doi:10.11959/j.issn.1000-436x.2016164

基于隐私匹配的服务代理发现方法

耿魁[1]，万盛[1]，李凤华[1,2]，何媛媛[2]，王瀚仪[2]
(1. 西安电子科技大学综合业务网理论与关键技术国家重点实验室，陕西 西安 710071；
2. 中国科学院信息工程研究所信息安全国家重点实验室，北京 100093)

摘 要：针对代理发现中用户对代理的性能、成本和安全性等方面的需求，以及需求匹配过程中的隐私保护问题，基于 Paillier 同态加密算法，提出一种新的综合考虑代理和用户属性及其偏好的私有数据信息匹配算法，包括建立基于欧氏距离的相似度函数、利用加密算法进行匹配、计算相似度和确定匹配的代理链 4 个步骤。该算法引入半可信主代理从全局层面管理所有子代理的业务类型和连接状况，并承担主要的计算开销，同时将欧氏距离与 Paillier 同态加密算法有机结合，支持具有偏好信息的多元属性数据匹配，能够有效保障用户和子代理的安全性。最终，通过安全性分析与性能仿真，证明所提出方案的安全性和有效性。

关键词：代理；服务代理发现；隐私匹配；同态加密
中图分类号：TP393 **文献标识码**：A

Privacy matching-based service proxy discovery scheme

GENG Kui[1], WAN Sheng[1], LI Feng-hua[1,2], HE Yuan-yuan[2], WANG Han-yi[2]
(1.State Key Laboratory of Integrated Service Networks,Xidian University, Xi'an 710071, China;
2.State Key Laboratory of Information Security, Institute of Information Engineering, CAS, Beijing 100093, China)

Abstract: According to user's requirements on the proxy performance, cost and safety in proxy discovery, and the privacy-preserving issue during the process of the demand private matching, a new private matching algorithm was presented based on Paillier homomorphic encryption, comprehensively considered the attributes of user and agents and their priority. It included four steps: building the similarity function based on Euclidean distance, carrying on the private matching by encryption algorithm, calculating the similarity and screening the proxy chain. Proposed scheme introduced semi trusted primary proxy from the global level, which is to manage all the sub proxy's business type and the connection status, and do the main computational overhead.At the same time, Euclidean distance and Paillier homomorphic encryption algorithm were combined to support the multivariate attributes with priority to match, which can effectively protect the privacy of user and sub proxy. Finally, security analysis and evaluation results show the effectiveness and safety of proposed scheme.

Key words: proxy, service proxy discovery, privacy matching, homomorphic encryption

1 引言

随着网络技术的快速发展和信息共享系统的大量应用，泛在网络中局域网、互联网、物联网和移动网等多种网络广泛互连互通，信息呈爆炸式增长，出现了过载现象，服务数量的大规模性、服务描述的异构性以及设备服务的资源高度受限性和移动性等特征日益凸显，为了向用户准确地提供所

收稿日期：2016-03-04；**修回日期**：2016-07-28
通信作者：李凤华，lfh@iie.ac.cn
基金项目：国家自然科学基金委—广东联合基金资助项目（No.U1401251）；国家高技术研究发展计划（"863"计划）基金资助项目（No.2015AA016007）；国家自然科学基金资助项目（No.61502489）
Foundation Items: The Key Program of the National Natural Science Foundation of China-Guangdong Union Foundation (No.U1401251), The National High Technology Research and Development Program of China (863 Program) (No.2015AA016007), The National Natural Science Foundation of China (No.61502489)

需的信息和服务，朋友发现、代理发现和服务发现等基于匹配的应用应运而生，成为泛在网络的重要应用，已受到了学术界和工业界的广泛关注。用户与活跃在邻近区域的代理及时通信，找到在性能、成本和安全性等方面最优的路径，更快速、更准确地获取信息和服务。然而，用户和代理通过交换服务属性集合优化访问路径的同时，用户和代理的处理能力、时延、能耗、服务费用和抗攻击能力等敏感信息将暴露给对方，从而为双方的数据安全和隐私信息带来潜在威胁。因此，对用户个人隐私和代理敏感数据的保护刻不容缓。

针对私有数据匹配问题，近年来大量文献提出了解决方法。大多数方法假设每个匹配参与方拥有一个属性集合，采用匹配参与双方共同属性的个数为相似度，共同属性越多，用户越匹配，整个匹配过程不泄露双方个人敏感信息，但是，这些方案存在一定的局限性。匹配函数只考虑了共同属性的个数，忽视了每个用户对各个属性偏好程度的差异；基于交换加密、同态加密等加密方法的隐私信息匹配方案安全性较高，但是，计算开销较大，难以在移动终端上运行；基于布隆过滤器、混淆等非加密方法的隐私信息匹配方案计算开销较小，但是，大量随机数导致通信量大幅增加，并且安全性较弱。因此，将复杂的计算工作交由第三方代理来完成是一种常见的解决思路。

为了解决上述问题，本文提出一种与属性偏好相关的私有数据信息匹配算法，其主要贡献如下。

1) 将 Paillier 同态加密应用于代理发现算法的设计中，针对用户的个性化服务请求提供最匹配的作业代理链。

2) 引入主代理从全局层面管理所有子代理的业务类型和连接状况，并在利用 Paillier 同态加密方法的语义安全性保证算法安全性的前提下，由主代理承担主要的计算开销，以缓解用户终端的计算压力。

3) 支持具有偏好信息的多元属性数据匹配，更细粒度地满足用户及代理在服务性能和隐私保护方面的需求。同时，执行加密运算前将属性的偏好程度转化为二元值，有助于降低计算开销。

2　相关工作

本文涉及的代理发现这一概念，是建立在第三方不可信的基础上进行的，因而本文方案的主要工作是在隐私保护的前提下实现代理发现。能够保障信息不被泄露的算法有很多种，大体分为加密算法和非加密算法，本文采用的是同态加密算法中的 Paillier 算法，在方案中起到了隐私保护的关键作用。

有不少学者会通过非加密的算法实现隐私保护，以便减轻执行过程中的消耗。Zheng 等[1]提出了一个基于位置隐私保护的握手协议，利用布隆过滤器进行一对多的匹配，构建时空位置标记，从无线电在设备的周围环境中捕获的信号（Wi-Fi 等）来标记位置，以防止位置欺骗。Sun 等[2]也根据布隆过滤器进行基于共同兴趣爱好的朋友隐私匹配，虽然上述方案均未采用加密算法，使整体方案的消耗很小，但是却有匹配精确度的偏差。Lu 等[3]针对移动医疗急救场景提出 SPOC 协议，并在方案中引入随机数，通过混淆来保证隐私安全，然而利用穷举法，还是有信息被破解的可能。

出于对安全性的考虑，更多学者会利用加密算法来实现高强度的隐私保护。例如，Agrawal 等[4]提出了基于交换加密的双方秘密共享方案，Vaidya 等[5]提出了 N 方秘密共享方案，Niu 等[6]提出 P-match，利用交换加密进行基于邻近的朋友发现，并考虑到用户的兴趣权重因素。而在各种加密方案中，同态加密方法更是不胜枚举，Li 等[7]利用同态加密和二维多项式，实现多方匹配的语义安全。Kerschbaum 等[8]也利用同态加密，在双方不泄露个人信息的前提下求出集合交集。Li 等[9]基于邻近的移动社交网络，提出了分布式的保护隐私信息匹配方案 FindU，建立 3 个不同的隐私级别，利用 Shamir 秘密共享方法结合同态加密，进行安全多方计算，实现多方匹配，可以抵抗多方同谋的情形。Zhang 等[10]利用 Paillier 同态加密方法，根据不同的隐私需求设计不同的方案，并考虑到了用户对不同兴趣的偏好程度，由此建立了基于邻近的隐私信息匹配协议。Thapa 等[11]针对在线社交网络，利用 Paillier 加密算法设计了 3 个不同隐私级别的非对称的基于邻近的隐私信息匹配方案，其社交网络中的用户不仅通过朋友关系彼此相连，还按社会团体进行分类，并提出新的相似度函数。Chu 等[12]提出了隐私保护的二部匹配框架，基于二部图结构中的匈牙利方法和 Paillier 同态加密方法，来寻找最佳匹配，在保障用户隐私的前提下，利用服务器的计算能力对多媒体进行分析研究。文献[13]提出了一对一基于

社交网络隐私信息匹配的方案，运用 ElGamal 和 Paillier 加密算法，使用户双方可以不同时在线进行匹配。Wang 等[14]针对社交网络中的群组隐私匹配，利用 Paillier 加密算法，选出与用户的兴趣集合最匹配的群组。Niu 等[15]同样基于 Paillier 算法，考虑到用户属性的权重来进行朋友发现。上述在不同场景下的文献，均采用了 Paillier 加密算法，可见该算法的实用性与安全性。

3 隐私匹配机制

3.1 预备知识

本文采用同态加密算法——Paillier 加密算法，来保护用户和各个子代理的隐私信息，由于 Paillier 算法对数据加密后，具有同态性质，所以可直接通过对加密后的信息进行操作，而不需要已知明文，这样保障了在交换信息过程中的安全性。下面介绍 Paillier 算法的具体操作。

1) 密钥生成。任意选取 2 个大素数 p、q，使 $GCD(pq,(p-1)(q-1))=1$。记 $N=pq$，$\lambda=LCM(p-1,q-1)$。选取随机整数 $g\in\mathbb{Z}^*_{N^2}$，使 $GCD(L(g^\lambda \bmod N^2),N)=1$，其中 $L(x)=\dfrac{x-1}{N}$。则用户 U 的公钥为数对<N,g>，私钥为 λ。

2) 加密。设需要加密的信息为 $k\in\mathbb{Z}_N$，加密方选取随机数 $r\in\mathbb{Z}^*_N$，则可得到加密后的密文，记作 $E(k \bmod N, r \bmod N)=g^k r^N \bmod N^2$。

3) 解密。设需要解密的密文为 $c\in\mathbb{Z}^*_{N^2}$，则解密方通过公钥与私钥，计算出明文，记作 $D(c)=\dfrac{L(c^\lambda \bmod N^2)}{L(g^\lambda \bmod N^2)} \bmod N$。

对任意 k_1，k_2，r_1，$r_2\in\mathbb{Z}_N$，Paillier 算法具有以下性质。

1) $E(k_1,r_1)E(k_2,r_2)=E(k_1+k_2,r_1r_2) \bmod N^2$。

2) $E^{k_2}(k_1,r_1)=E(k_1k_2,r_1^{k_2}) \bmod N^2$。

3) $E(k_1,r_1)r_2^N=E(k_1k_2,r_1^{k_2}) \bmod N^2$。

3.2 参数设置

用户在请求服务时，请求信息交给第三方代理后（主代理），主代理将对服务请求信息进行分解，服务请求将被多个不同类型、不同级别的子代理进行处理，形成一个作业代理链。根据不同下级子代理的类型、级别、业务能力以及代理之间的连通性，实际上可选的下级子代理有多种方案，为完成用户请求作业的下级代理所组成的作业代理链也存在多条。

在具体执行时，主代理可通过子代理之间的业务属性连通性，发现多条可完成用户服务请求的作业代理链。主代理需要根据用户提出的服务要求（例如时延、费用、安全等级等），以及各个子代理的服务属性值，找出一条最符合用户需求的作业代理链。

假设用户 U 有 m 个服务要求，记为 $(Q_1,Q_2,\cdots,Q_m)$。用户 U 针对每项服务要求 Q_j，用正整数来衡量其服务属性值，记为 $u_j\in[0,w-1]$，w 为约定的最大取值，则用户 U 的整体服务属性值记为 $U=\langle u_1,u_2,\cdots,u_m\rangle$。对于每条作业代理链，设其上含有 n 个子代理 $(V_1,V_2,\cdots,V_n)$，对不同的作业代理链 n 值可以不同，而每个子代理 V_i 针对服务要求 $(Q_1,Q_2,\cdots,Q_m)$ 的整体服务属性值记为 $V_i=\langle v_{i,1},v_{i,2},\cdots,v_{i,m}\rangle$，$v_{i,j}\in[0,w-1]$。

对任意 $x\in[0,w-1]$，定义函数 $h(x)=\langle x_1,x_2,\cdots,x_{w-1}\rangle$，其中，当 $1\leqslant l\leqslant x$ 时，$x_l=1$；当 $x\leqslant l\leqslant w-1$ 时，$x_l=0$。定义用户 U 的服务属性权重向量为 $\hat{\boldsymbol{u}}=\langle h(u_1),h(u_2),\cdots,h(u_m)\rangle=<\hat{u}_1,\ \hat{u}_2,\cdots,\hat{u}_{(w-1)m}>$，每个子代理 V_i 的服务属性权重向量为 $\hat{\boldsymbol{v}}_i=\langle h(v_{i,1}),h(v_{i,2}),\cdots,h(v_{i,m})\rangle=\langle \hat{v}_{i,1},\ \hat{v}_{i,2},\cdots,\ \hat{v}_{i,(w-1)m}\rangle$ 由定义可知，对 $\forall 1\leqslant i\leqslant n$，$1\leqslant s\leqslant (w-1)m$，$\hat{v}_{i,s}$、$\hat{u}_s$ 均为 0 或 1。

本文将用户 U 与子代理 V_i 的服务属性权重向量，在不泄露具体信息的情况下,借助主代理进行匹配，使用户得到评价值 $f(U,V_i)$，算出用户与整条作业代理链的整体评价值，并与其他作业代理链进行比较。评价值越小，表示该作业代理链与用户的服务越接近。本文中选取 $f(U,V_i)=\sum_{j=1}^{m}u_i-v_{i,j}$ 作为衡量评价值的函数。

在本方案中，忽略通信信道的安全问题，着重考虑来自内部参与者的攻击。假定内部参与信息交换传输的三方：用户、主代理、多个子代理，是诚实但好奇（HBC, honest but curious），即参与者能诚实的执行既定方案，不会对信息进行伪造，但却会尽可能分析方案执行过程中所有获得的信息，试图获取其他参与者的隐私信息。

基于上述攻击模型，当本文方案执行结束，用户、主代理、多个子代理，不能知道关于其他参与

者服务属性的任何信息。

3.3　系统架构

本文提出基于 Paillier 同态加密算法的作业代理链发现方法，其系统架构如图 1 所示。以用户通过一个不完全可信的第三方代理来完成餐饮、医疗等服务，下面举例对系统中相关背景进行介绍。

1) 主代理需要先对用户的请求服务进行分解。例如用户准备吃午饭，并给出了其对价钱、时间的要求，主代理则需要为用户查找交通类和餐饮类的子代理来完成该工作。

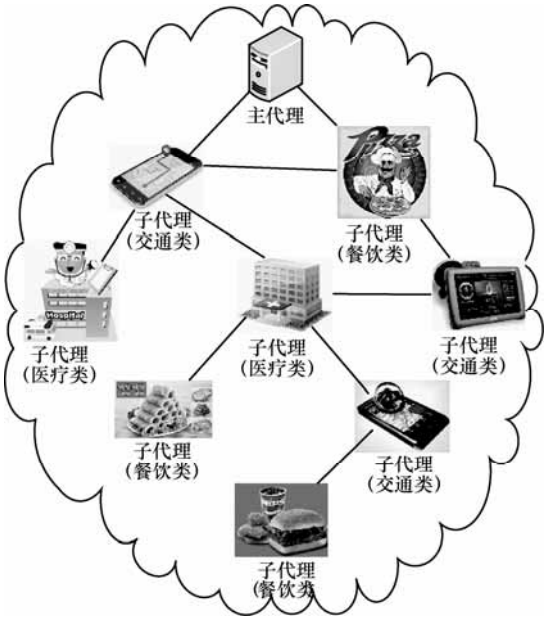

图 1　系统架构

2) 主代理通过广播方式，将其需要交通类和餐饮类的信息发送给其邻近的子代理，各个子代理根据自身的业务类型和连通性，将确认信息反馈给主代理。

3) 主代理根据接收的信息，综合出可选的作业代理链，然后根据同态加密算法，对用户和子代理的密文信息进行处理，并根据用户最终反馈的结果选择最合适的作业代理链来完成用户的请求作业。

3.4　匹配算法

由于用户与每个子代理的匹配过程都是相同的，所以为了简便，下面以子代理 V_i 为例，来对匹配过程进行说明。设子代理 V_i 的整体服务属性值为 $\langle v_{i,1}, v_{i,2}, \cdots, v_{i,m} \rangle$，服务属性权重向量为

$$\hat{\boldsymbol{v}}_i = \langle h(v_{i,1}), h(v_{i,1}), \cdots, h(v_{i,1}) \rangle = \langle \hat{v}_{i,1},\ \hat{v}_{i,2}, \cdots,\ \hat{v}_{i,(w-1)m} \rangle$$

显然，

$$f(U,V_i) = \sum_{j=1}^{m} |u_i - v_{i,j}| = \sum_{s=1}^{(w-1)m} |\hat{u}_s - \hat{v}_{i,s}|$$

因为 $|\hat{u}_s - \hat{v}_{i,s}|$ 等于 1 或 0，所以

$$\begin{aligned} f(U,V_i) &= \sum_{s=1}^{(w-1)m} |\hat{u}_s - \hat{v}_{i,s}|^2 \\ &= \sum_{s=1}^{(w-1)m} \hat{u}_s^{\,2} - 2\sum_{s=1}^{(w-1)m} \hat{u}_s\ \hat{v}_{i,s} + \sum_{s=1}^{(w-1)m} \hat{v}_{i,s}^{\,2} \\ &= \sum_{s=1}^{(w-1)m} \hat{u}_s^{\,2} + \sum_{s=1}^{(w-1)m} \hat{v}_{i,s}^{\,2} - 2\hat{\boldsymbol{u}} \end{aligned}$$

为保护用户和各个子代理的隐私信息，本文通过同态加密，用户使用自身的加密公钥对服务要求进行加密，同时，各个作业代理链上的子代理也使用用户的加密公钥加密自身的服务属性值。用户 U 和子代理 V_i 分别已知 $\sum_{s=1}^{(w-1)m} \hat{u}_s^{\,2}$ 和 $\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}^{\,2}$，为了让用户知道 $f(U,V_i)$，只需要在保证隐私的前提下，让用户计算 $\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}^{\,2} - 2\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}}_i$ 的数值即可，由此衍生出本文的具体方案，操作如下。

1) 用户 U 首先用 Paillier 加密算法给自己的服务属性权重向量加密。U 的服务属性权重向量为 $\hat{\boldsymbol{u}} = \langle h(u_1), h(u_2), \cdots, h(u_m) \rangle = \langle \hat{u}_1,\ \hat{u}_2, \cdots,\ \hat{u}_{(w-1)m} \rangle$，由定义可知，$\hat{u}_s = 1$ 对任意 $s \in I_u = \{s \mid (j-1)(w-1) < s \leqslant (j-1)(w-1) + u_j, 1 \leqslant j \leqslant m\}$ 成立，而当 $s \notin I_u$ 时，$\hat{u}_s = 0$。

U 选择一组随机数 $r_s \in \mathbb{Z}_N$，利用 U 的公钥 pub=<N,g>，对任意 $s \in [1, (w-1)m]$ 计算 $E(\hat{u}_s, r_s)$，然后将这组被加密的数据 $\{E(\hat{u}_s, r_s)\}_{s=1}^{(w-1)m}$、公钥 pub 以及用户的服务要求 *ReqData* 传送给主代理。

2) 在收到用户的服务请求信息 *ReqInfo*=<*ReqData*, $\{E(\hat{u}_s, r_s)\}_{s=1}^{(w-1)m}$, *pub*>后，主代理根据请求信息 *ReqInfo* 以及各个子代理的业务属性、连通性，采用广播方式，获取到多条作业代理链，对于每条代理链上的每个子代理 V_i，进行如下操作。

3) 主代理将 *ReqInfo* 发送给 V_i，V_i 对自己的服务属性权重向量进行处理。V_i 的服务属性权重向量为 $\hat{\boldsymbol{v}}_i = \langle h(v_{i,1}), h(v_{i,1}), \cdots, h(v_{i,1}) \rangle = \langle \hat{v}_{i,1},\ \hat{v}_{i,2}, \cdots,\ \hat{v}_{i,(w-1)m} \rangle$，其中，$\hat{v}_{i,s} = 1$ 对任意 $s \in I_{v_i} = \{s \mid (j-1)(w-1) < s \leqslant (j-1)(w-1) + v_{i,j}, 1 \leqslant j \leqslant m\}$ 成立，而当 $s \notin I_{v_i}$ 时，$\hat{v}_{i,s} = 0$。

子代理利用服务请求信息 *ReqInfo*，计算

$$E(\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}}_i, \prod\nolimits_{s \in I_{v_i}} r_s) = E\ (\sum\nolimits_{s \in I_{v_i}} \hat{u}_s, \prod\nolimits_{s \in I_{v_i}} r_s)$$

$$=\prod_{s\in I_{v_i}} E(\hat{u}_s, r_s) \bmod N^2$$

子代理 V_i 选取随机数 $d_i \in \mathbb{Z}_N$，利用用户公钥 <N, g>对 $\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2$ 进行加密，得到 $E(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2, d_i)$。子代理将 $E(\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i, \prod_{s\in I_{v_i}} r_s)$ 和 $E(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2, d_i)$ 的值发送给主代理。

4) 主代理接收子代理的数据后计算

$$E((N-2)\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i, e_i) = E^{N-2}(\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i, \prod_{s\in I_{v_i}} r_s) \bmod N^2$$

其中，$e_i = (\prod_{s\in I_{v_i}} r_s)^{N-2} \bmod N$。然后得到

$$E(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2 - 2\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i, d_i e_i)$$
$$= E(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2 + (N-2)\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i, d_i e_i)$$
$$= E(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2, d_i)\ E((N-2)\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i, e_i) \bmod N^2$$

在此之后，主代理将每条作业代理链上所有子代理的信息相乘，计算

$$E(\sum_{i=1}^{n}(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2 - 2\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i), \prod_{i=1}^{n} d_i e_i)$$
$$= \prod_{i=1}^{n} E(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2 - 2\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i, d_i e_i)$$

并将结果发给用户。

5) 用户使用自己的私钥，解密

$$E(\sum_{i=1}^{n}(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2 - 2\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i), \prod_{i=1}^{n} d_i e_i)$$

得到

$$\sum_{i=1}^{n}(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2 - 2\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i)$$

因为用户已知 $\sum_{s=1}^{(w-1)m} \hat{u}_s{}^2$，因此可以直接求出用户对该条作业代理链的总体评价值

$$f(U;V_1,V_2,\cdots,V_n) = \frac{1}{n}\sum_{i=1}^{n} f(U,V_i)$$
$$= \frac{1}{n}\sum_{i=1}^{n}(\sum_{s=1}^{(w-1)m} \hat{u}_s{}^2 + \sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2 - 2\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i)$$
$$= \sum_{s=1}^{(w-1)m} \hat{u}_s{}^2 + \frac{1}{n}\sum_{i=1}^{n}(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2 - 2\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i)$$

6) 设有 h 条作业代理链，第 i 条代理链的总评价值记为 $f_i(U)$，则用户将所有的作业代理链的总评价值进行比较，选取最小的第 z 条作业代理链，则 $z = \arg(\min\{f_i(U)\}_{i=1}^{h})$，并将第 z 条代理链的信息反馈给主代理，主代理将选用该作业代理链来处理用户的请求数据 *ReqData*。

4 安全性分析与性能仿真

4.1 安全性分析

在方案执行之后，用户将会得知每条作业代理链的评价值，子代理和主代理不会得知相关的有效信息。

1) 用户 U 的安全性

在本方案中，用户只将公钥 *pub* 和加密信息 $\{E(\hat{u}_s, r_s)\}_{s=1}^{(w-1)m}$ 透露给了主代理和子代理，后者并不知道用户的密钥，而 Paillier 同态加密算法是语义安全[16]的，因此在不知道密钥的前提下，主代理和子代理很难破解出用户的原有信息。

2) 子代理的安全性

在整个方案中，子代理 V_i 将 $E(\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i, \prod_{s\in I_{v_i}} r_s)$ 和 $E(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2, d_i)$ 透露给主代理，如上所述，主代理在不清楚私钥的情况下不能得到子代理和主代理的服务属性信息以及它们之间的共有信息。

而知道私钥的用户 U，只从主代理处得知 $E(\sum_{i=1}^{n}(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2 - 2\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i), \prod_{i=1}^{n} d_i e_i)$，密钥解密后，得到 $\sum_{i=1}^{n}(\sum_{s=1}^{(w-1)m} \hat{v}_{i,s}{}^2 - 2\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i)$ 的值。尽管用户不知道 $\hat{v}_{i,s}{}^2$、$\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i$ 具体的值，但是如果已知相互独立的 $(v_{i,1}, v_{i,2}, \cdots, v_{i,m}), 1 \leqslant i \leqslant n$，就可以根据定义得到 $\{\hat{v}_{i,s}{}^2\}_{s=1}^{(w-1)m}$ 和 $\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{v}}_i$ 的值，所以实质上，用户得到的是一个含有 mn 个未知变量的方程，每个未知变量都有 w 个可能取值，所以，如果用户是恶意的，想要通过匹配探测子代理的信息，则 U 可能需要从 w^{mn} 次可能取值中筛选出子代理的信息，而它的计算量无疑是大量的，因此子代理的关键信息很难被用户知晓。

根据上述分析，在主代理不一定可信的情况下，用户和主代理得不到子代理的服务属性值，除了用户的加密公钥信息，子代理和主代理也无法拿到用户的服务属性取值。

4.2 结果分析

假设在本文的方案中，通过主代理，从 k 个子代理中发现了 h 条服务链，并设用户和子代理均有 m 个服务要求，每个要求的偏好取值区间为 [0,w−1]。在整个方案中，根据 Paillier 同态加密算法的特点，取模数 N 的长度为 1 024 bit，则主要涉及到 1 024-bit 模乘、2 048-bit 模乘、1 024-bit 模幂和 2 048-bit 模幂的运算，分别记为 mul_1、mul_2、$\exp_1$、

exp_2。考虑到主代理的运行配置远优于用户和子代理的时间，因此将主代理的上述运算记为 mul_3、mul_4、exp_3、exp_4，以示区别。本文选取了文献[17]和文献[18]的方案，从离线计算量、在线计算量、通信量和执行时间的角度与本方案来量化对比。由于文献[17]与文献[18]的方案均为两方匹配，所以为使 3 个方案具有可比性，考虑在加入第三方的条件下使发起方与 k 个响应者进行匹配，3 个方案理论上的计算开销和通信量如表 1 所示。

本文的仿真运算将以便携式计算机（处理器：Intel Core P8700，2.53 GHz；内存：4 GB RAM；操作系统：Ubuntu 14.04）作为用户和子代理的运行配置，以服务器（处理器：Intel Xeon E5-2690，2.9 GHz；内存：32 GB RAM；操作系统：CentOS 6.3）作为主代理的运行配置。在便携式计算机上执行 mul_1、mul_2、exp_1、exp_2 运算，以及在服务器上执行 mul_3、mul_4、exp_3、exp_4 运算的最大时间、最小时间、平均时间、中位数以及标准差如表 2 和表 3 所示。

根据实际情况，取 w=10。图 2(a)表示的是方案离线的计算量，根据理论值可以看出，离线计算量只与 m 有关，并且本文方案与文献[18]方案计算量相同，略小于文献[17]方案。由于在整个过程中，只有用户进行了离线计算，所以该取值只与服务要求个数 m 有关，使 m 在 20 到 100 之间变化，从图中可以看出，随着 m 值的增大，整个方案的计算量也在不断增加，但是 3 个方案的时间差别并不大，通过计算可以知道，当 m=100 时，离线计算时间低于 5.2 s。

图 2(b)表示的是 3 种方案在线的计算量与子代理个数 k 的关系，本文使 k 值在 1 000 到 6 000 之间变动。显然，k 值增加，在线的计算时间也增长，尽管在线计算开销很大，但是本文方案中主代理相应的承担了绝大部分的计算，减轻了子代理和用户的负担，因此从图中可以明显的看出本文方案的计算量小于文献[17,18]方案的计算开销。当 k=6 000 时，本方案在线计算时间约为 2.3 s，文献[17,18]方案在线时间分别约为 3.1 s 和 3 s。

表 1　　**计算开销与通信量**

方案	参与方	离线计算开销	在线计算开销	通信量/bit
本文方案	用户	$2m(w-1)exp_1+m(w-1)mul_2$	$hexp_2$	$m(w-1)2\,048+1\,184$
	主代理	—	$kexp_4+kmul_4$	$k[m(w-1)2\,048+1\,184]+2\,048h$
	子代理	—	$m(w-1)mul_2+2exp_1$	4 096
文献[17]方案	用户	$2[m(w-1)+1]exp_1+[m(w-1)+1]mul_2$	$kmul_2,kexp_2$	$m(w-1)2\,048$
	主代理	—	—	$k[m(w-1)+1]2\,048$
	子代理	$m(w-1)mul_1$	$m(w-1)exp_2+m(w-1)mul_1+2exp_1+mul_2$	2 048
文献[18]方案	用户	$2m(w-1)exp_1,m(w-1)mul_2$	$kexp_2$	$m(w-1)2\,048+1\,184$
	主代理	—	—	$k[m(w-1)2\,048+1\,184]+2\,048k$
	子代理	—	$m(w-1)mul_2+exp_1$	2 048

表 2　　**便携式计算机运算执行时间**

运算	平均值	最大值	最小值	中位数	标准差
mul_1/μs	0.352 3	0.43	0.33	0.34	1.3×10^{-3}
mul_2/μs	2.562	2.64	2.53	2.55	5.4×10^{-4}
exp_1/ms	2.833	3.02	2.79	2.828	0.097
exp_2/ms	5.022	10.535	4.357	4.373	2.917

表 3　　**服务器运算执行时间**

运算	平均值	最大值	最小值	中位数	标准差
mul_3/μs	0.272 6	0.90	0.21	0.23	0.014 2
mul_4/μs	2.083	4.81	1.67	1.75	0.466 7
exp_3/ms	1.957	3.69	1.475	1.713	0.31
exp_4/ms	3.76	6.9	2.92	3.24	0.76

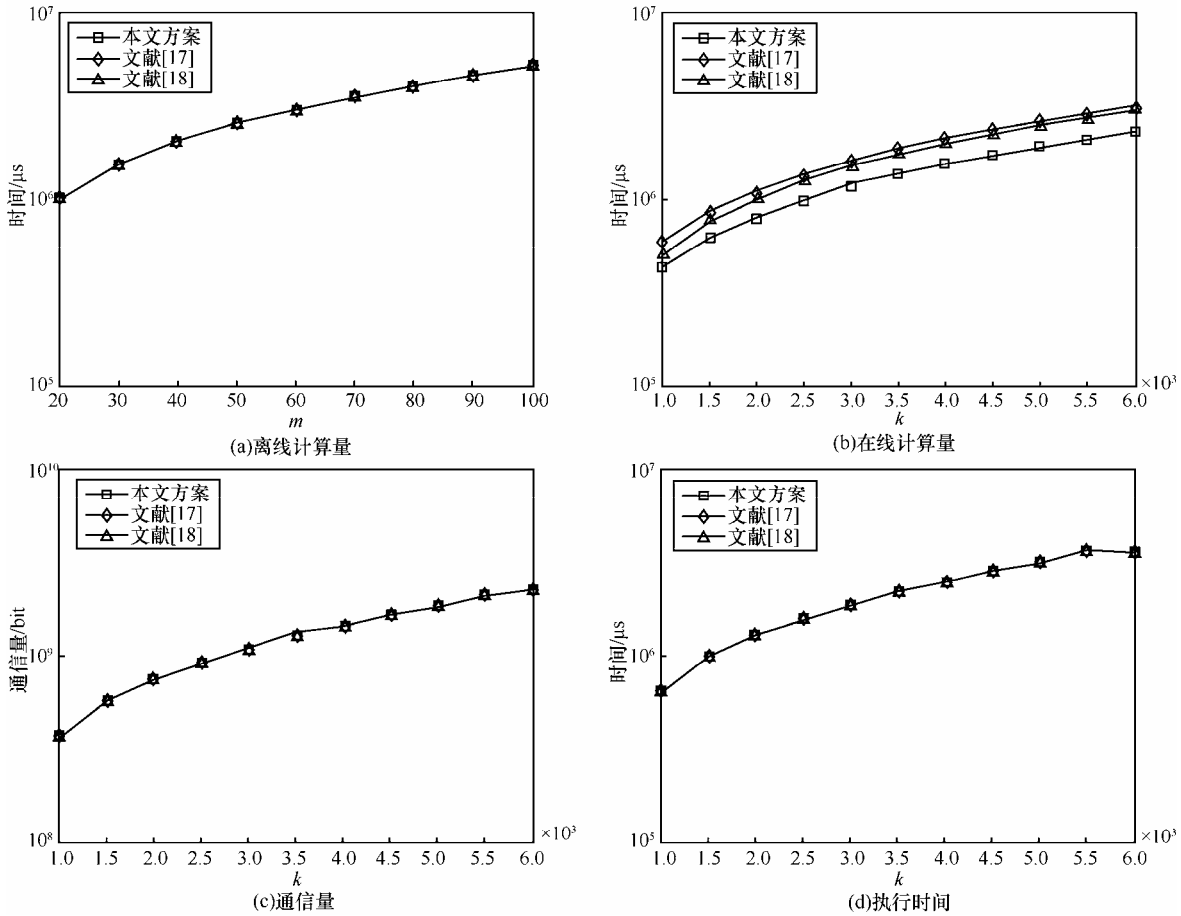

图 2　不同参数取值下的不同方案间性能比较

图 2(c)表示的是 3 种方案的通信开销与 k 的关系图，k 在 1 000 到 6 000 之间变动，随着 k 的增加，传输消耗不断增加，从图中难以分辨 3 种方案的差别，利用理论数据计算，当 k=6 000 时，本方案、文献[17]和文献[18]的方案通信量分别为 2.219 3×10^9 bit、2.224 5×10^9 bit 和 2.231 6×10^9 bit，显然，本文方案的通信量小于其他两者。

图 2(d)表示 3 种方案执行时间随 k 的变化，图像趋势同样为递增，虽然对比方案与本文方案执行时间处于同一量级，但是具体数值还是有所差别，同样通过理论数值计算，得到当 k=6 000 时，本文方案、文献[17]和文献[18]的方案的执行时间分别约为 3.798 2 s、3.819 2 s 和 3.829 8 s。并且尽管基于大传输量的条件，总体的执行时间仍然在可控范围内。

根据对理论值与图像的分析，方案消耗与服务要求的个数和子代理的个数均为正相关。本文方案通过增加第三方主代理，转移了用户和子代理的消耗，可以使整体的消耗和执行时间均有所减少，并与其他方案比较，本文方案在在线计算量方面有较明显的优越性。

5　结束语

本文针对代理发现中用户对代理的性能、成本和安全性等方面的需求，以及需求匹配过程中的隐私保护问题，基于 Paillier 同态加密算法，提出一种新的综合考虑代理和用户属性及其偏好的私有数据信息匹配算法，该算法引入半可信主代理，从全局层面管理所有子代理的业务类型和连接状况，并承担主要的计算开销，同时将欧氏距离与 Paillier 同态加密算法有机结合，支持具有偏好信息的多元属性数据匹配，能够有效保障用户和子代理的安全性。通过安全性分析，表明了方案的正确性和安全性，并通过性能仿真进一步验证方案的有效性和高效性。

参考文献：

[1]　ZHENG Y, LI M, LOU W, et al. Location based handshake and private proximity test with location tags[J]. IEEE Transactions on Dependable

and Secure Computing, 2015:1.

[2] SUN J, ZHANG R, ZHANG Y. Privacy-preserving spatiotemporal matching[C]//32th IEEE International Conference on Computer Communications. c2013: 800-808.

[3] LU R, LIN X, SHEN X. SPOC: a secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency[J]. IEEE Transactions on Parallel and Distributed Systems, 2013, 24(3): 614-623.

[4] AGRAWAL R, EVFIMIEVSKI A, SRIKANT R. Information sharing across private databases[C]//30th ACM SIGMOD International Conference on Management of Data. c2003: 86-97.

[5] VAIDYA J, CLIFTON C. Secure set intersection cardinality with application to association rule mining[J]. Journal of Computer Security, 2005, 1(13): 593-622.

[6] NIU B, ZHU X, ZHANG T, et al. P-match: priority-aware friend discovery for proximity-based mobile social networks[C]//10th IEEE International Conference on Mobile Ad-Hoc and Sensor Systems. c2013: 351-355.

[7] LI R, WU C. An unconditionally secure protocol for multi-party set intersection[M]. 5th Springer Applied Cryptography and Network Security, c2007: 226-236.

[8] KERSCHBAUM F. Outsourced private set intersection using homomorphicencryption[C]//7th ACM Symposium on Information, Computer and Communications Security. c2012: 85-86.

[9] LI M, YU S, CAO N, et al. Privacy-preserving distributed profile matching in proximity-based mobile social networks[J]. IEEE Transactions on Wireless Communications, 2013, 12(5): 2024-2033.

[10] ZHANG R, ZHANG J, ZHANG Y, et al. Privacy-preserving profile matching for proximity-based mobile social networking[J]. IEEE Journal on Selected Areas in Communications, 2013, 31(9): 656-668.

[11] THAPA A, LI M, SALINAS S, et al. Asymmetric social proximity based private matching protocols for online social networks[J]. IEEE Transactions on Parallel and Distributed Systems, 2015, 26(6): 1547-1559.

[12] CHU W T, CHANG F C. A privacy-preserving bipartite graph matching framework for multimedia analysis and retrieval[C]//5th ACM International Conference on Multimedia Retrieval, c2015: 243-250.

[13] JECKMANS A, TANG Q, HARTEL P. Poster: privacy-preserving profile similarity computation in online social networks[C]//18th ACM Conference on Computer and Communications Security. c2011: 793-796.

[14] WANG B, LI B, LI H. Gmatch: secure and privacy-preserving group matching in social networks[C]//55th IEEE Global Communications Conference. c2012: 726-731.

[15] NIU B, HE Y, LI F, et al. Achieving secure friend discovery in social strength-aware PMSNs[C]//34th IEEE Military Communications Conference, c2015: 947-953.

[16] PAILLIER P. Public-key cryptosystems based on composite degree residuosity classes[J]. Springer Cryptology-EUROCRYPT, 1999, 547(1):223-238.

[17] RANE S, SUN W, VETRO A. Privacy-preserving approximation of L1 distance for multimedia applications[C]//Multimedia and Expo (ICME). c2010: 492-497.

[18] ZHANG R, ZHANG R, SUN J, et al. Fine-grained private matching for proximity-based mobile social networking[C]//INFOCOM. c2012: 1969-1977.

作者简介：

耿魁（1989-），男，湖北红安人，西安电子科技大学博士生，主要研究方向为网络安全。

万盛（1987-），男，江苏南通人，西安电子科技大学博士生，主要研究方向为网络安全与隐私保护。

李凤华（1966-），男，湖北浠水人，博士，中国科学院信息工程研究所副总工、研究员、博士生导师，主要研究方向为网络与系统安全、信息保护、隐私计算。

何媛媛（1985-），女，湖北松滋人，中国科学院信息工程研究所博士生，主要研究方向为信息安全、隐私保护。

王瀚仪（1994-），女，吉林省吉林市人，中国科学院信息工程研究所博士生，主要研究方向为网络安全。