隐私保护的参与式感知数据分享与访问方案

刘树波 王 颖 刘梦君
(武汉大学空天信息安全与可信计算教育部重点实验室 武汉 430072)
(武汉大学计算机学院 武汉 430072)

摘 要 随着移动设备的发展,参与式感知具有广泛的应用前景。由于参与式感知的用户主体是具有社会属性的人,因此,其面临许多传统传感器网络未曾遇到的问题,用户在采集与共享数据过程中的安全与隐私问题便是其中之一,在用户与用户交互过程中怎样使用户通过单次交易就能获得全部必需的数据,以及在获得全部必需数据的同时如何保证用户的身份隐私和偏好隐私是用户十分关心的问题,也是参与式感知应该解决的问题。首先,通过采用双线性映射和盲签名来保护用户的身份隐私;其次,采用布隆过滤器,使用户通过单次数据交易就能获得全部必需的数据,同时保护用户的偏好隐私不被泄露给匹配失败的数据提供者;最后,通过分析表明了该方案的安全性和可行性。

关键词 参与式感知,身份隐私和偏好隐私,双线性映射,盲签名,布隆过滤器
中图法分类号 TP393.08 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.6.031

Privacy-preserving Data Sharing and Access Control in Participatory Sensing

LIU Shu-bo WANG Ying LIU Meng-jun
(Key Laboratory of Aerospace Information Security and Trusted Computing Ministry of Education, Wuhan University, Wuhan 430072, China)
(School of Computer, Wuhan University, Wuhan 430072, China)

Abstract With the development of mobile devices, participatory sensing has a broad application prospect. As the main users of participatory sensing are persons social attributes, participatory sensing is facing many problems which don't encounter in conventional sensor networks. The security and privacy for users to collecte and share data with others are one of the most important issues. One of the most concerned problems is how to get all the necessary data through single transaction and how to keep identity privacy and preference privacy for users to get all the necessary data when users interact with others. Meanwhile, if participatory sensing wants to be developed, the problems should be solved first. The scheme uses bilinear mapping and blind signature to protect the identity privacy of users, and uses bloom filter to make users get all the necessary data through single transaction and protect the preference privacy of users from data providers at the same time if the matching request fails between users and data providers. Finally, performance analysis shows the security and feasibility of proposed scheme.

Keywords Participatory sensing, Identity privacy and preference privacy, Bilinear mapping, Blind signature, Bloom filter

1 引言

随着移动智能终端的迅速普及和移动互联网的快速发展,智能终端用户出于个人或经济兴趣,有意识地将自己或手机内置传感器(GPS、加速度、时间、图像、温度等)采集到的信息通过社交媒体等移动互联网媒介共享,构成了一个感知能力强大的巨型"传感器网络[1]",催生了一类特殊的无线传感器应用——参与式感知[2]。

与传统的传感器网络相比,参与式感知主要有着部署费用低、覆盖范围广、感知数据类型和内容丰富灵活等优点[3]。近几年来,关于参与式感知的应用及研究大量涌现[4-9]。然而,由于参与式感知的用户主体是具有社会属性的人,因此其面临着许多传统传感器网络未曾遇到的问题,用户在数据的采集与共享过程中的安全与隐私问题便是其中之一。目前大部分的应用与研究都集中关注于数据的采集与共享过程[4,5],现有的安全研究也局限于数据的传输安全[6,7]以及用户与服务器交互过程中的隐私保护问题[8,9]。用户之间交互过程中的安全与隐私问题还鲜有关注。

一般而言,单个用户智能终端的功能有限,用户可能无法采集到自己需要的某些数据,而服务器上存储的往往只是用户共享的部分低精度稀疏数据,此时,用户可以向其周围用户请求获取更高精度的感知数据。出于安全与个人隐私考虑,

到稿日期:2014-07-16 返修日期:2014-09-12 本文受国家973计划项目(2011CB302306),中央高校基本科研业务费专项资金(211-274230),国家自然科学基金(41371402),水利部"948"项目(201044),湖北省水利厅农村饮用水水资源远程监控项目资助。
刘树波(1970—),男,教授,博士生导师,主要研究方向为信息安全、物联网、嵌入式系统,E-mail:liu.shubo@whu.edu.cn(通信作者);王 颖(1991—),女,硕士生,主要研究方向为信息安全、嵌入式系统,E-mail:543988022@qq.com;刘梦君(1988—),男,博士生,主要研究方向为移动计算与无线网络、移动社交与分布式系统中的安全及隐私,E-mail:lmj_whu@163.com。

用户在向周围用户提出数据请求过程中，并不希望将自己的个人身份信息泄露给他人，并且用户总是期望通过单次交易就获取全部自己所需的数据。而其它系统用户亦不愿意提供无偿服务。这就需要一个能保护个体隐私，且能够进行身份验证的有偿服务方案来解决上述问题。

传统的隐私保护方案如伪名[10]、K-匿名[11]、群签名[12]等通常采用一个集中的管理者来隔离用户，保障系统安全并消除用户交互过程中的个人隐私信息泄露风险。由于集中的管理者常常不能抵御单点攻击，一旦失效，将严重损害整个系统用户的隐私和安全[13]。而在参与式感知中，用户移动性强，静态的验证方式[11,12]难以适用，节点计算能力有限，过于复杂的方案不切实际，这些限制使得如何分布式地验证用户身份的正确性和有效性面临着巨大的挑战。此外，在验证用户身份合法性之后，如何使得合法用户通过单次交易就能获得全部必需的数据亦是一个需要解决的重要问题。

本文首先将上述用户身份验证问题归结为一个隐私保障的用户访问控制问题，并使用双线性映射和盲签名技术来解决这一问题。其次，在不泄露用户隐私前提下，将合法用户通过单次交易就能获得全部必需的数据问题归结为一个隐私保护的用户感知数据轮廓匹配问题，并使用布隆过滤器技术来解决这一问题。理论分析表明，本方案能够解决上述的参与感知环境下用户数据共享过程中的安全与隐私问题。

2 相关研究

与本文有关的研究主要集中在匿名访问控制以及隐私保护的集合交集两个问题上。

对于匿名访问控制问题，文献[10]采用伪名的方式通过不停更换伪名来防止其他节点通过报文分析追踪或者定位用户，但是伪名也存在一些威胁：恶意用户能通过不断变换伪名来从事各种攻击、欺诈或逃避事故责任。同时由于采用伪名机制，必须有可信第三方对每个用户的伪名进行分配，对第三方依赖性强，另外用户必须对自己拥有的所有伪名进行存储，从而浪费了存储空间，这对于资源有限的移动终端来说是不划算的。

文献[11]采用 k-匿名技术，将用户信息隐藏在至少其他 $k-1$ 个难以区分的用户中，从而使得攻击者不能分辨出敏感信息是出自 k 个用户中的哪一个。k-匿名技术的实现依赖于可信第三方，可信第三方完成 k 个用户的匿名。但在这种方案中，当遇到查询高峰时第三方容易成为系统的瓶颈，并且第三方容易成为攻击者的重点攻击目标，一旦第三方被攻破，系统中的用户将无隐私可言。另外 k 个用户的聚合操作计算量较大，耗费的时间也较长，对于资源有限的移动终端是不可行的。

文献[12]采用群签名，一个群体中的任意一个成员可以以匿名的方式代表整个群体对消息进行签名。但是群签名管理员的权力过大，可以追踪到签名者的身份。同时群签名的效率比较低，同样不适用于资源有限的移动终端。

对于隐私保护的集合交集问题，在文献[13]中，Agrawal 等人提出了一个可交换加密协议来解决 PSI 和 PSCI 问题，实现了两个数据集中交集的运算。可交换加密使用了一对加密函数 f 和 g，且 $f(g(x))=g(f(x))$，该函数的特点是加密结果与顺序无关，例如 $f_e(x)=x^e \bmod p$，其中 p 是一个安全质数。而 Agrawal 等人提出的协议是一个双方单向协议，即参与双方只有一方能知道交集，而另一方得不到任何信息。此外，该协议不能抵御恶意用户攻击。

文献[14]中，Xie 等人在 Agrawal 的可交换加密算法的基础上提出了一个移动社交网络中的匹配协议，协议分成 3 个步骤：1. 初始化，用户进行注册，获取唯一的身份标识；2. 对属性进行认证，用户将自己的属性上传到验证服务器，得到属性的验证信息；3. 执行匹配协议。同时，他们将算法进行扩展，使其足以抵御一定恶意攻击。首先，对属性元素进行签名认证，以避免敌手在交互过程中随意选择属性，从而避免伪造和扫描攻击；其次，交互双方都能得到交集，而不再是单向协议；最后，从交互双方行为出发，分析了一部分恶意攻击，但是计算量大。

文献[15]采用异或操作作为可交换加密函数，极大地减少了计算开支，但安全性有所降低。

文献[16]提出了一个基于双线性映射的匹配算法，该文献将匹配运用到了疾病监控的具体案例中，使具有相同病症的人可以进行通讯，分享信息。该系统设计成 3 个部分：系统参数设置、病人注册和分配密钥以及具有相同病症的病人之间进行匹配。该算法的核心是利用双线性映射函数的性质，通过双方各自计算来进行匹配。但该算法只适用于匹配一个属性的场景，难以扩充到多属性的应用中。

目前针对于匿名访问控制的方案在认证过程中多数需要可信第三方的参与。而在参与式感知中，用户的移动性较强，因此需要一个可隔离可信第三方的匿名认证方式，使用户能够分布式地进行匿名身份认证。而对于隐私保护的集合交集问题，目前的方案都存在一定的不足，如：计算量大[13,14]、安全性低[15]、应用场景受限[16]等。对于资源有限的移动端来说，使用这些方案可行性不大。因此需要一个计算量小、通信代价小，且能优先保证用户隐私的解决方案。

3 模型与假设

3.1 系统模型

如图 1 所示，本文的系统模型主要由一个可信第三方、用户（包括数据提供者和数据使用者）和一个网络拥有者构成，用户既可以是数据提供者也可以是数据使用者。可信第三方主要进行网络中网络拥有者和用户身份的认证。图 1 中的数据提供者利用自己智能终端的传感器类型采集数据，是数据的提供者。数据使用者由于自身智能终端或者网络的限制，只能获得有限的数据，但是又需要一些其他数据信息，因此向周围用户进行数据请求，是数据的使用者。网络拥有者对整个网络进行管理，数据使用者需要在网络拥有者处购买服务令牌才能从其他数据提供者处有偿获得数据。

本文中的服务交易是通过系统的令牌来进行的，此处的令牌相当于货币的作用，一个令牌只能被使用一次，同时令牌跟用户身份有一一对应关系。因此在交易过程中，数据提供者需要对令牌进行匿名验证。

在本文中，因为主要是讨论当用户无法从服务器处获得所需数据时与周围用户之间的匿名交易，所以不考虑用户与服务器之间的数据交互，只考虑用户与用户之间的数据交互过程。用户与服务器之间仍存在注册和身份验证过程。

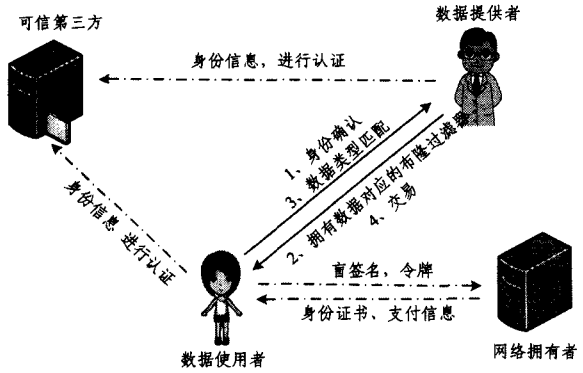

图1 系统模型

3.2 安全模型

假设网络拥有者会根据用户的付款信息，给予相应的令牌，但是他可能会对用户的身份信息和用户需要的数据类型感兴趣。假设用户是理性、诚实而好奇的。理性是指获得满足自己所需数据时才会付出相应代价；诚实而好奇是指每个用户都希望隐藏自己的身份和偏好隐私，但是却希望知道其他用户的身份和偏好隐私。可信第三方是完全可信的。

3.3 问题描述

3.3.1 匿名用户访问控制

在参与式感知交互过程中，数据提供者需要对发送请求的用户的身份进行验证，判断是否为合法用户，合法用户才有资格与之进行后续交互操作，同时数据使用者也不希望将自己的身份隐私暴露给数据提供者，所以数据提供者需要对数据使用者进行匿名的身份验证。另外由于系统主要是通过交易令牌来进行数据交易过程的，而令牌与用户的身份是一一对应关系，同时数据使用者在交互过程中需要保证自己的身份隐私，因此数据提供者需要对令牌正确性进行验证的同时保证数据使用者的身份不会暴露给数据提供者。

整个过程中的安全目标：

(1)数据提供者需要对数据使用者的身份合法性进行匿名验证。

(2)在数据交易过程中，数据提供者需要对数据使用者的令牌进行验证，但是又不能将令牌与数据使用者的身份对应起来，需要保证数据使用者的身份隐私。

3.3.2 隐私保护的感知数据交易

由于一块令牌只能使用一次，用户都希望使用当前令牌能够获得自己所需数据。用户在向数据提供者发送交易请求时不知道数据提供者是否拥有自己需要的全部数据类型，因此需要进行数据类型的匹配度验证。与此同时，用户不希望自己需要的数据类型被不满足条件的数据提供者知晓，而数据提供者(满足条件的和不满足条件的)也不希望已拥有的数据类型被无关人员知晓，因此在匹配过程中，需要保证交易双方的偏好隐私。

定义整个系统中的数据类型对应的属性集合为一个固定长度的向量集合 $A=\{A_1,A_2,\cdots,A_n\}$，n 是系统预先设定的值，由于每个用户既可以是数据使用者也可以是数据提供者，因此每个用户都拥有两个向量集合 NA_i 和 PA_i。NA_i 表示用户需要的数据类型对应的属性集合，PA_i 表示用户拥有的数据类型对应的属性集合。

进一步定义两个用户 U_i(Alice)和 U_j(Bob)，Alice 的智能终端功能有限，但是 Alice 希望从 Bob 处获得一些其智能终端能力之外的数据，假设 Alice 需要的数据类型对应的属性集合为 $NA_i=\{A_{i1},A_{i2},\cdots,A_{in}\}$，其中 Alice 需要的数据类型的相应位为1，其它位为0，向量长度为 n。Bob 拥有的数据类型对应的属性集合为 $PA_i=\{A_{j1},A_{j2},\cdots,A_{jn}\}$，其中 Bob 拥有的数据类型的相应位为1，其它位为0，向量长度为 n。Alice 向 Bob 发送请求，Bob 验证了 Alice 身份的合法性后与之继续交互。由于 Alice 希望能直接从一个数据提供者处获得全部所需的数据，因此 Alice 将需要的数据类型对应的属性集合 NA_i 和 Bob 拥有的数据类型对应的属性集合 PA_i 进行匹配，如果 Bob 拥有的数据类型满足 Alice 需要的数据类型，即 Alice 能够通过一次交易获得所有必需的数据，则 Alice 与 Bob 进行交易，获得想要得到的数据；如果 Bob 没有 Alice 想要的全部的数据，则 Alice 终止与 Bob 的交易，并继续与其它用户进行上述相同过程直到找到一个满足交易条件的用户，即找到一个拥有 Alice 全部所需数据类型的用户，然后 Alice 与满足交易条件的用户进行交易。

整个过程中的安全目标：

(1)整个匹配交易过程中，Alice 的身份是保密的，Bob 无法知道 Alice 的具体身份，即满足 Alice 的身份隐私。

(2)在 Alice 判断 Bob 是否拥有全部需要的数据之前，Bob 不知道 Alice 需要的数据类型，即满足 Alice 的偏好隐私。

(3)Bob 的全部数据类型对 Alice 是保密的，如果匹配成功，Alice 只能得到自己想要的数据类型，即满足 Bob 的偏好隐私。

(4)Alice 不知道 Bob 的具体身份，只需要知道 Bob 是否有自己想要的数据类型，即满足 Bob 的身份隐私。

4 具体方案

4.1 设计思想

根据3.3节问题描述，用户在交互过程中首先需要避免自己的身份信息泄露，随后在交易过程中需要保证在获取自己需要的数据类型时，不泄露自己的偏好隐私。双线性映射技术由于具有安全性高、密钥量小和无需第三方实时参与的特点，因此十分适用于参与式感知下这种资源受限的分布式用户身份的匿名验证。另外，通过采用计算量小的盲签名对令牌进行加密，使用户的身份信息与令牌没有直接明显的对应的关系。而对于用户自身的偏好隐私，3.3节已经将其建模成了向量的集合交集问题，在这种情况下，采用时间和空间效率高的布隆过滤器来进行交易验证，即可解决匹配过程中的用户数据偏好隐私问题。

4.2 相关知识

4.2.1 双线性映射

假设 G_1 和 G_2 是两个阶都为大素数 p 的乘法群，映射 e：$G_1\times G_2\rightarrow G_2$ 称为双线性映射，则映射 e 满足如下性质：

(1) 双线性。对于 $\forall P,Q\in G_1$ 和 $\forall a,b\in Z_p$，都有 $e(P^a,Q^b)=e(P,Q)^{ab}$。

(2) 非退化性。$\exists P,Q\in G_1$，使得 $e(P,Q)\neq 1$。

(3) 可计算性。对于 $\forall P,Q\in G_1$，都会有一个有效的多项式时间算法来计算 $e(P,Q)$。

4.2.2 盲签名

盲签名(Blind Signature)[17]是一种数字签名的方式，在

消息内容被签名之前，对于签名者来说消息是不可见的。盲签名满足两条性质：(1)签名者对其签署的消息是不可见的，即签名者不知道他所签署消息的具体内容；(2)签名消息不可追踪，即当签名消息被公布之后，签名者无法知道这是他哪次签署的。盲签名过程如下：

(1)盲化消息：接收者将待签数据进行盲变换，把变换后的盲数据发送给签名者，$m'=mk^e \bmod n$，k 是随机选择的数，满足条件 $gcd(k,n)=1$。

(2)签名消息：签名者对盲化后的消息进行签名，$s'=(m')^d \bmod n$。

(3)除盲消息：接收者对签名做除盲操作，得到的就是签名者对原数据的盲签名，$s=s' \cdot k^{-1} \bmod n$。

盲签名有效的原因是：

$k^{ed}=k \bmod n$

$s=s' \cdot k^{-1}=(m')^d k^{-1}=m^d k^{ed} k^{-1}=m^d \bmod n$

4.2.3 布隆过滤器

布隆过滤器(bloom filter)[18]是一个二进制向量和一系列随机映射函数，它具有很好的空间和时间效率，被用来检测一个元素是否在一个集合中[19-21]。布隆过滤器原理如下。

假设有一个 ω 位的布隆过滤器，属性集合为 $\{s_i\}_{i=1}^{d}$，初始化时，布隆过滤器的每一位都为 0。k 个不同的 *hash* 函数集合 $\{h_j(\cdot)\}_{j=1}^{k}$，每一个 *hash* 函数的输出结果为 $[1,\omega]$。将所有的 $\{h_j(s_i)\}_{j=1}^{k}$ 对应的位置 1。判断一个元素 e 属不属于集合，就检查 $\{h_j(e)\}_{j=1}^{k}$ 对应的所有位是否为 1，如果所有位都为 1，则元素 e 属于集合；否则不属于集合。

4.3 方案描述

围绕着 3.3 节中的两大问题，本文方案主要包括系统初始化、用户请求、属性匹配和交易 4 个部分。系统初始化主要是用户和网络拥有者进行身份验证和注册，得到相应的安全参数以及用户向网络拥有者购买令牌；用户请求阶段主要是对请求者的身份进行验证；属性匹配阶段主要是请求者将自己需要的数据类型对应的属性和数据提供者拥有的数据类型对应的属性进行匹配，判断是否满足交易条件；交易就是在认证和匹配都成功之后用户之间的数据传输过程。

4.3.1 系统初始化

(1)相关参数生成

网络拥有者生成公钥$\langle n,e\rangle$和私钥 a。n 是由两个随机大质数 p,q 计算得来的：$n=pq$；e 与 ϕ 互质，$\phi=(p-1)(q-1)$ 且 $1<e<\phi$，满足 $1<a<\phi$ 且 $ea=1 \bmod \phi$。同时选取阶都为 p 的乘法群 G_1 和 G_2 以及满足双线性映射性质的双线性映射 $e':G_1 \times G_2 \to G_2$。令 g、h 为 G_1 的两个生成元，并设置 $g_1=g^a$。同时网络拥有者选择一个 *hash* 函数 $H:\{0,1\}^* \to Z^*$。网络拥有者将 $P_{pub}=\{n,e,G_1,G_2,e',g,h,g_1,H\}$ 公开，$p_{pri}=\langle p,q,a\rangle$保密。$\langle n,e\rangle$主要用来对网络拥有者的签名进行认证。

网络拥有者将公钥 $P_{pub}=\{n,e,G_1,G_2,e',g,h,g_1,H\}$ 身份信息(ID)发送给可信第三方，可信第三方对网络拥有者的身份进行验证，生成证书 $C_0=\langle\{n,e,G_1,G_2,e',g,h,g_1,H\}、ID、OI_g\rangle$($OI_e$ 为其他信息包括证书有效时间和可信第三方的相关信息)，同时对证书进行签名。用户将身份信息(ID_i)发送给可信第三方，可信第三方对用户的身份进行验证，然后为每个合法用户 U_i 随机选择一个 λ 位的整数 m_i，$0 \leqslant m_i \leqslant 2^{\lambda}-1 \leqslant n-1$，生成证书 $C_i=\langle m_i^*、OI_e\rangle$($m_i^*$ 是可信第三方用私钥对 m_i 进行加密后的结果，OI_e 为其他信息包括证书有效时间和可信第三方的相关信息)，同时对证书进行签名。可信第三方将用户 U_i 对应的签名后的证书 $C_i=\langle m_i^*、OI_e\rangle$和 m_i 发回给用户。

(2) 用户公私钥生成

任何用户在加入系统时，都需要与网络拥有者进行交互，生成用户的公私钥对($K_{pub}^{U_i}$，$K_{pri}^{U_i}$)。

1. 用户 U_i 随机选择大整数 a_i 并保密当作私钥。同时计算其公开值 $R_i=h^{a_i}$，然后将(R_i，C_i)发送给网络拥有者，其中 C_i 为用户 U_i 的身份证书。

2. 网络拥有者收到(R_i，C_i)之后对 C_i 进行验证，若不合法，则拒绝 U_i 加入系统；反之，计算 $Q_i=H(C_i)$ 以及用户 U_i 的公钥 $H_i=(R_i)^{1/(a-Q_i)}=h^{a_i/(a-Q_i)}$。然后将 H_i 给用户 U_i。

3. 用户 U_i 得到 H_i 后首先对其进行真实性认证，即判断 $e'(H_i^{1/a_i},g_1 g^{-Q_i})=e'(h,g)$是否成立，若等式成立，则 U_i 的公私钥对($K_{pub}^{U_i}$，$K_{pri}^{U_i}$)=(H_i，a_i)；否则，公私钥对生成失败。

(3) 令牌获取

1. 用户 U_i 随机选择整数 k_i，$0 \leqslant k_i \leqslant n-1$ 且 $gcd(n,k_i)=1$。计算 $m_i'=m_i k_i^e \bmod n$。并将 m_i'和支付信息发送给网络拥有者。

2. 网络拥有者对用户的支付信息进行验证，然后计算 $\sigma_{m_i}'=(m_i')^a \bmod n$，并将 σ_{m_i}'发送给用户。

3. 用户 U_i 收到σ_{m_i}'之后，计算 $\sigma_{m_i}=k_i^{-1}\sigma_{m_i}' \bmod n$，作为网络拥有者对 m_i 的签名。用户 U_i 将$\langle m_i,\sigma_{m_i}\rangle$作为交易的令牌。

4.3.2 用户请求

Alice 将自己的身份证书 C_i、公钥 H_i 和私钥公开值 R_i 发送给 Bob，Bob 计算 $Q_i=H(C_i)$，判断 $e'(H_i,g_1 g^{-Q_i})=e'(R_i,g)$ 是否成立，若成立，则认为 Alice 是合法的，接受 Alice 的匹配请求。

4.3.3 属性匹配

为了增强布隆过滤器的隐私保护程度，采用带密钥的布隆过滤器[22]，即采用密钥 K 来扩展映射函数 $\{h_j(\cdot)\}_{j=1}^{k}$，在利用集合 S 生成布隆向量的过程中，利用 $h_j(s \infty K)$代替 $h_j(s)$ 将布隆过滤器的相应位置 1，其中 $s \in S$，∞表示串联。如果不知道 K，任何人都不能从布隆向量中推断出原始数据的信息。

因为用户对应的 m_i 只有交易双方的用户知晓，本文将用户对应的 m_i 当作密钥 K 来使用。另一方面 Bob 不知道自己是否满足 Alice 的匹配请求，所以，Bob 希望自己拥有的数据类型也对 Alice 保密，因此在匹配过程中，Bob 生成布隆过滤器的函数与 Alice 生成布隆过滤器的函数不完全相同。

匹配过程：

假设 $\mathcal{F}$ 为一个很大的公开的 *hash* 函数池。

(1)Bob 用自己的 ID_j 作索引，利用 $H(ID_j)$ 在 *hash* 池 $\mathcal{F}$ 里面选择 k 个 *hash* 函数，从得到的 k 个 *hash* 函数里面选择 l 个($l<k$)，同时选取 $k-l$ 个不在 $\mathcal{F}$ 里面的 *hash* 函数，通过选取的 k 个 *hash* 函数，计算自己属性集合 $PA_i=\{A_{j1},A_{j2},\cdots,A_{jn}\}$ 对应的 ω 位 *bloom filter*(BF_B)。构造 BF_B 时将 n 位属性集合中对应位为 1 的序号进行哈希操作。

如果 $PA_j=\{0,1,1,1,0,1,1,0,0,0\}$，则分别对{2,3,4,6,7}进行 $\{h_j(\cdot)\}_{j=1}^{k}$ 计算。如'2'：分别计算 $\{h_j(2\infty m_i)\}_{j=1}^{k}$，并分别将 BF_B 中对应的位置 1。其余元素采用相同的操作。

Alice 构造布隆过滤器也采用同样的方法。

(2)Bob 将 $H(ID_j)$和 BF_B 发送给 Alice。

(3)Alice 基于 $H(ID_j)$选取的 k 个 *hash* 函数，构造自己的属性集合 $NA_i=\{A_{i1},A_{i2},\cdots,A_{in}\}$ 对应的 *bloom filter* (BF_A)。然后计算在 BF_A 和 BF_B 中同时为 0 的数目(n_0)，计算属性匹配(即 A_i 中元素属于 B_j)的个数。

$$\hat{m}_A=\frac{2kn-\omega(\ln\omega-\ln n_0)}{l} \tag{1}$$

其中，n 为总的属性个数。

等式的正确性将在 5.4 节中进行分析。交互模型如图 2 所示。

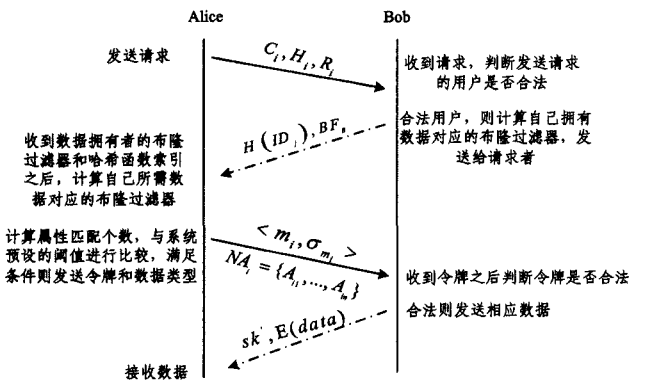

图 2　交互模型

4.3.4　交易

Alice 计算属性匹配个数 $\hat{m}_A$ 之后，如果 $\hat{m}_A\geqslant\tau$，τ 为系统选定的阈值，则 Alice 将自己的令牌$\langle m_i,\sigma_{m_i}\rangle$和需要的数据类型 $NA_i=\{A_{i1},A_{i2},\cdots,A_{in}\}$发送给 Bob。

Bob 在收到令牌之后对令牌的合法性进行验证。计算 $m_i=(\sigma_{m_i})^e \bmod n$ 是否成立，若不成立，则认为令牌不是合法的，拒绝交易。否则，Bob 根据文献[23]的方法检查令牌是否被使用过，如果没有被使用过，则 Bob 将 Alice 需要的数据发送给 Alice。

为了保证数据不被恶意用户截取或修改，采用数字信封的方式对数据进行加密。Bob 随机选择一个随机对称密钥 *sk*，通过对称加密算法利用 *sk* 对数据进行加密，同时，利用 Alice 的公钥 H_i 对 *sk* 进行加密得到 *sk'*，然后 Bob 将 *sk'* 和 Alice 需要的数据一起发送给 Alice。

Alice 收到数据包之后，利用自己的私钥 a_i 解密 *sk'* 得到 *sk*，然后利用 *sk* 解密加密之后的数据得到需要的数据。交易完成。

5　方案分析

5.1　匿名认证

系统参与者都需要把自己的身份信息(*ID*)提交给可信第三方进行认证，用户需要把身份证书 $C_i=\langle m_i^*、OI_e\rangle$、$m_i'=m_ik_i^e \bmod n$ 和 $R_i=h^{a_i}$ 提交给网络拥有者，网络拥有者进行验证之后才对合法用户的 m_i' 进行签名，得到 $\sigma_{m_i}'=(m_i')^a \bmod n$，整个过程中，网络拥有者知道用户是合法的，但是不知道用户的具体身份。

5.2　身份隐私

在用户发出匹配请求的同时将身份证书 C_i、公钥 H_i 和 $R_i=h^{a_i}$ 发送给数据提供者，数据提供者根据双线性映射的性质对用户的身份进行验证：判断 $e'(H_i,g_1g^{-Q_i})=e'(R_i,g)$是否成立。数据提供者能证明用户是否合法，但是不知道合法用户的具体身份信息。

证明用户身份合法之后，用户将令牌$\langle m_i,\sigma_{m_i}\rangle$发送给数据提供者，数据提供者对签名 σ_{m_i}'进行认证，计算 $m_i=(\sigma_{m_i})^e \bmod n$ 判断令牌是否合法，但是数据提供者无法将令牌与具体的用户身份对应起来，得不到任何跟用户身份有关的信息。排除网络拥有者和数据提供者共谋的情况，用户的身份隐私在整个过程中都得到了保护。

由于 *hash* 函数的性质，即使用户得到 $H(ID_j)$也无法计算出 ID_j，保护了数据提供者的身份隐私。

5.3　偏好隐私

在匹配成功之前，数据提供者对用户需要的数据类型是完全无知的。数据提供者选择 $k-l$ 个不在 $\mathcal{F}$ 里面的 *hash* 函数，并将这 $k-l$ 个 *hash* 函数对用户保密，在数据提供者将 $H(ID_j)$和 BF_B 发送给用户之后，用户不能直接计算出数据提供者拥有的数据类型[18]。数据提供者的偏好隐私在整个过程中都是保密的。

由于采用了带密钥的布隆过滤器，即使多个用户多次发送请求得到数据提供者不同的布隆向量，但是因为每个用户对应的 m_i 不同，所以仍无法由布隆向量反推出数据提供者的偏好隐私，因此可以抵抗恶意用户的联合攻击。

5.4　匹配正确性

假设每个元素都是等概率地 *hash* 到 ω 位的布隆过滤器的任何一位，与其他元素被 *hash* 到哪个位无关。m 是实际匹配的属性个数。

则对某一特定位，在一个元素由某个特定 *hash* 函数插入时没有被置位为 1 的概率为：

$$1-\frac{1}{\omega} \tag{2}$$

则 l 个 *hash* 函数全没有将其置位为 1 的概率为：

$$\left(1-\frac{1}{\omega}\right)^l \tag{3}$$

如果插入了 m 个元素，但都未将其置位为 1 的概率为：

$$\left(1-\frac{1}{\omega}\right)^{ml} \tag{4}$$

对于 BF_A 和 BF_B 中的每一位，对于同一个元素，通过 l 个哈希函数，BF_A 或 BF_B 中该位为 1 的概率为：

$$p=1-\left(1-\frac{1}{\omega}\right)^{ml}\approx 1-e^{-\frac{ml}{\omega}} \tag{5}$$

在这 l 个哈希函数之外该位被置为 1 的概率为：

$$q=1-\left(1-\frac{1}{\omega}\right)^{nk-ml}\approx 1-e^{-\frac{nk-ml}{\omega}} \tag{6}$$

因此，在 BF_A 和 BF_B 中，该位同时为 0 的概率为：

$$P_0=(1-p)(1-q)^2=e^{-\frac{ml}{\omega}}e^{-\frac{2(nk-ml)}{\omega}} \tag{7}$$

Alice 计算在 BF_A 和 BF_B 中同时为 0 的数目(n_0)，因此：

$$P_0=e^{-\frac{ml}{\omega}}e^{-\frac{2(nk-ml)}{\omega}}=\frac{n_0}{\omega} \tag{8}$$

通过式(8)得到

$$\hat{m}_A=\frac{2kn-\omega(\ln\omega-\ln n_0)}{l} \tag{9}$$

5.5 匹配精度

采用标准的(ε,δ)来评估匹配精度：

$$P[(1-\varepsilon)m\leqslant\hat{m}\leqslant(1+\varepsilon)m]>1-\delta \tag{10}$$

由文献[24]中的分析知，$\hat{m}$是关于 m 的(ε,δ)估计。其中 δ 满足：

$$\delta\geqslant\frac{\omega(e^{\frac{2nk}{\omega}}-(1+\frac{2nk}{\omega}))}{l^2\varepsilon^2m^2} \tag{11}$$

5.6 匹配开销

整个匹配过程是基于布隆过滤器的，没有采用很多复杂的加密操作，计算量小，交易双方(Alice 和 Bob)只需要进行 nk 个 *hash* 函数操作，复杂度为 $O(kn)$。匹配通信开销也很小，主要是 ω 位的布隆过滤器的传输，复杂度为 $O(\omega)$。与第 2 节中关于隐私保护的多属性集合交集现有的较安全的解决方案相比，本文方案的计算复杂度和通信开销都较小。

由于文献[15]安全性低，文献[16]只适用于特定应用场合，因此此处不进行开销对比。匹配开销对比如表 1 所列。

表 1 匹配过程开销对比

方案	计算复杂度	通信复杂度
文献[13]	O(2C(i+j))	O(i+2j)
文献[14]	O(2C(i+j)+2D)	O(i+2j)
本文	O(kn)	O(ω)

i,j 分别为交互双方拥有的属性数目，C 是利用 $f_e(x)=x^e \bmod p$ 加解密的开销，D 为文献[20]涉及的 Diffie-Hellman 计算开销。

5.7 单次交易获得全部数据

用户在交互过程中，需要根据计算出的匹配值$\hat{m}_A$ 与系统初始化阈值 τ 进行比较，只有$\hat{m}_A\geqslant\tau$ 时，用户才选择与相应的数据提供者交易，通过对阈值 τ 的选择，能够保证用户通过单次交易就能获取全部所需的数据。

由式(4)知，对于插入 n 个元素的集合，布隆过滤器中任意一位为 0 的概率为

$$p=(1-\frac{1}{\omega})^{nk}=e^{-\frac{nk}{\omega}} \tag{12}$$

将不属于集合中的元素误判为属于集合中的元素时，布隆向量所对应的 k 个位置必须全部为 1，则误判率为：

$$FPR=(1-p)^k=(1-e^{-\frac{nk}{\omega}})^k \tag{13}$$

当 m,ω 一定时，对式(13)求导，可知当 $e^{-\frac{nk}{\omega}}=\frac{1}{2}$，即 $k=\ln2\cdot\frac{\omega}{n}\approx0.7\cdot\frac{\omega}{n}$ 时，布隆过滤器的误判率最低，此时误判率的值为：

$$FPR=(1-\frac{1}{2})^k=2^{-k}=2^{-\ln2\cdot\frac{\omega}{n}}=0.6185^{\frac{\omega}{n}}$$

系统匹配过程中，交互双方(Alice 和 Bob)构造布隆过滤器采用的 *hash* 函数不完全相同，会对匹配结果产生一定影响。在此主要讨论在保证误判率最低的情况下 *hash* 函数不同个数 $t=(k-l)$ 对精度的影响。

设置 $\omega=2000,n=80$，则 $\ln2\cdot\frac{\omega}{n}\approx0.7\cdot\frac{\omega}{n}=17.5$，取 $k=17$，此时布隆过滤器的误判率最低。假设此时的误判率为 0。实验结果如图 3 所示。

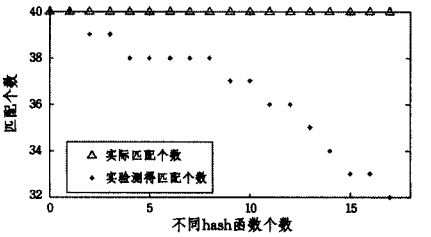

图 3 *hash* 函数不同个数$(k-l)$对精度的影响

由图 3 可知，当 *hash* 函数不同个数 $t=(k-l)$ 增大时，实验的匹配个数减少，但两者不呈现线性关系，即当 $t=k$ 时，匹配个数也不会减小到 0。

按照上述实验数据，$t=k=17$ 时，匹配个数为 32，误差为 $(40-32)/40=0.2$。此时只要设置系统初始的阈值 $\tau=32$ 就能使匹配成功。实际应用中，在可接受的误差范围内，合理地选择 *hash* 函数和阈值 τ，能使匹配精度满足应用要求，使得用户通过单次交易就能获取全部需要的数据。

结束语 在参与式感知的用户之间的交互过程中，为了实现既能保护用户隐私，又能使用户仅通过单次交易便能获得全部所需数据，本文提出了一个基于隐私保护的用户交互机制，采用双线性映射和盲签名来保护用户的身份隐私，采用改进后的布隆过滤器使得用户单次交易就能获得全部必需的数据，同时保护用户的偏好隐私不被泄露给匹配失败的数据提供者。通过分析表明，提出的方案既满足参与式感知用户与用户直接交互的隐私安全需求，同时又使得用户单次交易就能获得全部必需的数据。

参 考 文 献

[1] 任丰原，黄海宁，林闯. 无线传感器网络[J]. 软件学报，2003，(10)：1282-1290
Ren Feng-yuan，Huang Hai-ning，Lin Chuang. Wireless Sensor Networks[J]. Journal of Software，2003，(10)：1282-1290

[2] Burke J，Estin D，Hansen M，et al. Participatory. Sensing[OL]. http://citeseerx. ist. psu. edu/riewdoc/summary?doi=10. 1. 1. 122. 3024

[3] Christin D，Reinhardt A，Kanhere S S，et al. A survey on privacy in mobile participatory sensing applications[J]. Journal of Systems and Software，2011，84(11)：1928-1946

[4] Gaonkar S，Li J，Choudhury R，et al. Micro-Blog：Sharing and Querying Content through Mobile Phones and Social Participation[C]//Proceedings of the 6th ACM International Conference on Mobile Systems，Applications，and Services(MobiSys). 2008：174-186

[5] Kanjo E. NoiseSPY：A Real-Time Mobile Phone Platform for Urban Noise Monitoring and Mapping[J]. Mobile Networks and Applications，2010，15(4)：562-574

[6] Shilton K，Burke J，Estrin D，et al. Participatory Privacy in Urban Sensing[M]. MODUS，2008

[7] Ahmadi H，Abdelzaher T，Han J，et al. The sparse regression cube：A reliable modeling technique for open cyber-physical systems[C]//Proc. 2nd International Conference on Cyber-Physical Systems(ICCPS'11). 2011：87-96

[8] Zhang J，Ma J，Wang W，et al. A novel privacy protection scheme for participatory sensing with incentives[C]//2012 IEEE 2nd International Conference on Cloud Computing and Intelligent Systems(CCIS). IEEE，2012，3：1017-1021

(下转第 157 页)

Based on Intra Prediction Modes for Real-time Applications [OL]. 2012-04http://www.iaeng.org/publication/WCECS2012/WCECS2012_pp655-659.pdf

[3] Mehmood N, Mushtaq M. Blind Watermarking Scheme for H.264/AVC Based on Intra 4x4 Prediction Modes[J]. Lecture Notes in Electrical Engineering, 2012, 179: 1-7

[4] Yang Gao-bo, Li Jun-jie, He Ying-liang, et al. An information hiding algorithm based on intra-prediction modes and matrix coding for H.264/AVC video stream[J]. AEU-International Journal of Electronics and Communications, 2011, 65(4): 331-337

[5] Ko Man-Geun, Hong Jang-Eui, Suh Jae-Won. H.264/AVC error detection scheme using fragile data hiding in motion vector set [C]// 2012 IEEE International Conference on Consumer Electronics (ICCE). 2012: 237-238

[6] Su Yu-ting, Zhang Cheng-qian, Zhang Chun-tian. A video steganalytic algorithm against motion-vector-based steganography [J]. Signal Processing, 2011, 91(8): 1901-1909

[7] Jing Hui-yun, He Xin, Han Qi, et al. Motion Vector Based Information Hiding Algorithm for H.264/AVC against Motion Vector Steganalysis[J]. Intelligent Information and Database Systems Lecture Notes in Computer Science, 2012, 7197: 91-98

[8] 刘争艳. H.264/AVC视频流的信息隐藏技术研究[D]. 长沙: 湖南大学, 2009
Liu Zheng-yan. Researches on H.264/AVC Video Streaming Information Hiding Technology[D]. Changsha: Hunan University. 2009

[9] Kim Sung-Min, Kim Sang-Beom, Hong Youp-yo, et al. Data Hiding on H.264/AVC Compressed Video[J]. Lecture Notes in Computer Science, 2007, 4633: 698-707

[10] L in Xi-jie, Li Qing-bao, Wang Wei, et al. Information Hiding Based on CAVLC in H.264/AVC Standard[C]// 2012 Fourth International Conference on Multimedia Information Networking and Security (MINES). 2012: 900-904

[11] Zhu Hong-liu, Wang Rang-ding, Xu Da-wen. Information hiding algorithm for H.264 based on the motion estimation of quarter-pixel[C]// 2010 2nd International Conference on Future Computer and Communication(ICFCC). 2010: 423-427

[12] 马兰, 沈笑云, 等. 精通 Visual C++视频/音频编码技术[M]. 北京: 人民邮电出版社. 2008: 42-45
Ma Lan, Shen Xiao-yun, et al. Proficient in video/audio coding technology with VisualC++[M]. Beijing: Posts & Telecom Press. 2008: 42-45

[13] Logistic 混沌映射[EB/OL]. 2008-03-10. http://blog.sina.com.cn/s/blog_4511c21f01008p0k.html
Logistic Chaos Map [EB/OL]. 2008-03-10. http://blog.sina.com.cn/s/blog_4511c21f01008p0k.html

[14] jm86之get_block() 1/4亮度像素内插详述[EB/OL]. 2007-02-04. http:// blog.csdn.net/zhangji1983/article/details/1502 035
The detail of the 1/4 pixels interpolation with get_block() in jm 86[EB/OL]. 2007-02-04. http:// blog.csdn.net/zhangji1983/article/details/1502035

[15] JVT Reference Software version JM8.6[EB/OL]. 2012-04. http:// iphome.hhi.de/suehring/tml/download/old_jm/

[16] 唐贵进, 朱秀昌. H.264 JM模型中丢包算法的研究及改进[J]. 数字视频, 2011, 35(13): 6-8, 27
Tang Gui-jin, Zhu Xiu-chang. Research and Improvement of Packet Loss Algorithm in JM Model of H.264[J]. Digital Video, 2011, 35(13): 6-8, 27

(上接第144页)

[9] Ahmadi H, Pham N, Ganti R, et al. Privacy-aware regression modeling of participatory sensing data[C]// Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems. ACM, 2010: 99-112

[10] 朱晓玲. VANET安全和隐私保护机制研究[D]. 合肥: 合肥工业大学, 2013
Zhu Xiao-ling. Security and Privacy Preservation Mechanisms in Vehicular Ad Hoc Network[D]. Hefei: Hefei University of Technology, 2013

[11] Sweeney L. K-anonymity: a model for protecting privacy. International Journal on Uncertainty[J]. Fuzziness and Knowledge-based Syatems, 2002, 10(5): 557-570

[12] Camenisch J, Groth J. Group Signatures: Better Efficiency and New Theoretical Aspects[C]// SCN 2004, LNCS 3352, 2005: 120-133

[13] Agrawal R, Evfimievski A, Srikant R. Information Sharing Across Private Databases[C]// Proceedings of SIGMOD. 2003: 86-97

[14] Xie Q, Hengartner U. Privacy-Preserving Matchmaking for Mobile Social Networking Secure Against Malicious Users[C]// Proceedings 9th Int. Conf. on Privacy, Security. 2011: 252-259

[15] Li S D, Wang D S, Lou P, et al. Symmetric Cryptographic Solution to Yao's Millionaires' Problem and an Evaluation of Secure Multiparty Computations[J]. Information Sciences, 2008, 178(1): 244-255

[16] Liu R X, Lin X D, Liang X H, et al. Secure Handshake with Symptoms-matching[C]// The Essential to the Success of mHealthcare Social Network. BodyNets 2010, Corfu Island, Greece, September 2010: 10-12

[17] Chaum D. Blind signature system[C]// Advances in cryptology. Springer US, 1984: 153-153

[18] Bloom B. Space/time trade-offs in hash coding with allowable errors[J]. Comm. ACM, 1970, 13(7): 422-426

[19] Broder A, Mitzenmacher M. Network applications of bloom filters: A survey[J]. Internet Mathematics, 2005, 1(4): 485-509

[20] Zhao Y, Wu J. B-sub: A practical bloom-filter-based publish-subscribe system for human networks[C]// 2010 IEEE 30th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2010: 634-643

[21] Gremillion L L. Designing a Bloom filter for differential file access[J]. Communications of the ACM, 1982, 25(9): 600-604

[22] Qiu Ling, Li Ying-jiu, Wu Xin-tao. An Approach to Outsourcing Data Mining Tasks While Protecting Business Intelligence and Customer Privacy[C]// Proc. of the 6th IEEE International Conference on Data Mining. Hong Kong, China: [s.n.], 2006: 551-558

[23] Zhang Y, Ren K. DP2AC: Distributed Privacy-Preserving Access Control in Sensor Networks[C]// INFOCOM 2009. IEEE, 2009: 1251-1259

[24] Sun J, Zhang R, Zhang Y. Privacy-preserving spatiotemporal matching[C]// 2013 Proceedings IEEE INFOCOM. IEEE, 2013: 800-808