C. Predicting missing contacts in mobile social networks. *Pervasive and Mobile Computing*, *Mobile and Multimedia Networks*, 2012, 8(5): 698~716
- 9 Freedman M J, Nissim K, Pinkas B. Efficient private matching and set intersection. *Proceedings of the EUROCRYPT*, Interlaken, Switzerland, 2004
- 10 Kissner L, Song D. Privacy-preserving set operations. *Proceedings of the CRYPTO*, Santa Barbara, CA, 2005
- 11 Ye Q, Wang H, Pieprzyk J. Distributed private matching and set operations. *Proceedings of the ISPEC*, Sydney, Australia, 2008
- 12 Paillier P. Public-key cryptosystems based on composite degree residuosity classes. *Proceedings of the EUROCRYPT*, Prague, Czech Republic, 1999

[作者简介]



李金双, 男, 博士, 东北大学计算中心副教授, 主要研究方向为网络与信息安全、社会计算。



李凤云, 女, 博士, 东北大学计算中心讲师, 主要研究方向为网络与信息安全、无线传感器网络。

(收稿日期: 2013-07-30)