计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2017.20160034
54(7):1549-1556,2017

非加密方法安全计算集合包含关系

陈振华[1,2] 李顺东[3] 王道顺[4] 黄 琼[5] 董立红[1]
[1](西安科技大学计算机科学与技术学院 西安 710054)
[2](信息安全国家重点实验室(中国科学院信息工程研究所) 北京 100093)
[3](陕西师范大学计算机科学学院 西安 710062)
[4](清华大学计算机科学与技术系 北京 100084)
[5](华南农业大学数学与信息学院 广州 510642)
(chenzhenhua@snnu.edu.cn)

Protocols for Secure Computation of Set-Inclusion with the Unencrypted Method

Chen Zhenhua[1,2], Li Shundong[3], Wang Daoshun[4], Huang Qiong[5], and Dong Lihong[1]
[1](*School of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an* 710054)
[2](*State Key Laboratory of Information Security (Institute of Information Engineering, Chinese Academy of Sciences), Beijing* 100093)
[3](*School of Computer Science, Shaanxi Normal University, Xi'an* 710062)
[4](*Department of Computer Science and Technology, Tsinghua University, Beijing* 100084)
[5](*College of Mathematics and Informatics, South China Agricultural University, Guangzhou* 510642)

Abstract The most existing protocols for secure computation of set-inclusion are based on multiple public-key encryption algorithms and cannot either be computed publicly, which makes the computation complexity high and the application range limited as well. To cope with these problems, we design two protocols for secure computation of set-inclusion with the unencrypted method. Protocol 1 first transforms the original problem into the scalar-product problem, and then solves it with the hard problem in cryptography. Next, we present the practical protocol 2 under a certain scenario with a third untrusted party, in which we solve the set-inclusion problem combing the bilinear pairing with the hard problem in cryptography. Lastly, we give the security proof of two protocols, the performance analysis and comparison with others. Both of our protocols employ neither any public-key encryption algorithm nor multiple matching searches so as to avoid the intricate procedures of key-generation, encryption and decryption. This makes our protocols more efficient and concise than the previously known. In addition, we extend the secure computation of set-inclusion to a new application scenario, in which we can determine the set-inclusion relation publicly without revealing the privacy of both sets even if the untrusted third party exists.

Key words set-inclusion; secure computation; bilinear pairing; hard mathematics problem; public determination

收稿日期:2016-01-19;**修回日期**:2016-10-10

基金项目:国家自然科学基金项目(61272435);西安科技大学博士启动金项目(2015QDJ008);信息安全国家重点实验室开放课题基金项目(2016-MS-19)

This work was supported by the National Natural Science Foundation of China (61272435), the Research Fund for the Doctoral Program of Xi'an University of Science and Technology (2015QDJ008), and the Open Fund for State Key Laboratory of Information Security (2016-MS-19).

摘 要 针对已存在的安全计算集合包含关系的协议大多基于多次公钥加密算法,计算复杂性较高,并且不能公开计算,应用受限的问题.提出了2种非加密安全计算集合包含关系的协议.协议1首先将集合包含问题转化为向量内积问题;然后利用数学难解问题解决了此问题;最后针对不可信第三方存在的应用场景,利用双线性对和数学难解问题给出了可公开判断集合包含关系的实用性协议2.协议1和协议2都没有使用任何公钥加密方法,避免了前人方案中繁琐的公私钥产生和加解密过程以及多次匹配查找,因此更加高效而简洁.此外,协议2开拓了保密判断集合关系的新应用场景.

关键词 集合包含;安全计算;双线性对;数学难题;公开判断

中图法分类号 TP309.7

安全多方计算最早由Yao[1]提出,是指在不泄漏各方的输入数据(隐私性)的条件下,能正确完成输入数据的函数计算(正确性).现实问题中涉及到保护隐私的合作计算都可以归结到安全多方计算的范围.因此它在保护隐私的质量评估[2]、定位查找[3]、数据挖掘[4]、数据查询[5]、外包计算[6]等方面有着广泛的应用.

安全计算集合关系属于安全多方计算的一个分支,要求保密地判断一方的集合和另一个集合的关系,却不泄露2个集合数据的隐私,它是保护隐私的数据查询和匹配中的常见问题.针对于此,Freedman等人[7]利用多项式的值研究了集合相交、交集的势等问题;Kissner等人[8]利用多项式的性质;Clifton等人[9]利用随机数分别研究了集合交集、并集、交集的势等问题;夏峰等人[10]基于LWE(learning with errors)困难问题解决了集合相等、相交等问题.这些文献大多集中在集合交集、并集、交集的势等问题,对于集合的包含问题研究并不多.

安全计算集合包含关系要求在不泄露2个集合数据隐私的情况下,保密地判断一方的集合是否包含在另一个集合中.这个问题在现实中有着非常广泛的应用.例如2014年Guo等人[11]提出了基于集合包含关系的加密,并用于不经意传输.在这个加密算法中,只有满足集合包含关系的成员才能恢复消息.但存在的问题是:一方的集合未加保护是在公开信道上传递的,这就导致了集合中部分信息泄露,存在安全漏洞.若能进行全隐私的集合包含关系的判定,就能解决此方案中出现的问题,很好地改进方案.因此,本文对安全计算集合包含关系进行的研究,有着重要的现实意义.若能有效解决安全计算2个集合的包含问题,并以此为基础模块,就能建立很多其他密码学上的方案和应用.

针对安全计算集合包含关系的问题,以往的学者们提出了一些解决方案.李顺东等人[12]利用可交换的加密方案通过求交集的势,从而判断了集合是否包含,该方案实质上是将集合包含问题转化成百万富翁问题,利用可交换加密进行多次逐一匹配查找,若集合很大的话,导致匹配查找的次数也很大,效率较低;Kissner等人[8]将集合表示成多项式,利用门限同态加密算法解决了集合包含问题,但此方案需要门限解密,复杂性很高;李荣花等人[13]也将集合表示成多项式,利用同态加密的方法判断集合的包含关系.虽然比Kissner等人[8]的效率有所提高,但也利用了多次加解密算法,方案还是不够简洁.

由于这些方案都使用了多次公钥加解密算法,而公钥加解密算法,一般效率较低.此外,如果集合的规模很大,多次逐一匹配查找会使得方案的效率进一步降低.

针对以上方案中存在的问题,我们充分利用多项式性质,并结合密码学中的困难问题,重新设计了2个保密判断集合包含关系的协议,贡献有3点:

1) 提出了新的转化方法.本文将集合包含问题转化成向量内积问题,并利用数学难解问题解决了原问题.

2) 拓展了新的应用场景.本文的协议2将传统模式下只能两方保密计算集合关系的场景拓展到任何不可信第三方可参与计算的场景,为可公开保密判断的应用环境提供了新的技术.

3) 提高了效率.本文方案没有使用任何公钥加密算法,避免了前人方案中烦琐的公私钥产生和加解密过程以及多次匹配比较,提高了效率.

1 预备知识

1.1 数学难解问题

1) n-Diffie-Hellman inversion(n-DHI)问题[14]:给定 $g^{a}, g^{a^2}, \cdots, g^{a^n} \in G^{n+1}$,计算 $g^{1/a} \in G$.

2) n-Diffie-Hellman inversion(n-DHI)假设[14]:不存在多项式时间算法可以解决 n-DHI 问题.

1.2 双线性对

G,G_1 为 2 个同为素数阶 q 的乘法群,e 为一个线性映射,$e:G\times G\rightarrow G_1$,$g$ 为 G 的一个生成元,若 e 满足 3 种性质:

1) 双线性性. 对于任意的 $a,b\in q$,$e(g^a,g^b)=e(g,g)^{ab}$;

2) 非退化性. $e(g,g)\neq 1$;

3) 可计算性. 对于任意的 $P,Q\in G$,$e(P,Q)$可有效计算.

则 e 为群 G,G_1 上的双线性对.

1.3 内积协议

定义 1. 内积问题. Alice 拥有向量 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)$,Bob 拥有向量 $\boldsymbol{Y}=(y_1,y_2,\cdots,y_n)$. 在不揭示双方向量隐私的条件下,两者合作计算出内积 $\langle\boldsymbol{X},\boldsymbol{Y}\rangle=x_1y_1+x_2y_2+\cdots+x_ny_n$ 的值.

内积问题是安全多方计算中的基本问题,针对该问题构造的内积协议是构建安全多方计算协议的基本模块. 内积协议最早由 Atallah 等人[15]提出,后来有很多文献[16-22]分别根据不同程度的复杂性和安全性提出了不同的方法.

1.4 安全多方计算的安全性

1) 半诚实参与者

安全多方计算的协议运行环境分为半诚实参与者模型和恶意攻击者模型[23],半诚实参与者指协议方将诚实地执行协议,不会篡改输入和输出信息,但可能会保留计算的中间结果,试图推导出协议之外的信息或者他人的信息.

2) 半诚实模型下的安全性定义

Goldreich[23]利用比特承诺和零知识证明理论设计了一个编译器,这个编译器可以将在半诚实参与者条件下保密计算函数 f 的协议 π 自动生成在恶意参与者条件下也能保密计算 f 的协议 π'. 新的协议 π'可以迫使恶意参与者以半诚实方式参与协议的执行,否则就会被发现. 因此大多数情况下,我们只设计半诚实模型下的协议. 当我们设计出所需要的半诚实模型下的安全多方协议时,只要按照 Goldreich[23]的通用转化方法就可以将原协议转化为恶意模型下的新协议. 基于这一结论,本文也只给出半诚实模型下的协议和相应的安全性模拟范例.

设 $f(x,y)$为概率多项式函数,π 是计算 f 的协议,设 Alice 拥有 x,Bob 拥有 y,他们要在不暴露 x,y 的前提下,合作计算函数 $f(x,y)=(f_1(x,y),f_2(x,y))$. 计算的目的是为了让 Alice 和 Bob 分别得到函数 f 的 2 个分量 $f_1(x,y),f_2(x,y)$. Alice 在执行协议 π 的过程中所得到的视图记为 $view_1(x,y)$,输出记作 $output_1(x,y)$;同理,Bob 的视图记为 $view_2(x,y)$,输出记作 $output_2(x,y)$. Goldreich 在文献[23]中给出计算不可区分性的半诚实参与者的安全两方计算的定义,表述如下:

定义 2. 我们说协议 π 保密地计算了 $f(x,y)$,如果存在概率多项式时间模拟器 S_1 与 S_2,使得式(1)、式(2)同时成立:

$$\{(S_1(x,f_1(x,y)),f_2(x,y))\}\stackrel{c}{=}\{(view_1(x,y),output_2(x,y))\},\quad(1)$$

$$\{(f_1(x,y),S_2(y,f_2(x,y)))\}\stackrel{c}{=}\{(output_1(x,y),view_2(x,y))\},\quad(2)$$

其中,$\stackrel{c}{=}$表示计算不可区分.

此定义说明了任何一方参与者视图中的信息只能从自己输入和所获得的输出中得到,即说明任何一方参与者视图中不包含额外的信息,这样就保证了在协议执行过程中,任何一方得不到其他方的私有信息. 因此要证明一个两方计算协议是保密的,就必须构造使得式(1)和式(2)成立的模拟器 S_1 与 S_2.

2 问题的描述和转化

2.1 问题的描述

安全计算集合包含关系:已知 Alice 拥有集合 $X=\{x_1,x_2,\cdots,x_n\}$,Bob 拥有集合 $Y=\{y_1,y_2,\cdots,y_m\}$,其中 $m\leqslant n$. 除了 2 个集合的势以外,在不泄露 X 和 Y 任何信息的情况下,Alice 和 Bob 都想知道集合 X 和集合 Y 是否存在包含关系,即是否 $Y\subseteq X$.

2.2 问题的转化

Alice 拥有集合 $X=\{x_1,x_2,\cdots,x_n\}$,则此集合可表示为一个 n 次的多项式:

$$f(x)=(x-x_1)(x-x_2)\cdots(x-x_n)=a_0+a_1x+\cdots+a_nx^n=\sum_{i=0}^{n}a_ix^i.$$

定理 1. 记以上多项式 $f(x)$系数组成的向量 $\boldsymbol{a}'=(a_0,a_1,\cdots,a_n)$,某一个元素 y 对应的向量记为 $\boldsymbol{b}'=(1,y,y^2,\cdots,y^n)$,若 $y\in X\Leftrightarrow f(y)=0\Leftrightarrow\langle\boldsymbol{a}',\boldsymbol{b}'\rangle=0$.

推论 1. 记以上多项式 $f(x)$系数组成的向量 $\boldsymbol{a}'=(a_0,a_1,\cdots,a_n)$,另一个集合 $Y=\{y_1,y_2,\cdots,y_m\}$ 对应的向量记为

$$\boldsymbol{y}'=(m,(y_1+y_2+\cdots+y_m),(y_1^2+y_2^2+\cdots+y_m^2),\cdots,(y_1^n+y_2^n+\cdots+y_m^n)).$$

则有：

$$Y\subseteq X\Leftrightarrow(y_1\wedge y_2\wedge\cdots\wedge y_m)\in X\Leftrightarrow$$
$$f(y_1)+f(y_2)+\cdots+f(y_m)=0\Leftrightarrow\langle\boldsymbol{a}',\boldsymbol{y}'\rangle=0.$$

证明. 1) 充分性

$$Y\subseteq X\Rightarrow f(y_1)+f(y_2)+\cdots+$$
$$f(y_m)=0\Rightarrow\langle\boldsymbol{a}',\boldsymbol{y}'\rangle=0.$$

若 $Y\subseteq X$，则对于 Y 中每一个点 $y_i\in X$，根据定理 1，有 $f(y_i)=0$，则有：

$(f(y_1)=0)\wedge(f(y_2)=0)\wedge\cdots\wedge(f(y_n)=0)$，

即有 $f(y_1)+f(y_2)+\cdots+f(y_m)=0$，得到 $\langle\boldsymbol{a}',\boldsymbol{y}'\rangle=0$.

2) 必要性

$$\langle\boldsymbol{a}',\boldsymbol{y}'\rangle=0\Rightarrow f(y_1)+f(y_2)+\cdots+$$
$$f(y_m)=0\Rightarrow Y\subseteq X.$$

对集合 Y 中每一个点 y_i，得到向量 $\boldsymbol{b}_i'=(1,y_i,\cdots,y_i^n)$，$i=1,2,\cdots,m$. 根据定理 1，若 $\langle\boldsymbol{a}',\boldsymbol{b}'\rangle=0\Rightarrow y\in X$，即若 $\langle\boldsymbol{a}',\boldsymbol{b}_i'\rangle=0$，则 $y_i\in X$. 于是得到：

$$\langle\boldsymbol{a}',\boldsymbol{b}_1'\rangle=0,\text{则 } y_1\in X;$$
$$\langle\boldsymbol{a}',\boldsymbol{b}_2'\rangle=0,\text{则 } y_2\in X;$$
$$\vdots$$
$$\langle\boldsymbol{a}',\boldsymbol{b}_m'\rangle=0,\text{则 } y_m\in X.$$

根据内积的性质有：

$$\langle\boldsymbol{a}',\boldsymbol{b}_1'\rangle+\langle\boldsymbol{a}',\boldsymbol{b}_2'\rangle+\cdots+\langle\boldsymbol{a}',\boldsymbol{b}_m'\rangle=$$
$$\langle\boldsymbol{a}',\boldsymbol{b}_1'+\boldsymbol{b}_2'+\cdots+\boldsymbol{b}_m'\rangle=\langle\boldsymbol{a}',\boldsymbol{y}'\rangle=0.$$

于是得到 $\langle\boldsymbol{a}',\boldsymbol{y}'\rangle=0$ 时，有 $(y_1\wedge y_2\wedge\cdots\wedge y_m)\in X$. 即若 $\langle\boldsymbol{a}',\boldsymbol{y}'\rangle=0$，有 $f(y_1)+f(y_2)+\cdots+f(y_m)=0$，得到 $Y\subseteq X$.

由分析可以看出，要判断 2 个集合是否包含（即是否 $Y\subseteq X$），只要计算 2 个集合转化的向量内积是否为 0（即是否 $\langle\boldsymbol{a}',\boldsymbol{y}'\rangle=0$）. 证毕.

3 基本协议

协议假设所有的参与者都在半诚实模型下，网络之间传输都是公开信道. 基于 2.2 节的转化方法，我们给出基本协议 1

协议 1. 利用内积安全计算集合包含关系.

输入：Alice 保密输入集合 $X=\{x_1,x_2,\cdots,x_n\}$、Bob 保密输入集合 $Y=\{y_1,y_2,\cdots,y_m\}(m\leqslant n)$；

输出：Alice 和 Bob 都知道 $Y\subseteq X$ 或者 $Y\not\subset X$.

1) 利用 2.2 节的转化方法，Alice 将集合 X 表示成 n 次多项式 $f(x)$，$f(x)$ 的系数对应的向量记为 $\boldsymbol{a}'=(a_0,a_1,\cdots,a_n)$. Bob 将集合 Y 表示为向量：

$$\boldsymbol{y}'=(m,(y_1+y_2+\cdots+y_m),(y_1^2+y_2^2+\cdots+y_m^2),\cdots,$$
$$(y_1^n+y_2^n+\cdots+y_m^n)).$$

2) Alice 选取一个随机数 $r(r\neq0,1)$，并计算 $f(x)$ 系数的承诺 $g^{ra_0},g^{ra_1},\cdots,g^{ra_n}$ 发送给 Bob.

3) Bob 收到承诺后，计算 $(g^{ra_0})^m,(g^{ra_1})^{y_1+y_2+\cdots+y_n},(g^{ra_2})^{y_1^2+y_2^2+\cdots+y_n^2},\cdots,(g^{ra_n})^{y_1^n+y_2^n+\cdots+y_n^n}$，并计算乘积：

$$(g^{ra_0})^m(g^{ra_1})^{y_1+y_2+\cdots+y_m}(g^{ra_2})^{y_1^2+y_2^2+\cdots+y_m^2}\cdots$$
$$(g^{ra_n})^{y_1^n+y_2^n+\cdots+y_m^n}=$$
$$g^{r[ma_0+a_1(y_1+y_2+\cdots+y_m)+a_2(y_1^2+y_2^2+\cdots+y_m^2)+\cdots+a_n(y_1^n+y_2^n+\cdots+y_m^n)]}=$$
$$g^{r[f(y_1)+f(y_2)+\cdots+f(y_m)]}.$$

若 $g^{r[f(y_1)+f(y_2)+\cdots+f(y_m)]}=1$，则说明：

$$r[f(y_1)+f(y_2)+\cdots+f(y_m)]=0,$$

即：

$$f(y_1)+f(y_2)+\cdots+f(y_m)=0\Leftrightarrow\langle\boldsymbol{a}',\boldsymbol{y}'\rangle=0,$$

因此，$Y\subseteq X$；否则 $Y\not\subset X$.

4) Bob 把结果告诉 Alice.

分析：在协议 1 中，类似于 Feldman[24] 的 VSS (verifiable secret sharing) 方案中的承诺方法，利用离散对数保护 Bob 所持有的多项式 $f(x)$ 的系数 $a_0,a_1,\cdots,a_n$ 的隐私性，但是两者并不相同. Feldman 的原 VSS 方案中的多项式系数是随机选取，而且破解离散对数困难. 而我们这里的多项式系数是定值，因此为了保护 Alice 多项式系数的隐私性，加入随机数 r. 并且为了无法破解离散对数，随机数 r 要在一个较大数域中选择. 这样选择的 r 保护了 Alice 的多项式系数 $a_0,a_1,\cdots,a_n$ 的隐私性，但并不影响最后的计算结果 $f(y_1)+f(y_2)+\cdots+f(y_m)$.

4 应用实例：公开保密判断集合包含关系

协议 1 虽然保护了双方的隐私，但仍然只适用于传统模式：即只能由参与双方交互完成，第三方并不能参与，否则有一方隐私就会泄露. 但在一些场景中，比如云计算下的数据访问、加密数据搜索、可公开仲裁的抗抵赖等环境，第三方必须参与运算. 但由于第三方（服务器）是不可信的，用户并不想暴露自己的隐私，那么如何在不可信第三方存在的情况下设计安全计算集合关系的协议呢？针对此问题，我们将协议 1 拓展，并结合双线性对重新设计了适合此场景的协议 2.

协议假设两方和任何第三方验证者都是半诚实模型，网络之间都是公开信道，允许第三方和任何一方合谋.

协议 2. 不可信第三方存在场景下公开判断集合包含关系

输入:Alice 保密输入集合 $X=\{x_1,x_2,\cdots,x_n\}$、Bob 保密输入集合 $Y=\{y_1,y_2,\cdots,y_m\}(m\leqslant n)$;

输出:任何第三方(包括 Alice 和 Bob)都知道 $Y\subseteq X$ 或者 $Y\not\subset X$.

1) 利用 2.2 节的转化方法,Alice 将集合 X 表示为多项式 $f(x)$,$f(x)$的系数组成的向量记为 $\boldsymbol{a}'=(a_0,a_1,\cdots,a_n)$. Bob 将集合 Y 中每一个元素 y_i $(i=1,2,\cdots,m)$表示为向量 $\boldsymbol{y}_i'=(1,y_i,y_i^2\cdots,y_i^n)$,并得到向量:

$$\boldsymbol{y}'=(m,(y_1+y_2+\cdots+y_m),\cdots,(y_1^n+y_2^n+\cdots+y_m^n)).$$

2) Alice 任意选取一个随机数 $r(r\neq 0,1)$,公布 $f(x)$系数的承诺 $g^{ra_0},g^{ra_1},\cdots,g^{ra_n}$;Bob 对于集合 Y 中每个元素 $y_i(i=1,2,\cdots,m)$,公布向量 $\boldsymbol{y}_i'$ 的承诺:$g^{y_i},g^{y_i^2},\cdots,g^{y_i^n}$.

3) 任何第三方(包括 Alice 和 Bob)首先计算:$g^m,\prod_{i=1}^{m}g^{y_i}=g^{(y_1+y_2+\cdots+y_m)},\cdots,\prod_{i=1}^{m}g^{y_i^n}=g^{(y_1^n+y_2^n+\cdots+y_m^n)}$,计算:

$$e(g^{ra_0},g^m)=e(g,g)^{rma_0},$$

$$e(g^{ra_1},g^{(y_1+y_2+\cdots+y_m)})=e(g,g)^{ra_1(y_1+y_2+\cdots+y_m)},$$

$$\vdots$$

$$e(g^{ra_n},g^{(y_1^n+y_2^n+\cdots+y_m^n)})=e(g,g)^{ra_n(y_1^n+y_2^n+\cdots+y_m^n)},$$

再计算乘积:

$$e(g,g)^{rma_0}e(g,g)^{ra_1(y_1+y_2+\cdots+y_m)}e(g,g)^{ra_2(y_1^2+y_2^2+\cdots+y_m^2)}\cdots e(g,g)^{ra_n(y_1^n+y_2^n+\cdots+y_m^n)}=$$

$$e(g,g)^{r[(a_0+a_1y_1+\cdots+a_ny_1^n)+(a_0+a_1y_2+\cdots+a_ny_2^n)+\cdots+(a_0+a_1y_m+\cdots+a_ny_m^n)]}=$$

$$e(g,g)^{r[f(y_1)+f(y_2)+\cdots+f(y_m)]}.$$

若 $g^{r[f(y_1)+f(y_2)+\cdots+f(y_m)]}=1$,则说明:

$$r[f(y_1)+f(y_2)+\cdots+f(y_m)]=0,$$

即

$$f(y_1)+f(y_2)+\cdots+f(y_m)=0\Leftrightarrow\langle\boldsymbol{a}',\boldsymbol{y}'\rangle=0,$$

因此 $Y\subseteq X$;否则 $Y\not\subset X$.

4) 公布结果.

分析:和协议 1 同理,r 很好地保护了 Alice 的多项式系数向量 $\boldsymbol{a}'=(a_0,a_1,\cdots,a_n)$的隐私性.并不影响最后的计算结果 $f(y_1)+f(y_2)+\cdots+f(y_m)$. 但不同于协议 1 的是 Y 中每一个元素 $y_i(i=1,2,\cdots,m)$的承诺也需要公开.基于 n-DHI 难解问题,任何人无法从公开的承诺 $g,g^{y_i},g^{y_i^2},\cdots,g^{y_i^n}$ 中得到元素 y_i 的信息,否则 n-DHI 问题将被攻破,因此集合 Y 的隐私性也得到了保护.

备注:以上的协议 1、协议 2,唯一泄露的就是 2 个集合的势,但这并不是协议本身的缺陷.即使泄露了集合的势,但这并不影响集合包含的保密判定,只要 2 个集合不是特殊的空集,这唯一的信息对于判定集合是否包含关系,并不起任何作用.关于这一点,Du 等人在文献[25]专门有论证:在设计协议时,如果在降低完美安全性的程度上,允许泄露的信息并不影响方案的有效性,那么这就是可接受性安全.而可接受性安全要根据具体问题,具体设定.

5 安全性分析

本节我们应用 1.4 节的安全性模拟范例给出本文 2 个协议的安全性证明.由于协议 2 和协议 1 证明过程相似,为了节省篇幅,我们证明了协议 1、协议 2 只给出结论即可.

定理 2. 协议 1 保密地判定了集合成员关系.

证明:通过构造满足式(1)和式(2)的模拟器 S_1,S_2 来证明本定理.在本协议中

$$f_1(X,Y)=f_2(X,Y)=(Y\subseteq X)$$

或者

$$f_1(X,Y)=f_2(X,Y)=(Y\not\subset X).$$

假设 $f_1(X,Y)=f_2(X,Y)=(Y\subseteq X)$,构造模拟器 S_1. S_1 接受$(X,f_1(X,Y))$作为输入,工作步骤如下:

步骤 1. S_1 接受输入$(X,Y\subseteq X)$后,首先随机选取一个集合 $\overline{Y}=\{\bar{y}_1,\bar{y}_2,\cdots,\bar{y}_m\}$,使得 $f_1(X,Y)=f_1(X,\overline{Y})$;然后用$(X,\overline{Y})$来模拟.按照协议 1,根据 X 构造多项式系数$(a_0,a_1,\cdots,a_n)$,记作向量 $\boldsymbol{A}$.根据 $\overline{Y}$ 构造向量 $\bar{\boldsymbol{y}}'=(m,(\bar{y}_1+\bar{y}_2+\cdots+\bar{y}_m),(\bar{y}_1^2+\bar{y}_2^2+\cdots+\bar{y}_m^2),\cdots,(\bar{y}_1^n+\bar{y}_2^n+\cdots+\bar{y}_m^n))$,记作 $\boldsymbol{B}$.

步骤 2. S_1 选取较大随机数 $r'(r'\neq 0,1)$,计算多项式系数 $\boldsymbol{A}$ 的承诺$(g^{r'a_0},g^{r'a_1},\cdots,g^{r'a_n})$,记作向量 $\boldsymbol{A}'_{\text{commit}}$.

步骤 3. S_1 计算:

$$(g^{r'a_0})^m,(g^{r'a_1})^{(\bar{y}_1+\bar{y}_2+\cdots+\bar{y}_m)},(g^{r'a_1})^{(\bar{y}_1^2+\bar{y}_2^2+\cdots+\bar{y}_m^2)},\cdots,(g^{r'a_n})^{(\bar{y}_1^n+\bar{y}_2^n+\cdots+\bar{y}_m^n)},$$

再计算乘积:

$$(g^{r'a_0})^m(g^{r'a_1})^{(\bar{y}_1+\bar{y}_2+\cdots+\bar{y}_m)}\cdots(g^{r'a_1})^{(\bar{y}_1^2+\bar{y}_2^2+\cdots+\bar{y}_m^2)}=g^{r'[f(\bar{y}_1)+f(\bar{y}_2)+\cdots+f(\bar{y}_m)]},$$

得到 $g^{r'[f(\bar{y}_1)+f(\bar{y}_2)+\cdots+f(\bar{y}_m)]}$的结果,从而得到判断结果,记为 C'.

在本协议中,$view_1(X,Y)=\{X,\boldsymbol{A},r,A_{\text{commit}},C\}$,$S_1(X,\overline{Y})=\{X,\boldsymbol{A},r',A'_{\text{commit}},C'\}$.由于 $C=$

$(Y\subseteq X)$, $f_1(X,Y)=f_1(X,\overline{Y})$, 因此 $C'=(\overline{Y}\subseteq X)$, $C=C'$. 由于 $r\overset{c}{=}r'$, 因此 $A_{\text{commit}}\overset{c}{=}A'_{\text{commit}}$. 又因为 $output_2(X,Y)=f_2(X,Y)=(Y\subseteq X)$. 因此有:

$$\{(S_1(x,f_1(x,y)),f_2(x,y))\}\overset{c}{=}$$
$$\{(view_1(x,y),output_2(x,y))\}.$$

同理,用类似的方法可构造模拟器 S_2, 使得:

$$\{(f_1(x,y),S_2(y,f_2(x,y)))\}\overset{c}{=}$$
$$\{(output_1(x,y),view_2(x,y))\}.$$ 证毕.

定理 3. 在不可信第三方存在的情况下,协议 2 保密地判定了集合成员关系.

用定理 2 类似的方法可以证明定理 3,这里不再一一赘述.

6 效率分析与比较

本节给出本文协议和引言中相关文献[8,12-13]在计算复杂性、通信复杂性以及性能方面的比较.

1) 计算复杂度. 为了便于比较,统一本文和文献[8,12-13]中一方集合的势为 n, 另一方集合的势为 $m(n\geqslant m)$. 已存方案使用的多是公钥加密算法和匹配查找的方法,由于公钥加密算法的计算复杂性较高,而匹配查找的次数也制约着效率,因此以公钥加密算法的加解密次数和匹配查找次数作为衡量计算复杂性的指标. 文献[8]中的协议需要 n 次加密和 1 次门限解密,由于 1 次门限解密的复杂性远远大于 1 次单独解密的复杂性,为了有所区分,记 1 次门限解密为"**1**"次解密. 文献[12]中两方各需要 nm 次加密. 文献[13]中的协议需要 n 次加密和 1 次单独解密. 本文的 2 个协议均没有使用加解密算法.

2) 通信复杂度. 衡量通信复杂度的指标用协议交换信息的位数,或者用通信轮数(round),在多方保密计算研究中通常用轮数作为衡量通信复杂性的指标.

3) 性能. 在本文中以各个协议是否可公开计算、是否适合不可信第三方存在的场景作为衡量性能的指标.

综合以上分析,得到各个协议的效率比较如表 1 所示,性能比较如表 2 所示.

从表 1 可以看出,已存的所有方案都使用了公钥加密算法,而我们的 2 个协议并没有使用任何加解密算法,避免了烦琐的公私钥产生和加解密过程. 此外,我们的 2 个协议也没有使用任何匹配查找的方法,而文献[12]在集合很大时,最坏情况下需要查找比较 m 次,明显制约了效率. 因此从表 1 可以看出我们设计的协议计算复杂性和通信复杂性都较低,效率忧于其他方案.

Table 1 Efficiency Comparison Between Ours and the existing Schemes

表 1 本文协议与现有方案的效率比较

Scheme	Computation Cost		Communication Overhead
	Number of Encryption and Decryption	Number of Match and Search	
Ref [8]	$n+1$	0	4
Ref [12]	$2nm$	$1-m$	4
Ref [13]	$n+1$	0	3
Protocol 1	0	0	2
Protocol 2	0	0	3

Table 2 Performance Comparison Between Ours and the existing Schemes

表 2 本文协议与现有方案的性能比较

Scheme	Public Computation	Untrusted Third Party
Ref [8]	No	No
Ref [12]	No	No
Ref [13]	No	No
Protocol 1	No	No
Protocol 2	Yes	Yes

从表 2 可以看出,已存所有文献都只适用于传统模式:即只能参与者两方完成保密计算,并不适合不可信第三方存在的情况,也不能公开保密计算. 而我们在没有提高计算复杂性和通信复杂性的同时,首次针不可信第三方存在的应用场景,设计了实用型协议 2. 这使得我们开拓了安全计算集合包关系的新应用场景,比如云计算下的数据访问、加密数据搜索、可公开仲裁的抗抵赖等,也解决了引言中 Mu 等人[11]设计的加密方案中如何全隐私判断集合包含关系的问题.

7 总结和开放问题

集合包含关系的安全计算是安全多方计算中很重要的组成部分,但是这方面的研究并不多. 而已存在的方案大多利用了多次匹配查找和多次加解密算法,这在集合很大的情况下比较低效,并不实用. 本文将原问题转化为向量内积问题,然后基于数学难题和密码学知识解决了此问题. 我们设计的 2 个协

议,并没有使用任何加解密算法,避免了烦琐的公私钥产生和加解密过程,也没有使用匹配查找的方法,因此效率较高.此外,我们设计的协议2开拓了安全计算集合包含关系的新应用场景,具有较强的现实意义.

在本文的协议中,最重要的一个思想,就是将原问题转化.这种思想是安全多方计算的一个关键,因具体问题转化方法不同,直接制约着利用的密码学工具和方案的效率.本文将集合包含判断问题利用代数性质转化为内积的思想,进一步推广可以解决多种集合关系的保密判断,比如集合成员关系、集合的交集、并集等一类集合问题.反之,将这些安全多方计算中的集合关系嵌入到加密算法中,可以设计出不同以往的带有集合属性关系的加密方案.这些工作,我们都已经在陆续研究中.

我们的2个协议,为了提高效率,以接受性安全作为牺牲代价,有一定的信息泄露,虽然不影响协议的执行和有效性,但并未取得完美性安全.因此,可以进一步考虑如何设计关于此问题的完美安全性下的高效协议.以上这些问题都将是未来进一步研究的课题.

参 考 文 献

[1] Yao A C. Protocols for secure computations [C] //Proc of the 23rd IEEE Symp on Foundations of Computer Science. Piscataway, NJ: IEEE, 1982: 160-164

[2] Freudiger J, Rane S, Brito A E, et al. Privacy preserving data quality assessment for high-fidelity data sharing [C] //Proc of the ACM Workshop on Information Sharing & Collaborative Security. New York: ACM, 2014: 21-29

[3] Li Xiangyang, Jung T. Search me if you can privacy-preserving location query service [C] //Proc of IEEE INFOCOM'13. Piscataway, NJ: IEEE, 2013: 2760-2768

[4] Yang Jing, Zhao Jiashi, Zhang Jianpei. A privacy preservation method for high dimension data mining [J]. Acta Electronica Sinica, 2013, 41(11): 2187-2192 (in Chinese)
(杨静,赵家石,张健沛.一种面向高维数据挖掘的隐私保护方法[J].电子学报,2013,41(11):2187-2192)

[5] Samanthula B K, Elmehdwi Y, Howser G, et al. A secure data sharing and query processing framework via federation of cloud computing [J]. Information Systems, 2015, 48(c): 196-212

[6] Kerschbaum F. Privacy-preserving computation [G] //LNCS 8319: Proc of the Annual Privacy Forum. Berlin: Springer, 2014: 41-54

[7] Freedman M J, Nissim K, Pinkas B. Efficient private matching and set intersection [G] //LNCS 3027: Proc of Advances in Cryptology (EuroCrypt'04). Berlin: Springer, 2004: 1-19

[8] Kissner L, Song D. Privacy-preserving set operations [G] //LNCS 3621: Proc of Advances in Cryptology (Crypt'05). Berlin: Springer, 2005: 241-257

[9] Clifton C, Kantarcioglu M, Lin X D, et al. Tools for privacy-preserving distributed data mining [J]. ACM SIGKDD Explorations Newsletter, 2002, 4(2): 28-34

[10] Xia Feng, Yang Bo, Zhang Mingwu, et al. Secure two-party computation for set intersection and set equality problems based on LWE [J]. Journal of Electronics and Information Technology, 2012, 34(2): 462-467 (in Chinese)
(夏峰,杨波,张明武,等.基于LWE的集合相交和相等的两方保密计算[J].电子与信息学报,2012,34(2):462-467)

[11] Guo Fuchun, Mu Yi, Willy S. Subset membership encryption and its applications to oblivious transfer [J]. IEEE Trans on Information Forensics & Security, 2014, 9(7): 1098-1107

[12] Li Shundong, Si Tiange, Dai Yiqi. Secure multi-party computation about set-inclusion graph-inclusion [J]. Journal of Computation Research and Development, 2005, 42(10): 1647-165 (in Chinese)
(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):1647-165)

[13] Li Ronghua, Wu Chuankun, Zhang Yuqin. Secure computation protocols for testing the inclusion relation of sets [J]. Chinese Journal of Computers, 2009, 32(7): 1337-1346 (in Chinese)
(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):1337-1346)

[14] Mitsunari S, Sakai R, Kasahara M. A new traitor tracing [J]. IEICE Trans on Fundamentals of Electronics, Communications and Computer Sciences, 2002, 85(2): 481-484

[15] Atallah M J, Du W L. Secure multi-party computational geometry [G] //LNCS 2125: Proc of the Workshop on Algorithms and Data Structures. Berlin: Springer, 2001: 1165-1179

[16] Liu Fang, Ng W K, Zhang Wei. Secure scalar product for big-data in MapReduce [C] //Proc of 2015 1st IEEE Int Conf on Big Data Computing Service and Applications. Piscataway, NJ: IEEE, 2015: 120-129

[17] Calvino A, Ricci S, Domingo-Ferrer J. Privacy-preserving distributed statistical computation to a semi-honest multi-cloud [C] //Proc of IEEE Conf on Communications and Network Security. Piscataway, NJ: IEEE, 2015: 506-514

[18] Zhu Youwen, Takagi T. Efficient scalar product protocol and its privacy-preserving application [J]. International Journal of Electronic Security and Digital Forensics, 2015, 7(1): 1-19

[19] Mohammed N, Alhadidi D, Fung B, et al. Secure two-party differentially private data release for vertically partitioned data [J]. IEEE Trans on Dependable and Secure Computing, 2014, 11(1): 59-71

[20] De I, Tripathy A. A secure two party hierarchical clustering approach for vertically partitioned data set with accuracy measure [G] //Proc of the Advances in Intelligent Systems and Computing. Berlin: Springer, 2014: 153-162

[21] Sheng Gang, Wen Tao, Guo Quan, et al. Privacy preserving inner product of vectors in cloud computing [J]. International Journal of Distributed Sensor Networks, 2014, 6: 1-6

[22] Dong Changyu, Chen Liqun. A fast secure dot product protocol with application to privacy preserving association rule mining [G] //LNCS 8443: Proc of the Advances in Knowledge Discovery and Data Mining. Berlin: Springer, 2014: 606-617

[23] Goldreich O. Foundations of Cryptography: Basic Applications [M]. London: Cambridge University Press, 2004: 599-729

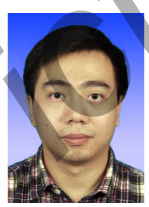
[24] Feldman P. A practical scheme for non-interactive verifiable secret sharing [C] //Proc of the 28th IEEE Annual Symp on Foundations of Computer Science. Piscataway, NJ: IEEE, 1987: 427-438

[25] Du Wenliang, Zhan Zhijun. A practical approach to solve secure multi-party computation problems [C] //Proc of the 2002 Workshop on New Security Paradigms. New York: ACM, 2002: 127-135

Chen Zhenhua, born in 1976. PhD. Associate professor. Her main research interests include secret sharing and secure multi-party computation, etc.

Li Shundong, born in 1963. PhD, professor. His main research interests include information hiding and secure multi-party computation, etc.

Wang Daoshun, born in 1964. PhD, associate professor. His main research interests include key management, digital watermarking and multimedia security, etc.



Huang Qiong, born in 1981. PhD, professor. Member of CCF and Chinese Association for Cryptologic Research. His main research interests include cryptography and information security.

Dong Lihong, born in 1968. PhD, professor. Her main research interests include scientific computation and system engineer, etc.