外包数据库系统中隐私匹配与包含关系的安全计算协议

蒋亚军[1,2] 杨 波[1] 张明武[1] 陈旭日[3]

(华南农业大学信息学院 广州 510642)[1] (湖南科技学院计算机与通信工程系 永州 425100)[2]

(上海大学计算机工程与科学学院 上海 200072)[3]

摘 要 针对外包数据库系统中的隐私匹配问题,提出了基于分布式环境的安全计算协议(协议1):数据所有者采用Mignotte秘密共享方案将数据集外包,用户与第三方服务提供者交互,通过加法同态加密与秘密重构构造判别式,以判别式的值是否为零来判断用户的数据集的元素是否属于数据所有者的数据集,最终实现隐私匹配。此外,在协议1的基础上还提出了一种判断用户数据集是否包含于数据所有者的数据集的协议(协议2)。在半诚实模型下,采用基于模拟器的方法证明了两个协议的安全性。

关键词 外包数据库,秘密共享,隐私匹配,模拟器

中图法分类号 TP309 **文献标识码** A

Secure Computation Protocol for Private Matching and Inclusion Relation against Outsourced Database System

JIANG Ya-jun[1,2] YANG Bo ZHANG Ming-wu[1] CHEN Xu-ri[3]

(College of Informatics, South China Agricultural University, Guangzhou 510642, China)[1]

(Department of Computer and Communication Engineering, Hunan University of Science and Engineering, Yongzhou 425100, China)[2]

(College of Computer Engineering and Science, Shanghai University, Shanghai 200072, China)[3]

Abstract The secure computation protocol based on distributed environment, namely Protocol 1, was proposed for private matching against outsourced database system. The data owner adopted Mignotte's secret sharing scheme to outsource a dataset. The user interacted with some third-party service provider to determine if some elements of the user's dataset belonged to the data owner's dataset by means of additive homomorphic encryption and secret reconstruction to construct discriminant and with the value of discriminant being zero or not, and ultimately realized private matching. In addition, the other protocol was also proposed to determine whether the user's dataset was included in the owner's dataset, namely protocol 2. In the semi-honest model, the security of the two protocols was proved by simulator.

Keywords Outsourced databases, Secret sharing, Private matching, Simulator

1 引言

随着计算机网络技术的迅速发展,数据库外包已成为一个重要的趋势。数据所有者将数据管理业务委托给第三方服务提供者,外包服务提供者提供软硬件和网络资源,为数据所有者及用户提供远程的数据库创建、存储、更新和查询等数据管理服务,从而优化数据所有者的资源配置,降低软硬件投资成本,减轻了管理和维护任务[1,2]。

隐私匹配问题即集合交集的安全计算问题[3,4]。例如Alice有数据集$A=\{a_1,a_2,\cdots,a_n\}$,Bob有数据集$B=\{b_1,b_2,\cdots,b_m\}$,Alice和Bob之间的一个隐私匹配协议使得Bob获取$A\cap B$,但是Bob不能获取除此之外数据集A中的任何其他元素信息。在外包数据库系统的分布式环境中,我们考虑Alice将数据集A外包给第三方服务提供者,Bob通过与服务提供者交互获取$A\cap B$,但是Bob不能获取服务提供者的任何外包信息以及除$A\cap B$之外数据集A中的任何其他元素信息。而安全计算领域中的集合包含关系判断,则只输出包含关系的结果,即B是否包含于A,而不泄露其他任何信息。

1.1 相关工作

通常解决集合运算问题有两种途径:一般多方安全计算协议[5,6]和专门设计的多方安全计算协议。然而,一般多方安全计算协议由于主要考虑一般性,因此效率通常不高。

文献[3]最先提出了一种专门的集合隐私匹配协议,该协议通过构造一个根集包含所求交集的多项式来实现,当多项式的阶数很高时通过哈希组分配降低多项式阶数来提高计算效率。受文献[3]的启发,文献[7,8]通过改进多项式得到了效率更高的隐私匹配协议。

文献[9]利用交换加密提出了一种通过计算两方交集的势来判定集合包含关系的协议,该协议还能判定不相交关系以及相交非包含关系。文献[10]在文献[3]的基础上提出了

到稿日期:2010-04-13 返修日期:2010-07-12 本文受国家自然科学基金(60773175,60973134),现代通信国家重点实验室基金(9140C1108020906),广东省自然科学基金(9151064201000058)资助。

蒋亚军(1975-),男,博士生,副教授,主要研究方向为网络与信息安全、可信计算,E-mail:jyjall@163.com;杨 波(1963-),男,教授,博士生导师,主要研究方向为密码学与信息安全、可信计算;张明武(1970-),男,博士,副教授,主要研究方向为网络与信息安全、可信计算;陈旭日(1971-),男,博士生,副教授,主要研究方向为网络与信息安全、网格计算。

一种判定集合包含关系的协议，该协议利用多项式表示集合，通过计算多项式的值来判断集合是否具有包含关系，该协议可以避免泄漏两方集合交集的势。

随着数据库外包技术的发展，如何将安全多方计算的集合运算应用到外包数据库系统的分布式环境中也开始引起了人们的注意。文献[11]首先提出了一种以 Shamir 秘密共享方案实现数据集外包的分布式环境中的隐私匹配协议，该协议根据 Shamir 秘密共享方案中秘密重构的特点构造判别式来实现。然而由于用户直接与多个服务提供者交互，因此降低了协议的安全性。本文的工作受文献[11]的启发，借签文献[3,10]的思想，在以 Mignotte 秘密共享方案实现数据集外包的分布式环境中提出了隐私匹配与包含关系的安全计算协议。相比文献[11]，它具有更好的安全性。

1.2 本文的结果

针对外包数据库系统中的隐私匹配问题，提出了一个分布式环境下的隐私匹配协议(协议 1)。在该协议中，数据所有者采用 Mignotte 秘密共享方案进行数据集外包，用户与某个第三方服务提供者产生一轮通信，通过加法同态加密与秘密重构构造判别式，以判别式的值是否为零来判断用户的数据集中的元素是否属于数据所有者的数据集，最终实现隐私匹配。同时在协议 1 的基础上提出了一种判断用户数据集是否包含于数据所有者的数据集的协议(协议 2)，服务提供者通过对中间结果的置换，阻止了用户获取两个数据集的交集，通过构造判别式判断用户的数据集是否属于数据所有者的数据集。

2 基础知识

2.1 概念

定义 1(计算不可区分) 假设集合 $X=\{X_n\}$ 和 $Y=\{Y_n\}$ 分别是随机变量 X_n，Y_n 的序列($n\in\{1,2,\cdots,M\}$)，其中 X_n，Y_n 取遍长度为 $poly(n)$ 的字符串。X 和 Y 是计算不可区分的(记作 $X\equiv^c Y$)当且仅当对于每一个多项式时间算法 A，每个 $c>0$，都存在某个整数 K，对于所有的 $n\geqslant K$，有 $|\Pr[A(X_n)=1]-\Pr[A(Y_n)=1]|<1/n^c$。其中，$\Pr[A(x)=1]$ 是 A 在输入 x 时输出 1 的概率。

定义 2(加法同态加密) 加法同态加密方案由密钥产生算法 Gen、加密算法 E、解密算法 D 组成。密钥产生算法 Gen 的输入为一个整数 τ(通常称为安全参数)，输出为一对公私钥(pk,sk)，并满足同态性质：对于两个给定的密文 $c_1=E(m_1)$，$c_2=E(m_2)$，在不需要解密的情况下，可以计算出 m_1+m_2 的密文 $c'=E(m_1+m_2)=E(m_1)\otimes E(m_2)$；对于一个给定的密文 $c=E(m)$ 和一个常数 a，在不需要解密的情况下，可以计算出 am 的密文 $c'=E(am)=E(m)^a$。本文所采用的同态加密方案是语义安全的，即概率多项式时间敌手对任意两个密文是计算不可区分的。

2.2 基于数据库外包的分布式环境

外包数据库系统由三类成员组成：(1)数据所有者 DO；(2)服务提供者 SP；(3)用户 U。数据所有者 DO 是数据库的所有者，并将数据库外包给服务提供者 SP(本文中假设有 w 个服务提供者，表示为 $SP_1,SP_2,\cdots,SP_w$)，服务提供者 SP 负责外包数据库的存储和管理，响应用户 U 提出的计算请求，用户 U 不与数据所有者 DO 直接进行交互，而是直接与某服务提供者进行交互计算。

2.2.1 共享分配阶段

本文中假设数据所有者 DO 有数据集 $A=\{a_1,a_2,\cdots,a_n\}$，通过(k,w)-Mignotte 秘密共享方案将此数据集分配给 w 个服务提供者 $SP_1,SP_2,\cdots,SP_w$。

设 w 为大于 2 的正整数，选择 w 个两两互质的正整数 $d_1<d_2<\cdots<d_w$，令

$$\alpha=\prod_{j=1}^{k}d_j,\beta=\prod_{j=0}^{k-2}d_{w-j}$$

要求 $\beta<a_i<\alpha(1\leqslant i\leqslant n),2\leqslant k\leqslant w$。

对于每个数据项 a_i，SP_j 的共享为 $y_{i,j}$，其中 $y_{i,j}=a_i \bmod d_j$，$1\leqslant i\leqslant n,1\leqslant j\leqslant w$。则 SP_j 关于数据集 A 的共享为：

$$(d_1,d_2,\cdots d_w,y_{1,j},y_{2,j},\cdots,y_{n,j}),1\leqslant j\leqslant w$$

2.2.2 秘密重构阶段

根据(k,w)-Mignotte 秘密共享方案，其中任意 k 个服务提供者(记为 $SP_{t_1},SP_{t_2},\cdots,SP_{t_k}$)都能够恢复数据集 A，但少于 k 个服务提供者将不能恢复数据集 A。

令 $D=\prod_{j=1}^{k}d_{t_j}$，每个服务提供者 S_{t_j} 计算

$$\begin{cases} D_{t_j}=D/d_{t_j} \\ D'_{t_j}\equiv D_{t_j}^{-1}(\bmod d_{t_j}) \\ u_{i,t_j}=y_{i,t_j}\cdot D_{t_j}\cdot D_{t_j}' \end{cases} \quad 1\leqslant i\leqslant n$$

根据中国剩余定理，k 个服务提供者共同计算的(k,w)-Mignotte 秘密重构公式为：

$$a_i=\left(\sum_{j=1}^{k}u_{i,t_j}\right) \bmod D,1\leqslant i\leqslant n$$

性质 1 对于上述 D_{t_j}，D'_{t_j}，有

$$\sum_{j=1}^{k}D_{t_j}\cdot D'_{t_j}=1 \bmod D$$

证明：在共享分配阶段，假设 $a_i=1$，由 $y_{i,j}=a_i \bmod d_j$，则 $y_{i,j}=1,1\leqslant j\leqslant w$，显然

$$u_{i,t_j}=y_{i,t_j}\cdot D_{t_j}\cdot D'_{t_j}=D_{t_j}\cdot D'_{t_j}$$

又 $a_i=\left(\sum_{j=1}^{k}u_{i,t_j}\right) \bmod D$，所以

$$\sum_{j=1}^{k}D_{t_j}\cdot D'_{t_j}=1 \bmod D$$

证毕。

3 基于数据库外包的安全计算协议

协议的成员如果按照协议要求进行运算和发送消息，仅只保留计算的中间结果试图推导出其他成员的输入，则称为半诚实成员；协议成员若不按照协议的明确要求，任意地偏离协议规定的运算，包括拒绝参与协议、替换本地输入或者过早中止协议等，则称为恶意成员。协议的成员若都是半诚实的，则称为半诚实模型；协议的成员若是恶意的，则称为恶意模型。本文协议建立在半诚实模型下。

3.1 协议 1：隐私匹配的协议

在外包数据库系统的分布式环境中，用户 U 通过与某个服务提供者交互求取 $A\cap B$。具体协议描述如下：

输入：用户 U 具有通过密钥产生算法 Gen 产生的同态加密算法 E 的公私钥(pk,sk)，以及数据集 $B=\{b_1,b_2,\cdots,b_m\}(\beta<b_\ell<\alpha)$；每个服务提供者 $SP_j(1\leqslant j\leqslant w)$ 具有数据集 A 的共享 $(d_1,d_2,\cdots d_w,y_{1,j},y_{2,j},\cdots,y_{n,j})$。

输出：用户 U 知道 $A\cap B$。

协议步骤：

1) U 向某一个服务提供者(假设为 SP_{l_1})发出隐私匹配请求，并将公钥 pk 以及同态加密密文$\{E_{pk}(b_1),E_{pk}(b_2),\cdots,E_{pk}(b_m)\}$发送给 SP_{l_1}。

2) SP_{l_1} 向其他任意的 $k-1$ 个服务提供者(设为 SP_{l_2}，SP_{l_3}，$\cdots$，SP_{l_k})发出合作响应隐私匹配的请求，并将 pk 及$\{E_{pk}(b_1),E_{pk}(b_2),\cdots,E_{pk}(b_m)\}$广播给这 $k-1$ 个服务提供者。

3) k 个服务提供者协同计算。

(1) 每个服务提供者 SP_{l_j} 计算

$$f_{\ell,i,j}=E_{pk}(D_{l_j}\cdot D'_{l_j}(b_\ell-y_{i,l_j}))$$

式中，$\ell=1,2,\cdots,m$，$i=1,2,\cdots,n$，$j=1,2,\cdots,k$。然后 SP_{l_j} ($2\leqslant j\leqslant k$)将 $f_{\ell,i,j}$并发送给 SP_{l_1}。

(2) SP_{l_1} 计算 $f_{\ell,i}=(\prod_{j=1}^{k}(f_{\ell,i,j})^{\lambda})$，其中 λ 为随机数，$\ell=1,2,\cdots,m$，$i=1,2,\cdots,n$。然后 SP_{l_1} 将 $f_{\ell,i}$发送给 U。

4) U 计算 $f_\ell=\prod_{i=1}^{n}D_{sk}(f_{\ell,i})\bmod D$，其中 $\ell=1,2,\cdots,m$。如果 $f_\ell=0$，则有 $b_\ell\in A$，否则，$b_\ell\notin A$。

3.2 协议 2：判断包含关系的协议

在外包数据库系统的分布式环境中，用户 U 通过与某个服务提供者交互，判断 B 是否包含于 A。具体协议描述如下：

输入：用户 U 具有通过密钥产生算法 Gen 产生的同态加密算法 E 的公私钥(pk,sk)，以及数据集 $B=\{b_1,b_2,\cdots,b_m\}$($\beta<b_\ell<\alpha$)；每个服务提供者 SP_j($1\leqslant j\leqslant w$)具有数据集 A 的共享$(d_1,d_2,\cdots d_w,y_{1,j},y_{2,j},\cdots,y_{n,j})$。

输出：用户 U 知道 B 是否包含于 A。

协议步骤：

1) U 向某一个服务提供者(假设为 SP_{l_1})发出判断包含关系的请求，并将公钥 pk 以及同态加密密文$\{E_{pk}(b_1),E_{pk}(b_2),\cdots,E_{pk}(b_m)\}$发送给 SP_{l_1}。

2) SP_{l_1} 向其他任意的 $k-1$ 个服务提供者(设为 SP_{l_2}，SP_{l_3}，$\cdots$，SP_{l_k})发出合作响应判断包含关系的请求，并将 pk 及$\{E_{pk}(b_1),E_{pk}(b_2),\cdots,E_{pk}(b_m)\}$广播给这 $k-1$ 个服务提供者。

3) k 个服务提供者协同计算。

(1) 每个服务提供者 SP_{l_j} 计算

$$f_{\ell,i,j}=E_{pk}(D_{l_j}\cdot D'_{l_j}(b_\ell-y_{i,l_j}))$$

式中，$\ell=1,2,\cdots,m$，$i=1,2,\cdots,n$，$j=1,2,\cdots,k$。然后 SP_{l_j} ($2\leqslant j\leqslant k$)将 $f_{\ell,i,j}$并发送给 SP_{l_1}。

(2) SP_{l_1} 计算 $f_{\ell,i}=(\prod_{j=1}^{k}(f_{\ell,i,j})^{\lambda})$，其中 λ 为随机数，$\ell=1,2,\cdots,m$，$i=1,2,\cdots,n$。然后 SP_{l_1} 通过随机置换函数 π 将 mn 个 $f_{\ell,i}$进行置换，将得到的 mn 个 $f_{\pi(\ell,i)}$发送给 U。

4) U 计算 $f_\ell=\prod_{i=1}^{n}D_{sk}(f_{\pi(\ell,i)})\bmod D$，其中 $\ell=1,2,\cdots,m$。然后计算 $f=\sum_{\ell=1}^{m}f_\ell$。如果 $f=0$，则 B 包含于 A，否则 B 不包含于 A。

4 安全和效率分析

4.1 安全性

定理 1 假设同态加密方案具有语义安全性，则协议 1 在半诚实模型中具有安全性。

1) 正确性

根据协议 1 步骤，U 得到密文：

$$\begin{aligned}f_{\ell,i}&=(\prod_{j=1}^{k}(f_{\ell,i,j})^{\lambda})\\&=E_{pk}(\sum_{j=1}^{n}\lambda\cdot D_{l_j}\cdot D'_{l_j}(b_\ell-y_{i,l_j}))\\&=E_{pk}(\sum_{j=1}^{n}\lambda\cdot b_\ell\cdot D_{l_j}\cdot D'_{l_j}-\sum_{j=1}^{n}\lambda\cdot D_{l_j}\cdot D'_{l_j}\cdot y_{i,l_j})\\&=E_{pk}(\lambda\cdot b_\ell-\lambda\cdot\sum_{j=1}^{n}D_{l_j}\cdot D'_{l_j}\cdot y_{i,l_j})\ (\text{根据性质 1})\end{aligned}$$

所以由(k,w)-Mignotte 秘密重构公式，可得：

$$\begin{aligned}f_\ell&=\prod_{i=1}^{n}D_{sk}(f_{\ell,i})\bmod D\\&=\prod_{i=1}^{n}(\lambda\cdot b_\ell-\lambda\cdot\sum_{j=1}^{n}D_{l_j}\cdot D'_{l_j}\cdot y_{i,l_j})\bmod D\\&=\prod_{i=1}^{n}(\lambda\cdot(b_\ell-a_i))\bmod D\end{aligned}$$

式中，$i=1,2,\cdots,n$，$\ell=1,2,\cdots,m$。

显然，如果 $b_\ell\in A$，当且仅当 $f_?=0$，否则 f_ℓ 为一个随机数。

2) 隐私性

如果 U 是敌手，模拟器 Simulator 以 U 的输入输出为输入，Simulator 构造 $A'=\{a_1',a_2',\cdots,a_n'\}$，使得 $A'\cap B=A\cap B$，SP_j 关于数据集 A' 的共享为$(y'_{1,j},y'_{2,j},\cdots,y'_{n,j})$，$1\leqslant j\leqslant w$。模拟步骤如下：

(1) Simulator 把$\{E_{pk}(b_1),E_{pk}(b_2),\cdots,E_{pk}(b_m)\}$发送给服务提供者 SP_{l_1}。

(2) Simulator 计算

$$f'_{\ell,i,j}=E_{pk}(\lambda\cdot D_{l_j}\cdot D'_{l_j}(b_\ell-y'_{i,l_j}))$$

$$\begin{aligned}f'_{\ell,i}&=(\prod_{j=1}^{k}(f'_{\ell,i,j})^{\lambda})\\&=E_{pk}(\lambda\cdot b_\ell-\lambda\cdot\sum_{j=1}^{n}D_{l_j}\cdot D'_{l_j}\cdot y'_{i,l_j})\end{aligned}$$

并将 $f'_{\ell,i}$发送给 U，其中 $\ell=1,2,\cdots,m$，$i=1,2,\cdots,n$。

(3) Simulator 计算

$$f'_\ell=\prod_{i=1}^{n}D_{sk}(f'_{\ell,i})\bmod D=\prod_{i=1}^{n}(\lambda\cdot(b_\ell-a'_i))\bmod D$$

如果 $f_\ell'=0$，则有 $b_\ell\in A$，$\ell=1,2,\cdots,m$。

(4) Simulator 输出 B，$A'\cap B$，$f'_{\ell,i}$ 与 f_ℓ'，其中 $\ell=1,2,\cdots,m$，$i=1,2,\cdots,n$。

显然，模拟结果中 B，$A'\cap B$ 与实际协议中 U 的输入输出相同，根据同态加密的语义安全性，$f'_{\ell,i}$，f_ℓ' 与 U 在实际协议中所看到的消息 $f_{\ell,i}$，f_ℓ 是计算不可区分的。因此模拟的结果与 U 在协议 1 中的结果是计算不可区分的。

如果与 U 交互的 SP_{l_1}，SP_{l_2}，$\cdots$，SP_{l_k} 是敌手，模拟器 Simulator 构造 $B'=\{b_1',b_2',\cdots,b_m'\}$，并计算出相应的密文$\{E_{pk}(b'_1),E_{pk}(b_2'),\cdots,E_{pk}(b_m')\}$。然后计算

$$f'_{\ell,i,j}=E_{pk}(D_{l_j}\cdot D'_{l_j}(b_\ell'-y_{i,l_j}))$$

$$\begin{aligned}f'_{\ell,i}&=(\prod_{j=1}^{k}(f'_{\ell,i,j})^{\lambda})\\&=E_{pk}(\lambda\cdot b_\ell'-\lambda\cdot\sum_{j=1}^{n}D_{l_j}\cdot D'_{l_j}\cdot y_{i,l_j})\end{aligned}$$

式中，$\ell=1,2,\cdots,m$，$i=1,2,\cdots,n$。

则在 SP_{l_1}，SP_{l_2}，$\cdots$，SP_{l_k} 串谋的情况下，可以得到 $f'_{\ell,i,j}$，$f'_{\ell,i}$，及$\{E_{pk}(b_1'),E_{pk}(b_2'),\cdots,E_{pk}(b'_m)\}$。根据同态加密的语义安全性，这与实际协议中的 $f_{\ell,i,j}$，$f_{\ell,i}$及$\{E_{pk}(b_1),E_{pk}(b_2),\cdots,E_{pk}(b_m)\}$是计算不可区分的。证毕。

(下转第 135 页)

[3] Damiani E, Vimercati S D C, et al. Managing and sharing servents' reputation in P2P systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2003, 15(4): 840-854

[4] Damiani E, Vimercati S D C, et al. A reputation-based approach for choosing reliable resources in peer-to-peer networks[C]// Proceedings of the 9th ACNM Conference on Computer and Communications Security(CCS'02). ACM, 2002: 207-216

[5] Jøsang A, Ismail R. The beta reputation system [C] // Proceedings of the 15th Bled Conf. on Electronic Commerce. Bled, Slovenia, 2002: 708-721

[6] Zhou R, Hwang K. PowerTrust: A robust and scalable reputation system for trusted P2P computing[J]. IEEE Transaction on Parallel and Distributed systems, 2007, 18(5)

[7] Wu Xu, He Jing-xia, Xu Fei. An enhanced trust model based on reputation for P2P networks[C]// IEEE International Conf. on Sensor Networks, Ubiquitous, and Trustworthy Computing. Taichung, Taiwan: IEEE Press, 2008: 67-73

[8] Xiong L, Liu L. A reputation-based trust model for P2P ecommerce communities [C] // Proceedings of IEEE International Conf. on Electronic Commerce, 2003

[9] Xiong L, Liu L. PeerTrust: Supporting reputation-based trust for peer-to-peer communities[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16

[10] Srivatsa M, Xiong L, Liu L. Trustguard: countering vulnerabilities in reputation management for decentralized overlay networks[C]// Proceedings of WWW. 2005

[11] Singh A, Liu L. Trust me: Anonymous management of trust relationships in decentralized P2P systems[C]// Proceedings of P2P'03. 2003

[12] Beth T, Borcherding M, Klein B. Valuation of trust in open networks[C]// Proceedings of Computer Security-ESORICS 94. 1994: 3-14

[13] Dellarocas C. Mechanisms for coping with unfair ratings and discriminatory behavior in online reputation reporting systems[C]// Proceedings of the 21st International Conf. on Information Systems(ICIS). Brisbane, Australia, 2000, 12

[14] 毕方明,等. 面向对等网络的主观逻辑信任模型[J]. 计算机工程与应用, 2009, 45(33): 99-102

[15] 田祥宏,等. P2P 环境下的一种混合式信任模型[J]. 计算机工程与应用, 2009, 45(31): 73-76

[16] http://www.guntella.com, 2003

(上接第 122 页)

定理 2 假设同态加密方案具有语义安全性,则协议 2 在半诚实模型中具有安全性。

证明:

1) 正确性

由定理 1 可知,如果 $b_\ell \in A$,当且仅当 $f_\ell=0$,则 B 包含于 A,当且仅当 $f=\sum_{\ell=1}^{m} f_\ell=0$。

2) 隐私性

证明同定理 1 隐私性证明类似,略。证毕。

这里需要指出的是协议 1 和协议 2 的判定泄漏了数据集 A 和 B 的势,并且协议 2 尽管采取了对中间结果的置换,阻止了用户获取两个数据集的交集,但仍然泄漏了交集的势。这些问题是协议 1 和协议 2 美中不足的地方。

4.2 效率

1) 协议 1 的效率

协议 1 需要一轮通信,通信量为 $o(kmn\tau)$ 比特;用户 U 做 $o(m)$ 次加密,$o(mn)$ 次解密;参与的每个服务提供者 SP_{l_j} ($1\leqslant j\leqslant k$) 做 $o(mn)$ 次指数运算和乘法运算。

2) 协议 2 的效率

由于协议 2 中的置换代价可以忽略,因此协议 2 与协议 1 的效率相等。

结束语 本文讨论了基于数据库外包的分布式环境下的隐私匹配问题,给出了一个具体的隐私匹配协议(协议 1),在此基础上提出了一种判断用户数据集是否包含于数据所有者的数据集的安全计算协议(协议 2)。并在半诚实模型下,证明了两个协议的安全性。

下一步的工作将在半诚实模型或者恶意模型下,研究基于数据库外包的分布式环境下的其他运算问题。

参 考 文 献

[1] Hacigümüs H, Mehrotra S, Iyer B. Providing database as a service[C]// Proceedings of the 18th International Conference on Data Engineering. San Jose, USA, IEEE Computer Society Press, 2002: 29-40

[2] Agrawal D, Abbadi A E, Emekci F, et al. Database management as a service: challenges and opportunities[C]// Proceedings of the 2009 IEEE International Conference on Data Engineering. Washington, DC, USA: IEEE Computer Society Press, 2009: 1709-1716

[3] Freedman M J, Nissim K, Pinkas B. Efficient private matching and set intersection[J]. Lecture Notes in Computer Science, 2004, 3027: 1-19

[4] Oleshchuk V A, Zadorozhny V. Secure multi-party computations and privacy preservation: results and open problems[J]. Telektronikk, 2007, 2: 20-26

[5] Yao A C. How to generate and exchange secrets[C]// The 27th IEEE Symposium on Foundations of Computer Science. Toronto, Ontario, Canada, IEEE Computer Society Press, 1986: 162-167

[6] Yao A C. Protocols for secure computations[C]// The 23th IEEE Symposium on Foundations of Computer Science. Chicago, USA, IEEE Computer Society Press, 1982: 80-91

[7] Kissner L, Song D. Privacy-preserving set operations[J]. Lecture Notes in Computer Science, 2005, 3621: 241-257

[8] Sang Y, Shen H, Tan Y, et al. Efficient protocols for privacy preserving matching against distributed datasets[J]. Lecture Notes in Computer Science, 2006, 4307: 210-227

[9] 李顺东,司天歌,戴一奇. 集合包含与几何包含的多方保密计算[J]. 计算机研究与发展, 2005, 42(10): 1647-1653

[10] 李荣花,武传坤,张玉清. 判断集合包含关系的安全计算协议[J]. 计算机学报, 2009, 32(7): 1337-1345

[11] Ye Q, Wang H, Tartary C. Privacy-preserving distributed set intersection[C]// Proceedings of the 3th International Conference on Availability, Security and Reliability. Barcelona, Spain: IEEE Computer Society Press, 2008: 1332-1339