保密计算交集对应元素和的最大值

马秀莲,张倦倦,李顺东
(陕西师范大学计算机科学学院,陕西西安 710119)

摘 要: 安全多方计算是国际密码学的研究热点之一,隐私集合问题是安全多方计算的重要研究方向. 本文提出了一个新的安全多方计算问题:Alice 和 Bob 分别拥有集合 $X=\{(v_i,x_i)\}_{i=1}^{l_1}$ 和 $Y=\{(w_j,y_j)\}_{j=1}^{l_2}$,他们想要保密计算 $v_i=w_j$ 时的最大值 $\max(x_i+y_j)$. 该问题在教育、网购等领域具有重要的理论和现实意义. 针对这个问题,我们在半诚实模型下提出了两个安全协议. 第一个协议基于编码方法和保密移位思想,适用于元组中元素数据范围已知的情况. 第二个协议利用 Paillier 密码算法和添加假元素的方法,适用于元素数据范围未知的情况. 最后,我们使用公认的模拟范式证明了两个协议是安全的.

关键词: 密码学;安全多方计算;和的最大值;同态加密;模拟范例
基金项目: 国家自然科学基金(No.61272435)
中图分类号: TP309 **文献标识码:** A **文章编号:** 0372-2112(2023)07-1835-07
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20221177

Maximum Value of Sum of Intersection Elements of Secret Calculation

MA Xiu-lian, ZHANG Juan-juan, LI Shun-dong
(*School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi* 710119, *China*)

Abstract: Secure multi-party computation is one of the research hotspots in the international cryptographic community. Privacy set problem is an important research direction of secure multi-party computation. In this paper, we propose a novel secure multi-party computation problem, i.e., Alice and Bob have two private sets $X=\{(v_i,x_i)\}_{i=1}^{l_1}$ and $Y=\{(w_j,y_j)\}_{j=1}^{l_2}$, and they want to compute $\max(x_i+y_j)$ where $v_i=w_j$. This problem has important theoretical and practical significance in education, online shopping and other fields. In response to this problem, we propose two security protocols in the semi-honest model. The first protocol, which is based on an encoding method and secure shift, is for the case where two tuples of elements of the private sets are elements of a known set whose cardinality is not large. The second protocol, which is based on Paillier cryptosystem and the method of adding fake elements, is for the case where two tuples of elements of the private sets are not known. Finally, we demonstrate the security of both protocols using recognized simulation paradigm.

Key words: cryptography; secure multi-party computation; maximum value of the sum; homomorphic encryption; simulation paradigm
Foundation Item(s): National Natural Science Foundation of China (No.61272435)

1 引言

数据共享促进了整个社会的发展,但人们希望在共享数据时能够保护敏感信息. 如何保护个人敏感数据并实现共享是信息时代需要解决的难题. 安全多方计算[1]可以很好地解决这一难题,能够在保护敏感数据的同时实现数据共享.

姚期智教授以百万富翁问题[1]开创了安全多方计算的先河. 随后,在诸多学者的不断推动下,安全多方计算成为密码学中非常重要的研究领域. 在研究过程中,Goldreich[2]等人设计了安全多方计算的通用方案,但对于具体问题,通用方案的计算复杂度和通信复杂度都很高. 因此,有必要针对具体问题设计一个安全高效的保密协议.

隐私集合问题是安全多方计算的重要问题. 关于隐私集合问题大致分为:判断元素是否属于集合[3,4]、保密计算集合交集[5,6]、保密计算交集的势[7,8]、保密判

收稿日期:2022-10-17;修回日期:2023-06-19;责任编辑:李勇锋

断集合间的关系[9,10]、保密计算交集元素和[11~14]等. 其中,文献[4]将集合表示成多项式,通过判断元素是否为多项式的根来确定元素与集合的关系. 文献[6]结合隐私求交方法和局部敏感哈希函数提出了高效的隐私保护链接协议. 文献[7]利用布隆过滤器和ElGamal密码算法设计了交集势的多方计算协议. 文献[11]利用DDH问题和Paillier密码算法计算集合交集的势和标识符关联的整数值的总和. 文献[12]利用Hash技巧和不经意传输技术,在保护交集元素情况下保密计算交集的势、交集权值的统计和以及权值的方差. 文献[13]结合投票场景,提出了隐私交叉加权和的协议. 文献[14]结合广告转化率,利用不经意传输和布隆过滤器保密计算了交集的势和对应整数值的总和. 由此可见,保密计算交集、交集和等研究已经有了比较丰富的成果,但目前还没有保密计算交集对应元素和的最大值的解决方案. 使用传统的隐私集合求交协议来保密计算交集对应元素和的最大值会泄露交集元素与交集的势. 采用交集和保密协议解决交集对应元素和的最大值问题也不适用,通过先计算交集和再计算最大值,则无法保护交集的和. 因此,本文在保护交集元素、元素和的情况下,设计了保密计算集合交集对应元素和的最大值协议.

本文研究集合交集对应元素和的最大值问题,具体描述为:一方拥有集合$X=\{(v_i, x_i)\}_{i=1}^{l_1}$,另一方拥有集合$Y=\{(w_j,y_j)\}_{j=1}^{l_2}$,其中$\{v_i\}$,$\{w_j\}$是关键字集合;$x_i,y_j$分别是关键字$v_i,w_j$对应的关联值. 双方想要保密计算相同关键字对应关联值之和的最大值,即$v_i=w_j$时,x_i+y_j的最大值. 我们将相同关键字对应关联值之和的最大值称为交集对应元素和的最大值. 该问题在现实生活中具有重要意义. 例如,在选择多门课程的学生中保密计算最高总成绩. 在保护用户购买记录情况下计算同时在多个平台购物用户的最高消费额等. 本文集合的每个元素由二元组组成,针对元组分量有无全集的情况分别设计了具体的协议. 本文主要贡献包括:(1)从实际应用中提炼出一个新问题,即保密计算属性交集对应的关联值之和的最大值;(2)利用新的编码和移位寄存器思想,解决了元组分量有全集限制下的保密计算交集对应元素和的最大值问题;(3)通过添加假元素的方法来混淆实际数据,在保护交集势及私有数据的条件下设计了元组分量无全集的协议. 通过模拟范例对协议的安全性进行了严格证明.

2 预备知识

2.1 安全性定义

半诚实模型半诚实参与者[15]严格遵守协议规定发送正确的信息,但在协议执行后试图推算出其他参与者隐私数据的额外信息. 本文研究半诚实模型下的两方问题. 假设Alice和Bob利用协议π保密计算多项式时间函数$f(x,y)$. 在执行协议过程中,Alice获得的信息序列记为

$$\text{view}_i^{\pi}(x,y)=\left(x,r_i,M_1^i,\cdots,M_i^t\right)$$

其中,r_i是Alice选的随机数,$M_i^j\left(j=1,\cdots,t\right)$表示Alice收到的第$j$个消息. Bob得到的信息序列也可类似定义.

半诚实模型下的安全性对于f,若存在多项式算法S_1、S_2(称多项式时间算法为模拟器),使得:

$$\{S_1(x,f_1(x,y))\}_{x,y}\overset{c}{\equiv}\{\text{view}_1^{\pi}(x,y)\}_{x,y} \tag{1}$$

$$\{S_2(y,f_2(x,y))\}_{x,y}\overset{c}{\equiv}\{\text{view}_2^{\pi}(x,y)\}_{x,y} \tag{2}$$

则称π保密地计算f,其中$\overset{c}{\equiv}$表示计算不可区分.

模拟范例模拟范例是安全多方计算证明安全性的一种常用方法. 证明协议在半诚实模型下是安全的,需构造满足(1)式和(2)式的模拟器来证明.

2.2 Paillier密码系统

Paillier加密算法是Paillier基于合数剩余类困难问题提出的概率公钥加密方法,具体描述如下:

密钥生成选取两个素数p,q,计算$N=p\times q,\lambda=\text{lcm}(p-1,q-1)$. λ是$p-1$和$q-1$的最小公倍数. 随机选择$g\in Z_N^*$,使得$\gcd\left(L\left(g^{\lambda}\bmod N^2\right),N\right)=1$,其中$L(x)=\frac{x-1}{N}$. 算法的公钥为$(g,N)$,私钥为$\lambda$.

加密过程对明文$m\in Z_N$,选取随机数$r\in Z_N^*$,计算:

$$c=g^m r^N \bmod N^2$$

解密过程对密文$c\in Z_{N^2}^*$,计算:

$$m=\frac{L\left(c^{\lambda}\bmod N^2\right)}{L\left(g^{\lambda}\bmod N^2\right)}\bmod N$$

加法同态性:

$$\begin{aligned}E(m_1)\times E(m_2)&=g^{m_1}r_1^N g^{m_2}r_2^N \bmod N^2\\&=g^{m_1+m_2}(r_1r_2)^N \bmod N^2\\&=E(m_1+m_2)\bmod N^2\end{aligned}$$

同时,进一步得到Paillier加密方案的下述性质:

$$E(m_1)^{m_2}\bmod N^2=E(m_1m_2)\bmod N^2$$

2.3 保密替换和保密移位

保密替换是文献[16]提出的一种密文上的操作,解决了最值问题及最小公倍数问题. 保密替换有非常多的用途,如同时计算最大值和最小值、排序等. 保密移位是文献[17]利用概率算法和移位寄存器提出的一种方法,解决了盲百万富翁问题[17],还可以用于判定三条边能否构成一个三角形. 这两种密文上的操作是安全多方计算中非常有用的方法,具体原理和操作方法请参考文献[16]和文献[17].

3　交集对应元素和的最大值(有全集)

问题描述:Alice有集合$X=\{(v_i, x_i)\}_{i=1}^{l_1}$,Bob有集合$Y=\{(w_j, y_j)\}_{j=1}^{l_2}$. 记$A=\{v_i\}_{i=1}^{l_1}$,$B=\{w_j\}_{j=1}^{l_2}$. 设$A,B\subseteq Q=\{q_1,q_2,\cdots,q_l\}$;$\{x_i\}_{i=1}^{l_1},\{y_j\}_{j=1}^{l_2}\subseteq U=\{u_1,u_2,\cdots,u_m\}(u_1<u_2<\cdots<u_m)$. 其中$Q$、$U$分别为关键字、关联值集合. Alice和Bob希望计算相同关键字的最大关联值之和,而不泄露其他信息.

若Q为学生姓名集合,U为课程成绩,x_i、y_j分别为语文、数学成绩,则可以保密计算选两门课学生的最高总成绩.

计算原理:Alice和Bob根据A,B和Q,U执行下面操作:

(1) Alice根据全集构造$X'=\{(q_i,x_i')\}_{i=1}^{l}$,其中,若$(q_i\in A)\cap(q_i=v_k)$,则$x_i'=x_k$,若$q_i\notin A$,则$x_i'=0$. Bob用同样方法对$Y$编码得到$Y'=\{(q_i, y_i')\}_{i=1}^{l}$.

(2) Alice对X'中关键字q_i对应的关联值x_i'编码为向量$\boldsymbol{V}_i=(v_{i1},v_{i2},\cdots,v_{i(2m)})$:若$q_i\notin A(x_i'=0)$,则$V=0$;若$x_i'=u_k$,则$v_{ik}=1,v_{it}=0(t\neq k)$. 这样Alice将$X'$编码成一个$l\times 2m$的矩阵发给Bob.

(3) Bob对矩阵进行操作:若$q_i\notin B$,则将矩阵的第i行元素全部替换为0;否则将第i行元素向右移动y_i'位. 操作完成后将矩阵每一列元素相加得到向量$V=(v_1,v_2,\cdots,v_{2m})$. 然后Bob将$V$的分量从后向前逐一发给Alice.

(4) Alice找到第一个不为0的数停止后面查找. 第一个不为0的数所对应的关联值为交集对应元素和的最大值.

以计算原理为基础,我们基于Paillier加密算法设计出有全集限制下的保密计算协议1.

协议1　交集对应元素和的最大值(有全集)

输入: Alice输入集合$X=\{(v_i,x_i)\}_{i=1}^{l_1}$,Bob输入集合$Y=\{(w_j,y_j)\}_{j=1}^{l_2}$. 记$A=\{v_i\}_{i=1}^{l_1}$,$B=\{w_j\}_{j=1}^{l_2}$. 设$A,B\subseteq Q$,$\{x_i\}_{i=1}^{l_1},\{y_j\}_{j=1}^{l_2}\subseteq U$. 其中$Q=\{q_1,q_2,\cdots,q_l\}$为关键字集,$U=\{u_1,u_2,\cdots,u_m\}(u_1<u_2<\cdots<u_m)$为关联值集合.

输出:$f(X,Y)=\max(x_i+y_j)(v_i=w_j)$.

(1) Alice运行Paillier算法,生成公钥和私钥.

(2) 根据Q和U,Alice构造新集合$X'=\{(q_i,x_i')\}_{i=1}^{l}$. 其中,若$(q_i\in A)\cap(q_i=v_k)$,则$x_i'=x_k$;否则$x_i'=0$. 同样Bob得到$Y'=\{(q_i,y_i')\}_{i=1}^{l}$.

(3) Alice对X'中关键字对应的关联值编码得到$\boldsymbol{V}_i=(v_{i1},v_{i2},\cdots,v_{i(2m)})$,并将所有$\boldsymbol{V}_i$组成一个$l\times 2m$的矩阵,加密后发送给Bob.

(4) Bob在加密的矩阵上进行操作:若Y'中的关键字q_i不在Y中,则将第q_i行元素全部替换为$E(0)$;否则将q_i行元素$E(x_i')$右移y_i'位去掉y_i'个密文,在第q_i行的最前y_i'个位置添加y_i'个0的密文得到$E(x_i'+y_i')$. Bob将全集$u_1,u_2,\cdots,u_{2m}$对应的每列相乘,然后从后往前逐一发给Alice.

(5) Alice使用私钥依次解密Bob发的密文,解密遇到$(x_k'+y_k')\neq 0(k\in\{1,2,\cdots,2m\})$时,终止协议. 此时$x_k'+y_k'$在全集$u_1,u_2,\cdots,u_{2m}$对应的值$\max(x_i+y_j)$为交集对应元素和的最大值.

3.1　具体协议

3.2　协议的正确性

定理1　协议1能正确地保密计算有全集限制的交集对应元素和的最大值.

证明　协议1在两个分量有全集情况下利用编码、保密替换、保密移位思想计算交集对应元素和的最大值. Alice将关键字q_i对应的关联值编码为向量$\boldsymbol{V}_i=(v_{i1},v_{i2}\cdots,v_{i(2m)})$,并将向量组成一个$l\times 2m$的矩阵,$v_{ij}=1(j\in\{1,2,\cdots,2m\})$对应的数据是$x_i'$,将向量$\boldsymbol{V}_i$整体右移$y_i'$,1对应的数据变成$x_i'+y_i'$. 从右往左的第一个不为0的分量在全集$u_1,u_2,\cdots,u_{2m}$中对应的数据为交集对应元素和的最大值$\max(x_i+y_j)$. 因此协议1有全集限制的条件下能正确计算交集对应元素和的最大值.

证毕

3.3　协议的安全性

定理2　协议1在半诚实模型下是安全的.

证明　下面在半诚实模型下,采用构造模拟器S_1和S_2来证明定理2. 在协议1中,

$$\text{view}_1^{\pi}(X,Y)=\{X,r_1,C\},$$

$$\text{view}_2^{\pi}(X,Y)=\{Y,r_2,E(X')\}$$

其中,$X=\{(v_i,x_i)\}_{i=1}^{l_1}$、$Y=\{(w_j,y_j)\}_{j=1}^{l_2}$分别是Alice和Bob的输入,$r_1$是Alice加密时选择的随机数集合,$C$是Bob发给Alice的向量,$C$由$E(x_{2m}'+y_{2m}'),\cdots,E(x_k'+y_k')$组成(假设$(x_k'+y_k')$是满足当$v_i=w_j$时,$x_i+y_j$的最大值). r_2是Bob计算时选择的随机数集合,$E(X)$是Alice发给Bob的密文矩阵.

首先,构造模拟器S_1模拟$\text{view}_1^{\pi}(X,Y)$来证明Bob数据的安全性,S_1模拟过程如下:

(1) 接收输入$(X,f(X,Y))$,根据$f(X,Y)$的值,S_1选择$Y'=\{(q_i,y_j')\}_{j=1}^{l_2}$,使得$f(X,Y)=f(X,Y')$;

(2) S_1根据协议1将Y'编码为集合$Y''=\{(q_i,y_i')\}_{i=1}^{l}$;

(3) S_1根据Y''对矩阵X'的第q_i行进行操作:若Y''中的关键字q_i不在Y'中,则将q_i行元素全部替换为0;否则将q_i行元素x_i右移y_i'位,得到$x_i'+y_i'$. 操作完成后,

S_1将矩阵的每一列相乘，从后往前选择$2m-k+1$个密文构成C'.

对于Alice而言，X、r_1是自己的数据，由于Paillier加密算法是语义安全的，所以真实环境得到的C和模拟得到的C'是计算不可区分的，依次解密得到的是前面是固定数量的0，最后是一个1. 令$S_1\left(X,f\left(X,Y\right)\right)=\left\{X,r_1,C'\right\}$，则有下式成立：

$$\{S_1(X,f(X,Y))\}_{X,Y}\overset{c}{\equiv}\{\text{view}_1^{\pi}(X,Y)\}_{X,Y}$$

类似地，构造模拟器S_2模拟$\text{view}_2^{\pi}(X,Y)$来证明Alice数据的安全性，S_2模拟过程如下：

（1）接收输入$\left(Y,f\left(X,Y\right)\right)$，根据$f(X,Y)$的值，$S_2$选择集合$X'=\{(q_i',x_i')\}_{i=1}^{l_1}$，使得$f(X,Y)=f(X',Y)$.

（2）S_2将X'编码为$X''=\{(q_i,x_i')\}_{i=1}^{l}$，加密得到$E(X'')=\{(q_i,E(x_i'))\}_{i=1}^{l}$.

对于Bob而言，S_2得到的$E(X'')$和真实环境得到的$E(X')$不可区分. 令$S_2\left(Y,f\left(X,Y\right)\right)=\left\{Y,r_2,E\left(X''\right)\right\}$，则有下式成立：

$$\{S_2(Y,f(X,Y))\}_{X,Y}\overset{c}{\equiv}\{\text{view}_2^{\pi}(X,Y)\}_{X,Y}$$

因此，协议1在半诚实模型下是安全的.

证毕

4 交集对应元素和的最大值(无全集)

问题描述：Alice有集合$X=\{(v_i,x_i)\}_{i=1}^{l_1}$，Bob有集合$Y=\{(w_j,y_j)\}_{j=1}^{l_2}$. 其中，$\{v_i\}_{i=1}^{l_1}$和$\{w_j\}_{j=1}^{l_2}$是关键字集合；$\{x_i\}_{i=1}^{l_1}$和$\{y_j\}_{j=1}^{l_2}$是关键字对应的关联值集合. Alice和Bob希望计算相同关键字的最大关联值之和，而不泄露其他信息.

计算原理：双方对关键字和关联值执行下面操作.

（1）Alice添加l_3个关键字及关联值$\left(v_a,0\right)\left(a=1,2,\cdots,l_3\right)$，得到$X'=\{(v_i,x_i)\}_{i=1}^{l_1+l_3}$. 其中$\{v_a\}_{a=1}^{l_3}\cap\{v_i\}_{i=1}^{l_1}=\varnothing$.

（2）Bob添加l_4个关键字及关联值$\left(w_b,0\right)\left(b=1,2,\cdots,l_4\right)$，得到$Y'=\{(w_j,y_j)\}_{j=1}^{l_2+l_4}$. 其中$\{w_b\}_{b=1}^{l_4}\cap\{w_j\}_{j=1}^{l_2}=\varnothing$.

（3）Bob根据Y'构造$\left\{\left(m_{ij},n_{ij}\right)\right\}$. 具体操作：将$X'$中的$v_i$与$Y'$的$w_j$进行相减得到$m_{ij}=v_i-w_j$；将$X'$中的$x_i$与$Y'$中的$y_j$相加得到$n_{ij}=x_i+y_j$，共计算$(l_1+l_3)(l_2+l_4)$次. Bob将$m_{ij}$，$n_{ij}$组成元组$A_{ij}=(m_{ij},n_{ij})$.

（4）Bob选择随机数r_{ij}，r_1，$r_2(r_1>0)$构造$B_{ij}=\left\{\left(m_{ij}\times r_{ij},n_{ij}\times r_1+r_2\right)\right\}(i=1,2,\cdots,(l_1+l_3);\ j=1,2,\cdots,(l_2+l_4))$. Bob将$B_{ij}$的位置随机置换为$C_{ij}$，并将$C_{ij}$的第一个分量发给Alice.

（5）Alice将第一个分量为0的对应位置$\{C_{ij}\}$发给Bob. 接着，Bob将对应的$\{C_{ij}\}$的第二个分量发给Alice. Alice找到最大数$\max\left(\left(x_i+y_j\right)\times r_1+r_2\right)$发给Bob.

（6）Bob在最大数基础上先减去r_2再除以r_1，得到交集对应元素和的最大值.

同样基于Paillier算法设计无全集限制下的保密协议2.

协议2 交集对应元素和的最大值(无全集)

输入：Alice输入$X=\{\left(v_i,x_i\right)\}_{i=1}^{l_1}$，Bob输入$Y=\{(w_j,y_j)\}_{j=1}^{l_2}$.

输出：$f(X,Y)=\max(x_i+y_j)(v_i=w_j)$.

(1) Alice运行Paillier算法，生成公钥和私钥.

(2) Alice添加l_3个关键字及关联值$\left(v_a,0\right)\left(a=1,2,\cdots,l_3\right)$，得到$X'=\{(v_i,x_i)\}_{i=1}^{l_1+l_3}$. 其中$\{v_a\}_{a=1}^{l_3}\cap\{v_i\}_{i=1}^{l_1}=\varnothing$. 然后，Alice加密$\{(E\left(v_i\right),E(x_i))\}_{i=1}^{l_1+l_3}$发给Bob.

(3) Bob添加l_4个关键字及关联值$\left(w_b,0\right)\left(b=1,2,\cdots,l_4\right)$，其中$\{w_b\}_{b=1}^{l_4}\cap\{w_j\}_{j=1}^{l_2}=\varnothing$. Bob得到$Y'=\{(w_j,y_j)\}_{j=1}^{l_2+l_4}$.

(4) Bob计算$E\left(v_i\right)E\left(N-w_j\right)$和$E\left(x_i\right)E\left(y_j\right)$，得到$\left\{\left(E\left(v_i-w_j\right),E\left(\mathrm{x}_i+y_j\right)\right)\right\}\left(i=1,2,\cdots,\left(l_1+l_3\right);j=1,2,\cdots,\left(l_2+l_4\right)\right)$.

(5) Bob选择随机数r_{ij}，r_1，$r_2(r_1>0)$计算$E(v_i-w_j)^{r_{ij}}$，$E(x_i+y_j)^{r_1}E(r_2)$. 利用Paillier同态性得到$\left\{\left(E\left(v_i-w_j\right)\times r_{ij},E\left(x_i+y_j\right)\times r_1+r_2\right)\right\}$. Bob随机置换加密后的集合元素，将分量$\left\{E\left(\left(v_i-w_j\right)\times r_{ij}\right)\right\}$发给Alice.

(6) Alice解密后将第一个分量为0的对应所有位置发给Bob. 接着，Bob将对应位置的第二个分量$E((x_i+y_j)\times r_1+r_2))$发给Alice. Alice解密找到最大数$\max\left(\left(x_i+y_j\right)\times r_1+r_2\right)$发给Bob.

(7) Bob在最大数的基础上先减去随r_2再除以r_1，得到$\max\left(x_i+y_j\right)$为交集对应元素和的最大值.

4.1 具体协议

4.2 协议的正确性

定理3 协议2在无全集限制下能正确地保密计算交集对应元素和的最大值.

证明 协议2利用Paillier同态性质和添加假元素的方法计算交集对应元素之和的最大值. Alice和Bob

添加假元素$(v_a,0)$和$(w_b,0)$. Bob选择随机数r_{ij},r_1,r_2计算得到$\{(v_i-w_j)\times r_{ij},(x_i+y_j)\times r_1+r_2\}\}$. Bob将计算后的集合随机置换后,将第一个分量发给Alice,这样保证了集合有交集,交集可能由Alice和Bob添加的假元素得到,但假元素第二个分量为0,不会对交集对应元素和的正确性有影响,从而不会对交集对应元素和的最大值有影响. Bob将交集的第二个分量发给Alice. 接着,Alice找到最大数$\max((x_i+y_j)\times r_1+r_2)$发给Bob,因为$r_1>0$,所以对于正数$x_1>x_2$,$x_1\times r_1+r_2>x_2\times r_1+r_2$是成立的. Bob用最大数减去$r_2$除以$r_1$得到交集对应元素和的最大值. 因此,协议2在无全集限制下能正确地保密计算交集对应元素和的最大值.

证毕

4.3 协议的安全性

定理4 协议2在半诚实模型下是安全的.

证明 协议2的安全性类似于协议1的安全性,用构造模拟器S_1和S_2来证明定理4. 在协议2中,

$$\text{view}_1^{\pi}(X,Y)=\{X,r_3,E(Y_1),E(Y_2)\},$$

$$\text{view}_2^{\pi}(X,Y)=\{Y,r_4,E(X),P_1,P_2\}$$

其中,$X=\{(v_i,x_i)\}_{i=1}^{l_1}$、$Y=\{(w_j,y_j)\}_{j=1}^{l_2}$分别是Alice和Bob的输入,$r_3$是协议2中Alice加密时选择的随机数集合,$E(Y_1),E(Y_2)$是分别是Bob第一次和第二次发给Alice的密文,$E(Y_1)=\{(E((v_i-w_j)\times r_{ij})\}$,$E(Y_2)=\{E((x_i+y_j)\times r_1+r_2))\}$. r_4是协议2中Bob加密和计算时选的随机数集合. $E(X)$是Alice发给Bob的密文,P_1是Alice根据第一个分量计算结果发给Bob的位置信息,P_2是Alice发给Bob的明文信息. 其中$E(X)=\{(E(v_i),E(x_i))\}_{i=1}^{l_1+l_3}$,$P_2=\max((x_i+y_j)\times r_1+r_2)$.

同样,S_1模拟$\text{view}_1^{\pi}(X,Y)$证明Bob的安全性,具体如下:

(1) 接收输入$(X,f(X,Y))$,根据$f(X,Y)$的值,S_1选择集合$Y'=\{(w_j',y_j')\}_{j=1}^{l_2+l_4}$,使得$f(X,Y)=f(X,Y')$.

(2) S_1在$\{(E(v_i),E(x_i))\}_{i=1}^{l_1+l_3}$基础上计算$E(v_i)E(N-w_j')$和$E(x_i)E(y_j')$,得到$\{(E(v_i-w_j'),E(x+y_j'))\}$. S_1选三个随机数r_{ij}',r_1',r_2'计算$E(v_i'-w_j')^{r_{ij}'}$和$E(x_i+y_j')^{r_1'}E(r_2')$,得到$\{E((v_i-w_j')\times r_{ij}'),E((x_i+y_j')^{r_1'}+r_2')\}$.

(3) S_1随机置换加密后的集合,得到$E(Y_1')=\{E((v_i-w_j')\times r_{ij}')\}$. S_1根据第一个分量为0的位置,得到$E(Y_2')=\{E((x_i+y_j')r_1'+r_2')\}$.

对于Alice而言,X、r_3是本身拥有的数据,模拟得到的$E(Y_1'),E(Y_2')$和真实得到的$E(Y_1),(Y_2)$是不可区分的. 令$S_1(X,f(X,Y))=\{X,r_3,E(Y_1),E(Y_2)\}$,则有下式成立:

$$\{S_1(X,f(X,Y))\}_{X,Y}\overset{c}{\equiv}\{\text{view}_1^{\pi}(X,Y)\}_{X,Y}$$

类似地,构造模拟器S_2模拟$\text{view}_2^{\pi}(X,Y)$,具体如下:

(1) 接收输入$(Y,(X,Y))$,根据$f(X,Y)$,S_2选择集合$X'=\{(v_i',x_i')\}_{i=1}^{l_1+l_3}$,使得$f(X,Y)=f(X',Y)$.

(2) S_2加密得到$E(X')=\{(E(v_i'),E(x_i'))\}_{i=1}^{l_1+l_3}$.

(3) S_2根据第一个分量$\{E((v_i'-w_j)\times r_{ij})\}$,得到第一个分量为0的位置$P_1'$. 然后,$S_2$在第二个分量中得到最大数$P_2'=\max((x_i'+y_j)\times r_1+r_2)$.

同理,对于Bob而言,模拟得到的$E(X'),P_1',P_2'$和真实得到的$E(X),P_1,P_2$是计算不可区分的. 令$S_2(Y,f(X,Y))=\{Y,r_4,E(X'),P_1',P_2'\}$,则有下式成立:

$$\{S_2(Y,f(X,Y))\}_{X,Y}\overset{c}{\equiv}\{\text{view}_2^{\pi}(X,Y)\}_{X,Y}$$

因此,协议2在半诚实模型下是安全的.

证毕.

5 效率分析

本文与文献[15~17]都研究集合交集的相关问题,但研究问题完全不同. 本文研究的保密计算交集对应元素和的最大值是一个新问题,尚未有解决方案. 因此无法将本文与现有方案对比. 下面我们对本文方案进行分析. 方案以模指数运算次数作为复杂性度量指标,其中Paillier算法每一次加密需要2次模指数运算,解密一次需要1次模指数运算.

5.1 计算复杂性

协议1中,Alice需要对$l\times 2m$的矩阵加密1次,解密次数跟关联值全集m有关,最好情况解密1次,最坏情况解密$2m$次,本文取平均m次. Bob在$l\times 2m$的矩阵上操作,因此模指数运算$l\times 2m$次,Bob不解密. 因此协议1的模指数运算次数为$4(l\times 2m)+m=8lm+m$(l,m分别为关键字集合、关联值集合的势).

协议2中,Alice、Bob加密次数跟添加的假元素个数l_3、l_4有关. Alice加密(l_1+l_3)次. Alice进行两轮解密:第一轮Alice解密$(l_1+l_3)(l_2+l_4)$次,第二轮解密跟交集的势有关. 因此解密次数为$\left|\{v_i\}_{i=1}^{l_1+l_3}\cap\{w_j\}_{j=1}^{l_2+l_4}\right|$. Bob进行两轮加密:加密自己数据共$(l_2+l_4)$次,第二轮加密跟添加假元素后交集的势有关,加密次数为$\left|\{v_i\}_{i=1}^{l_1+l_3}\cap\{w_j\}_{j=1}^{l_2+l_4}\right|$. Bob不解密. 因此,模指数运算为$2((l_1+l_3)+(l_2+l_4))+4\times\left|\{v_i\}_{i=1}^{l_1+l_3}\cap\{w_j\}_{j=1}^{l_2+l_4}\right|$($l_1$、$l_2$分别为集合$X$、$Y$的势,$l_3$、$l_4$分别为Alice和Bob添加假元素的个数).

5.2 通信复杂性

安全多方计算中通常用通信轮数衡量通信复杂度.

协议1的通信轮数与解密过程中第一个不为0的数有关，即跟关联值集合的势m有关. 最好情况解密1次，最坏情况解密$2m$次，本文取平均值，即解密m次得到第一个不为0的数. 因此，协议1需要$m+2$轮通信. 协议2需要4轮通信.

表1 协议计算复杂性与通信复杂性

	协议1	协议2
计算复杂性	$8lm+m$	$2((l_1+l_3)+(l_2+l_4))+4\times\left\lvert\{v_i\}_{i=1}^{l_1+l_3}\cap\{w_j\}_{j=1}^{l_2+l_4}\right\rvert$
通信复杂性	$m+2$	4

5.3 实验数据分析

测试环境：Windows 10 64位操作系统，处理器是Intel(R) Core(TM) i5-6600 CPU@3.30 GHz，内存是8 GB，在PyCharm用Python 3.6.4语言运行实现.

实验方法：本文协议均采用Paillier算法，设定素数的比特数为1 024，两个协议的执行时间跟集合$X=\{(v_i,x_i)\}_{i=1}^{l_1}$的势$l_1$，$Y=\{(w_j,y_j)\}_{j=1}^{l_2}$的势$l_2$有关. 所有实验设定$l_1$、$l_2$依次取5，10，…，50. 所有实验进行100次，统计平均值(忽略协议中预处理数据时间). 具体如下：

协议1的执行时间跟关键字集Q的势l和关联值集U的势m有关. l的取值和l_1同步，m的取值和l_2同步. 图1表示执行时间随全集势变化的二维图. 通过分析可知，执行时间随关键字集Q的势l和关联值集U的势m的增加而增长.

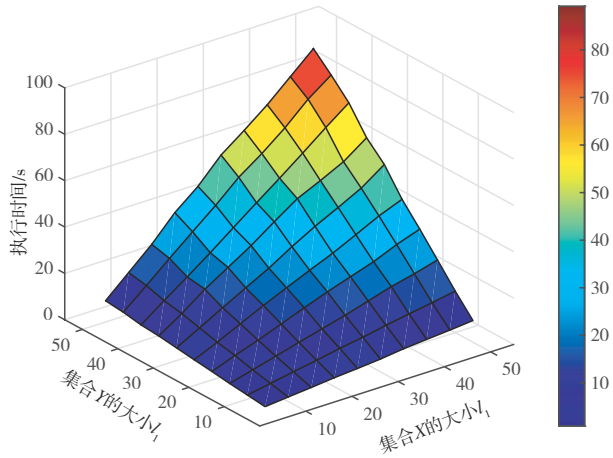

图1 有全集时执行时间随l_1、l_2的变化规律

协议2的执行时间不仅与X、Y的势l_1、l_2有关，还与假元素数量l_3和l_4有关. 所以实验2考虑假元素数量固定(取l_3=4，l_4=6)，实验3考虑假元素数量变化情况，即假元素数量从[1，10]中随机选择. 图2表示假元素固定时执行时间随l_1、l_2变化图. 图3表示假元素随机取值时，执行时间随(l_1+l_3)、(l_2+l_4)变化图.

通过分析图2和图3可知协议2中添加假元素越多，安全性越高，但效率越低. 因此使用协议2时需要

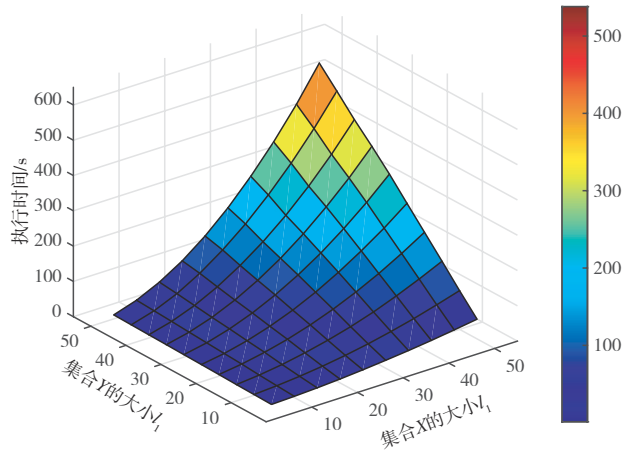

图2 无全集时执行时间随l_1、l_2的变化规律

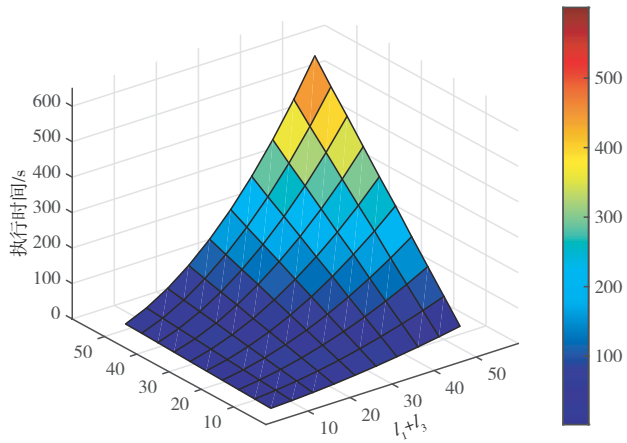

图3 执行时间随(l_1+l_3)、(l_2+l_4)的变化规律

综合考虑安全性和效率之间的平衡，需根据实际问题灵活选择假元素数量.

6 结论

保密计算集合交集对应元素和的最大值问题是安全多方计算中新的研究问题，具有重要的研究意义. 本文基于半诚实模型，在保护集合交集势的条件下，设计了有无全集两种协议，并利用模拟范例证明了协议的安全性. 本文所设计的协议可以实现保密计算集合交集对应元素和的最大值，但本文协议的效率比较低. 下一步我们将进一步研究云计算下高效的交集对应元素和的最大值协议.

参考文献

[1] YAO A C. Protocols for secure computations[C]//23rd Annual Symposium on Foundations of Computer Science (SFCS 1982). Piscataway: IEEE, 2008: 160-164.

[2] GOLDREICH O, MICALI S, WIGDERSON A. How to play ANY mental game[C]//The 19th Annual ACM Con-

ference on Theory of Computing. New York: ACM, 1987: 218-229.

[3] DOU J W, GONG L M, LI S D, et al. Efficient private subset computation[J]. Security and Communication Networks, 2016, 9(18): 5965-5976.

[4] 陈振华, 李顺东, 王道顺, 等. 非加密方法安全计算集合包含关系[J]. 计算机研究与发展, 2017, 54(7): 1549-1556.
CHEN Z H, LI S D, WANG D S, et al. Protocols for secure computation of set-inclusion with the unencrypted method[J]. Journal of Computer Research and Development, 2017, 54(7): 1549-1556. (in Chinese)

[5] SUN H L, WEI L F, WANG L B, et al. A trusted and privacy-preserving carpooling matching scheme in vehicular networks[J]. Journal of Information Security, 2022, 13(1): 1-22.

[6] YANG L, LI C, CHENG Y T, et al. Achieving privacy-preserving sensitive attributes for large universe based on private set intersection[J]. Information Sciences, 2022, 582: 529-546.

[7] ADAVOUDI JOLFAEI A, MALA H, ZAREZADEH M. EO-PSI-CA: Efficient outsourced private set intersection cardinality[J]. Journal of Information Security and Applications, 2022, 65: 102996.

[8] ALI M, MOHAJERI J, SADEGHI M R, et al. Attribute-based fine-grained access control for outscored private set intersection computation[J]. Information Sciences, 2020, 536: 222-243.

[9] YAO L, CHEN Z Y, WANG X, et al. Sensitive label privacy preservation with anatomization for data publishing[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 18(2): 904-917.

[10] SHU X K, YAO D F, BERTINO E. Privacy-preserving detection of sensitive data exposure[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(5): 1092-1103.

[11] RAGHUVIR Y A, GOVINDARAJAN S, VIJAYAKUMAR S, et al. Advancement on security applications of private intersection sum protocol[C]//Proceedings of the Future Technologies Conference (FTC) 2021, Volume 3. Cham: Springer International Publishing, 2021: 104-116.

[12] KOLESNIKOV V, KUMARESAN R, ROSULEK M, et al. Efficient batched oblivious PRF with applications to private set intersection[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2016: 818-829.

[13] YANG X C, YI X, KELAREV A, et al. A distributed networked system for secure publicly verifiable self-tallying online voting[J]. Information Sciences, 2021, 543: 125-142.

[14] ION M, KREUTER B, NERGIZ A E, et al. On deploying secure computing: Private intersection-sum-with-cardinality[C]//2020 IEEE European Symposium on Security and Privacy (EuroS&P). Piscataway: IEEE, 2020: 370-389.

[15] GOLDREICH O. The Fundamental of Cryptography: Basic Applications[M]. London: Cambridge University Press, 2004: 599-764.

[16] 杨晓艺, 李顺东, 亢佳. 保密替换及其在保密科学计算中的应用[J]. 计算机学报, 2018, 41(5): 1132-1142.
YANG X Y, LI S D, KANG J. Private substitution and its applications in private scientific computation[J]. Chinese Journal of Computers, 2018, 41(5): 1132-1142. (in Chinese)

[17] 李顺东, 张萌雨. 盲百万富翁问题的高效解决方案[J]. 计算机学报, 2020, 43(9): 1755-1768.
LI S D, ZHANG M Y. An efficient solution to the blind millionaires' problem[J]. Chinese Journal of Computers, 2020, 43(9): 1755-1768. (in Chinese)

作者简介

马秀莲　女,1997年9月出生于青海省西宁市. 现为陕西师范大学计算机科学学院硕士生. 主要研究方向为密码学与信息安全.
E-mail: xiulian@snnu.edu.cn

张倦倦　女,2000年2月出生于陕西省榆林市. 现为陕西师范大学计算机科学学院硕士生. 主要研究方向为密码学与信息安全.
E-mail: zjuanjuan@snnu.edu.cn

李顺东(通讯作者)　男,1963年12月出生于河南省平顶山市. 现为陕西师范大学计算机科学学院博士生导师. 主要研究方向为密码学与信息安全.
E-mail: shundong@snnu.edu.cn