集合交集元素和的保密计算

李顺东,赵雪玲,家珠亮
(陕西师范大学计算机科学学院,陕西西安 710119)

摘 要: 安全多方计算作为密码学的重要分支,长期以来主要致力于解决两方或多方参与者隐私数据的联合计算. 集合交集元素和的隐私计算作为安全多方计算中的科学计算问题,在保密计算广告转化率中具有重要作用. 我们利用保密替换和加密选择求集合的交集,结合 Lifted ElGamal 加密算法,研究了不同限制下(数据范围较小和数据范围较大)集合交集元素和多方保密计算. 本文方案解决两方计算时,Bob 只需从 Alice 发送的数据中选择数据,避免了复杂的模指数运算,且双方不需多次交互,降低了计算成本和通信次数. 多方参与计算时,根据加密选择和保密替换的性质,得到集合交集的密文,然后在密文上计算得到集合交集元素的和. 通过理论分析和实验证明,本文协议是高效的. 最后利用模拟范例证明本文协议是安全的.

关键词: 安全多方计算;集合交集元素的和;概率加密;加密选择;保密替换
基金项目: 国家自然科学基金(No.61272435)
中图分类号: TP309.2 **文献标识码:** A **文章编号:** 0372-2112(2023)01-0086-07
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20211444

Private Intersection-Sum Computation

LI Shun-dong, ZHAO Xue-ling, JIA Zhu-liang
(*School of Computer Science*, *Shaanxi Normal University*, *Xi'an*, *Shaanxi* 710119, *China*)

Abstract: As an important branch of cryptography, secure multi-party computation has long been mainly dedicated to solving the joint computation on private data owned by two or more parties. As a scientific computation problem in secure multi-party computation, secure intersection-sum computation plays an important role in privately computing advertising conversion rate. We use private substitution, encryption selection and lifted ElGamal cryptosystem to design secure intersection-sum protocols for different scenarios, that is, small data range and large data range for multi-party. When the proposed scheme is used to solve the secure intersection-sum for two-party, Bob only needs to choose data from the data sent by Alice without performing expensive modular exponentiations, and the two parties do not need to interact multiple times. The computational cost and communication times are reduced. In the multi-party protocols, all parties use encryption selection and private substitution to obtain some ciphertexts and perform computation on the ciphertexts to obtain the ciphertext of the intersection-sum. Theoretical analysis and experimental results show that our protocol is efficient. Finally, we use the simulation paradigm to strictly prove that our protocol is secure.

Key words: secure multi-party comptation; intersection-sum; probabilistic encryption; encryption selection; private substitution
Foundation Item(s): National Natural Science Foundation of China (No.61272435)

1 引言

网络迅速发展,为人们生活带来很大便利,但用户在注册信息或搜索信息时便会泄露自己的私有数据. 如何在保护用户私有数据前提下,得到想要结果,这是信息时代亟需解决的问题.

姚期智教授[1]于 1982 年提出百万富翁问题,开创了保密计算的先河,随后 Ben-Or 和 Goldwasser[2]继续推进,提出安全多方计算问题,即多个参与者的联合保密计算问题,Goldreich 等[3]设计了安全多方计算的通用方案. 长期以来,安全多方计算主要致力于两方或多方参与者数据的联合计算,让数据安全碰撞出更大价值,打破因私有数据独立存在,无法有效利用的壁垒. 保证计

收稿日期:2021-10-21;修回日期:2022-09-14;责任编辑:李勇锋

算结束时每个参与者私有数据都未泄露.

集合交集的保密计算在基因序列测试[4]、相似文档检索[5]、私有好友推荐[6]、指纹匹配[7]和衡量广告转化率[8]等方面具有重要作用. 广告转化率是衡量线上广告对产品销售影响的重要指标,用户在网上看到某公司在线广告,然后在该公司购买产品,此时便会形成广告转化率. 计算广告转化率的数据分别由两个公司掌握,广告公司有点击广告的用户名单,产品公司有购买产品的用户的名单. 为保护用户隐私,两公司都不愿泄露自己用户名单,却都想知道点击广告页面且购买产品的用户名单,此场景便是集合交集问题的保密计算. 但若两公司希望保密得到因点击广告而产生的消费总额,此问题是集合交集问题的扩展,但因为消费总额是一个聚合统计问题,除了集合元素本身,集合交集元素也不能泄露. 此时集合交集计算方案都不能直接应用于该问题的求解,在此场景下保密计算集合交集元素的和成为衡量广告转化率的新方法.

文献[8]基于Diffie-Hellman问题和Paillier加密系统解决了一方参与者拥有用户对应的标识符,另一方参与者拥有与标识符对应的整数值,计算的结果为标识符的交集,与交集标识符对应整数值的和. 文献[9]分别基于判定性Diffie-Hellman问题、不经意伪随机函数和Bloom过滤器设计了三个计算集合交集元素和的安全协议. 文献[10]在文献[9]的基础上,设计出恶意模型下集合交集元素和的保密计算协议. 但上述方案均泄露了集合交集的势,且由于文献[8~10]协议需两方参与者多次交互,因此很难扩展到多方参与计算时的情况.

针对以上问题,本文设计了有全集和无全集两种情况下集合交集元素和的保密计算协议. 其方法很容易解决多方参与计算的情况,因此给出了集合交集元素和的安全多方计算协议,两方是多方参与计算的特例. 我们的主要贡献为

(1) 利用加密选择和保密替换得到集合交集的密文,在集合交集密文上计算集合交集元素的和;

(2) 利用随机数混淆实际数据,将参与者元素对应为一个元组,为无全集限制的安全多方计算提供了新方案;

(3) 对已有集合交集元素和的协议进行改进,通过分析可知,我们的协议效率更高,扩展性更强;

(4) 将数据转化计算,为使用Lifted ElGamal抵抗合谋攻击加密较大数,无需计算较多离散对数提供了新方式.

2 预备知识

2.1 半诚实模型及其安全性

半诚实参与者[3]能严格遵守协议规定,向其他参与者发送正确信息,但会记录中间结果,在协议结束后试图推导其他参与者私有数据. 若所有参与者都是半诚实的,则将该模型称为半诚实模型. 半诚实模型中常用的安全性证明方法是模拟范例[11].

设有n个参与者$P_1,\cdots,P_n$分别输入$x_1,\cdots,x_n$,令$X=x_1,\cdots,x_n$,在协议中将$P_i(i=1,\cdots,n)$获得的序列记为:$\text{view}_i^{\pi}(X)=(x_i,r_i,M_i^1,\cdots,M_i^t)$,其中,$r_i$表示$P_i$抛硬币得到的结果,$M_i^j$表示$P_i$第$j$次获得的信息.

对部分参与者$I=\{P_{i_1},\cdots,P_{i_s}\}\subsetneq\{P_i,\cdots,P_n\}$,将获得序列记为:$\text{view}_I^{\pi}(X)=(I,\text{view}_{i_1}^{\pi}(X),\cdots,\text{view}_{i_s}^{\pi}(X))$.

定义1 参与者均为半诚实时,若存在模拟器S,对于任意$I=\{P_{i_1},\cdots,P_{i_s}\}\subsetneq\{P_i,\cdots,P_n\}$有:

$$\{S(I,(x_{i_1},\cdots,x_{i_s}),f_I(X))\}_X\overset{c}{\equiv}\{\text{view}_I^{\pi}(X)\}_X \tag{1}$$

则表示协议π保密计算函数f. 其中$\overset{c}{\equiv}$表示计算不可区分,构造满足式(1)模拟器证明协议安全性.

2.2 门限密码体制

门限密码体制[12]是安全多方计算中对抗合谋攻击的重要工具. 在门限密码体制中,n个参与者联合生成公钥,解密密钥由n个参与者联合持有. 公钥可以直接加密消息,但解密需要n个参与者合作才能正确解密.

2.3 Lifted ElGamal门限加密系统[13]

密钥生成:给定安全参数k,生成一个大素数p,以及Z_p^*的一个生成元g. n个参与者商定一个小素数ρ,$P_i(i=1,\cdots,n)$各自选取$k\in Z_p^*$作为自己的私钥,计算$K_i=g^{k_i}\bmod p$. 并联合生成公钥

$$h=\prod_{i=1}^{n}K_i\bmod p=\prod_{i=1}^{n}g^{k_i}\bmod p=g^{\sum_{i=1}^{n}k_i}\bmod p$$

加密:对于明文$m\in Z_p^*$,选择随机数r,计算

$$C=(C_1,C_2)=(g^r\bmod p,\rho^m h^r\bmod p)$$

联合解密:$P_i(i=1,\cdots,n)$计算$t_i=C_1^{k_i}\bmod p$,并将结果公布. n个参与者进一步计算

$$t=\prod_{i=1}^{n}t_i\bmod p=\prod_{i=1}^{n}C_1^{k_i}\bmod p=C_1^{\sum_{i=1}^{n}k_i}\bmod p$$

进一步计算$\rho^m=C_2t^{-1}\bmod p$,最后通过查询提前构造的对数表得到结果m.(注:Lifted ElGamal门限密码系统解密为ρ^m,需进一步计算离散对数,为简化计算,一般选择$\rho=2$. 同时,由于离散对数计算的困难性,当明文空间较小时提前构造离散对数表,解密后通过查询离散对数表获得计算结果.)

同态性质: Lifted ElGamal加密系统具有加法同态性. 对于明文消息$m_1,m_2\in Z_p^*$,有

$$E(m_1)E(m_2)=E(m_1+m_2)$$

2.4 密文的重随机化

密文的重随机化指参与者在不解密的情况下将消

息 m 的密文转化为另一个密文，相当于对明文 m 的重新加密. 本文通过重随机化使每个参与者都无法了解其余参与者选择了哪些数据，替换了哪些数据，进而实现保密替换. Lifted ElGamal 利用其加法同态性进行重随机化，只需计算 $E(m)\cdot E(0)$.

3 有全集限制的解决方案

3.1 有全集限制的多方保密计算

两方保密计算是多方参与计算的特例，我们先设计集合交集元素和的多方保密计算协议.

问题描述：设参与者 $P_i(i=1,\cdots,n)$ 分别拥有集合 $X_i=\{x_{i1},\cdots,x_{il_i}\}\subseteq Q$. 其中，$Q=\{q_1,\cdots,q_l\}$ 且满足 $q_1<\cdots<q_l$. n 个参与者想要计算他们集合交集元素的和，但不想泄露交集以及参与者的私有数据.

计算原理 1 P_1 将自己数据对应为 l 维数组. 后续参与者加入保密替换操作，结合加密选择求集合交集，进而求集合交集元素的和. 具体步骤如下：

（1）P_1 构造数组 $M_1=\{m_{11},\cdots,m_{1l}\}$，将 M_1 发送给 P_2 继续执行. 其中，

$$m_{1j}=\begin{cases}x_{1j}, & q_j\in X_1(j=1,\cdots,l)\\ 0, & q_j\notin X_1(j=1,\cdots,l)\end{cases}$$

（2）$P_i(i=2,\cdots,n-1)$ 通过修改 M_{i-1} 构造 M_i，然后将 M_i 发送给 P_{i+1} 继续执行. 修改 M_{i-1} 的原则为

$$m_{ij}=\begin{cases}m_{i-1,j}, & q_j\in X_i(j=1,\cdots,l)\\ 0, & q_j\notin X_i(j=1,\cdots,l)\end{cases}$$

（3）参与者 P_n 根据 X_n 在 M_{n-1} 中选择数据，计算 $\sum_{j=1}^{l_n}m_{n-1,j}$. 式中 $m_{n-1,j}$ 对应的 $q_j\in X_n(j=1,\cdots,l)$. 我们将计算结果 $\sum_{j=1}^{l_n}m_{n-1,j}$ 记为 $\mathrm{Sum}(X_1\cap\cdots\cap X_n)$. 则 $\mathrm{Sum}(X_1\cap\cdots\cap X_n)$ 即为集合交集元素的和.

上述计算过程如果在明文下实现，无法对参与者数据进行保密. 为抵抗合谋攻击，我们基于 Lifted ElGamal 门限加密系统设计出安全协议 1.

正确性：P_2 先加密选择，若 q_j 为 P_1 和 P_2 共有元素，则 P_2 选择的 $m_{1j}=x_{1j}=x_{2j}$；若 q_j 仅为 P_1 的元素，P_2 不会选择该元素；若 q_j 仅为 P_2 的元素，则 P_2 选择的 $m_{1j}=0$. 那么 P_1 和 P_2 集合的交集必然包含于 P_2 选出的数据中. 然后 P_2 保密替换，当 $q_j\notin X_2$ 时，使用 0 替换 m_{1j}，P_2 将属于 P_1 不属于自己元素对应的编码全部替换为 0，经过加密选择和保密替换后，P_2 得到的 M_2 中不为 0 编码对应的元素即为 P_1 和 P_2 集合交集. 依次类推，P_n 选择的数据或为 n 个参与者集合的交集或为 0. 所以 P_n 计算 $\prod_{j=1}^{l_n}E(m_{n-1,j})$，由加密系统的加法同态性可知：

协议 1 有全集限制的多方保密计算

输入： $P_i(i=1,\cdots,n)$ 分别输入 $X_i=\{x_{i1},\cdots,x_{il_i}\}\subseteq Q$

输出： $\mathrm{Sum}(X_1\cap\cdots\cap X_n)$

准备：n 个参与者利用 Lifted ElGamal 加密系统，得到私钥 k_i 和公钥 h，并提前根据全集生成离散对数表.

(1) P_1 加密 M_1 得到 $E(M_1)=\{E(m_{11}),\cdots,E(m_{1l})\}$，并将 $E(M_1)$ 发送给 P_2.

(2) $P_i(i=2,\cdots,n-1)$ 修改 $E(M_{i-1})$ 中数据构造 M_i. 当 $q_j\in X_i$ 时，保持 $E(m_{i-1,j})$ 数据不变，仅对其重随机化；当 $q_j\notin X_i$ 时，将 $E(m_{i-1,j})$ 采用 $E(0)$ 替换. 修改后将 $E(M_i)=\{E(m_{i1}),\cdots,E(m_{il})\}$ 发送给 P_{i+1}.

(3) P_n 根据 X_n 在 $E(M_{n-1})$ 中选择数据. 并计算

$$E(\mathrm{Sum}(X_1\cap\cdots\cap X_n))=\prod_{j=1}^{l_n}E(m_{n-1,j})$$

式中，$E(m_{n-1,j})$ 对应的 $q_j\in X_n$，P_n 将计算结果公布.

(4) n 个参与者联合解密，并根据离散对数表得到结果 $\mathrm{Sum}(X_1\cap\cdots\cap X_n)$.

$$E(\mathrm{Sum}(X_1\cap\cdots\cap X_n))=\prod_{j=1}^{l_n}E(m_{n-1,j})=E(\sum_{j=1}^{l_n}m_{n-1,j})$$

安全性：$P_i(i=2,\cdots,n-1)$ 重随机化加密选择的数据 $E(m_{i-1,j})$，相当于对选择数据对应明文重新加密. 保密替换时使用 $E(0)$ 替换 $E(m_{i-1,j})$. 因此 P_{i+1} 收到一个重新加密的数组，此时 P_{i+1} 无法判断 P_i 对选择了哪些数据，替换了哪些数据.

定理 1 协议 2 在半诚实模型下是安全的，能抵抗任意 $n-1$ 个参与者的合谋.

证明 协议中每位参与者的作用是相同的，且 $n-1$ 个参与者合谋是协议 2 最严重的合谋攻击，因此我们仅证明 P_1 对其余 $n-1$ 个参与者 $I=\{P_2,\cdots,P_n\}$ 构成的合谋是安全的.

在实际协议过程中，将 I 使用的随机数记为 r_I，I 作为一个整体得到的信息只有 P_1 发送的 $E(M_1)$ 和计算的 $\mathrm{Sum}(X_1\cap\cdots\cap X_n)$，因此合谋者 I 的 view 为

$$\mathrm{view}_I^{\pi}(X)=\{(X),E(M_1),r_I,\mathrm{Sum}(X_1\cap\cdots\cap X_n)\}$$

（1）对于输入 $(I,X_I,\mathrm{Sum}(X_1\cap\cdots\cap X_n))=\{I,(X_2,\cdots,X_n),\mathrm{Sum}(X_1\cap\cdots\cap X_n)\}$，$S$ 随机选择集合 $X_1'=\{x_{11}',\cdots,x_{1l_1}'\}\subseteq Q$ 使得

$$\mathrm{Sum}(X_1'\cap X_2\cap\cdots\cap X_n)=\mathrm{Sum}(X_1\cap\cdots\cap X_n).$$

（2）S 将 X_1' 编码为 M_1'，并加密为 $E(M_1')$.

（3）S 根据合谋者 I 的集合在 $E(M_1')$ 中选择数据，得到 $E(M_I^*)=\{E(M_{I1}^*),\cdots,E(M_{Il_I}^*)\}$，并计算

$$\mathrm{Sum}(X_1'\cap X_2\cap\cdots\cap X_n)=\prod_{j=1}^{l_I}E(m_{Ij})$$

在模拟器模拟协议中令：

$$S(I,X_I,\mathrm{Sum}(X_1\cap\cdots\cap X_n))=\{(I,(X_2,\cdots,X_n),E(M_1'),r_I,\mathrm{Sum}(X_1'\cap X_2\cap\cdots\cap X_n)\}$$

Lifted ElGamal 加密系统是语义安全的，所以 $E(M_1')$ 和

$E(M_1)$是不可区分的,因此,

$$S(I,X_I,\mathrm{Sum}(X_1\cap\cdots\cap X_n))_{X\in Q}\overset{c}{\equiv}\{\mathrm{view}_I^{\pi}(X_1,\cdots,X_n)\}_{X\in Q}$$

综上所述,协议2是安全的,能抵任意合谋.

3.2 有全集限制的两方保密计算

两方计算是多方计算的特例,在多方计算中若只有第一和最后一个参与者,协议就成为两方保密计算方案,我们只给出两方保密计算协议2.

协议2　有全集限制的两方保密计算

输入:Alice输入私有集合$X_1=\{x_{11},\cdots,x_{1l_1}\}\subseteq Q$, Bob输入私有集合$X_2=\{x_{21},\cdots,x_{2l_2}\}\subseteq Q$.

输出:$\mathrm{Sum}(X_1\cap X_2)$.

(1) Alice运行加密算法,生成公私钥.根据计算原理1中P_1的编码方式构造数组M_1,对M_1加密得到$E(M_1)$,将$E(M_1)$发送给Bob.

(2) Bob根据X_2在$E(M_1)$中选择数据.计算

$$E(\mathrm{Sum}(X_1\cap X_2))=\prod_{j=1}^{l_2}E(m_{1j})$$

将结果发送给Alice.式中,m_{1j}对应的$q_j\in X_2$.

(3) Alice使用私钥解密并公布结果$\mathrm{Sum}(X_1\cap X_2)$.

协议2的正确性和安全性由协议1保证.

4 无全集限制的解决方案

4.1 无全集限制的多方保密计算

第3节中,编码方式是连续的,此方法仅适合数据范围较小的情况. 因此,我们设计另一种编码方式,利用私有数据和编码相对应的方式,将参与者的每个数据编码为元组进行计算.

问题描述:假设n个参与者$P_i(i=1,\cdots,n)$,分别拥有集合$X_i=\{x_{i1},\cdots,x_{il_i}\}$. n个参与者想要保密计算他们交集元素的和,但不能泄露交集元素和参与者的具体数据.

计算原理2　本节主要思想是利用虚假元素隐藏参与者实际数据,将参与者每个数据对应为元组,结合保密替换和加密选择进行计算.

(1) P_1先给X_1中添加r_1个随机数,将X_1对应为l_1+r_1维数组M_1,M_1中元素$m_{1j}=(m_{1j}^1,m_{1j}^2)$. 其中,$m_{1j}^2$为$P_1$集合元素或添加的随机数,$m_{1j}^1$表示$m_{1j}^2$是否为$P_1$集合元素,若$m_{1j}^2$为$P_1$集合元素,则$m_{1j}^1=1$,若$m_{1j}^2$不是$P_1$集合元素,则$m_{1j}^1=1$. 即

$$m_{1j}=\begin{cases}(1,x_{1j}), & m_{1j}^2\in X_1\\(0,x_{1j}), & m_{1j}^2\notin X_1\end{cases}$$

P_1将M_1中数据置换后发送给P_2.

(2) $P_i(i=2,\cdots,n-1)$根据X_i修改数组M_{i-1}构造M_i. 当$m_{i-1,j}^2$为P_i集合的元素时,即$m_{i-1,j}^2\in X_i$,保持$m_{i-1,j}$数据不变,当$m_{i-1,j}^2$不是P_i集合的元素时,即$m_{i-1,j}^2\notin X_i$时,将$m_{i-1,j}$中$m_{i-1,j}^1$设置为0. 即

$$m_{ij}=\begin{cases}(m_{i-1,j}^1,m_{i-1,j}^2), & m_{i-1,j}^2\in X_i\\(0,m_{i-1,j}^2), & m_{i-1,j}^2\notin X_i\end{cases}$$

P_i将M_i随机化后发送给P_{i+1}.

(3) P_n根据X_n在M_{n-1}中选择对应元组$m_{n-1,j}$,将结果记为$M_n=(m_{n1},\cdots,m_{nl_n})$. 对于$j\in[1,l_n]$,$m_{nj}=(m_{nj}^1,m_{nj}^2)$,且选择的数据中$m_{nj}^2\in X_n$.

(4) 对$j\in[1,l_n]$,P_n将m_{nj}^2表示$m_{nj}^{2d'}\cdots m_{nj}^{20'}$,并根据$m_{nj}^2$的二进制和$m_{nj}^1$计算$m_{nj}^1\times m_{nj}^2$的二进制. 步骤为:对$k'\in[0',d']$,当$m_{nj}^{2k'}=1$时,保持$m_{nj}^1$不变,此时$m_{nj}^1\times m_{nj}^{2k'}=m_{nj}^1$,当$m_{nj}^{2k'}=0$时,令$m_{nj}^1\times m_{nj}^{2k'}=0$.

(5) 对$j\in[1,l_n]$,P_n将$m_{nj}^1\times m_{nj}^2$的结果表示为$m_{nj}^{d'}\cdots m_{nj}^{0'}$,然后将所有二进制对应位相加,并记为$m^k$. 即对$k'\in[0',d']$计算:$m^k=\sum_{j=1}^{l_n}m_{nj}^{k'}$.

P_n计算完所有m^k,记为$M=(m^d,\cdots,m^0)$,然后计算$m^d\times 2^d+\cdots+m^0\times 2^0$,将计算结果记为$\mathrm{sum}(X_1\cap\cdots\cap X_n)$. 即为集合交集元素的和.

命题1　存在n个十进制$a_1,\cdots,a_n$,对应二进制分别为$a_1^d\cdots a_1^0,\cdots,a_n^d\cdots a_n^0$. 计算$m^d=a_1^d+\cdots+a_n^d,\cdots,m^0=a_1^0+\cdots+a_n^0$,则有

$$a_1+\cdots+a_n=m^d\times 2^d+\cdots+m^0\times 2^0.$$

证明　根据二进制和十进制的转化我们有$a_1=a_1^d\times 2^d+\cdots+a_1^0\times 2^0,\cdots,a_n=a_n^d\times 2^d+\cdots+a_n^0\times 2^0$,将十进制直接相加,等于对应的二进制相加.

$$\begin{aligned}a_1+\cdots+a_n&=(a_1^d\times 2^d+\cdots+a_1^0\times 2^0)+\cdots+(a_n^d\times 2^d+\cdots+a_n^0\times 2^0)\\&=(a_1^d+\cdots+a_n^d)\cdot 2^d+\cdots+(a_1^0\cdots+a_n^0)\cdot 2^0\\&=m^d\cdot 2^d+\cdots+m^0\cdot 2^0\end{aligned}$$

转化后的$m^d,\cdots,m^0\in[0,n]$,便于使用Lifted ElGamal加密系统保密计算$m^d,\cdots,m^0$,进而计算$a_1+\cdots+a_n$.

在计算原理基础上,使用Lifted ElGamal加密系统,设计出无全集限制下的保密计算协议3.

正确性:协议3是协议1思想的拓展. 都是利用保密替换和加密选择求集合交集. 但协议3编码与协议1稍有不同,P_1编码时,首先给集合X_1添加任意个随机数,利用添加虚假元素混淆拥有的实际数据,然后将每一个数据$x_{1j}(j=1,\cdots,l_1)$和随机数对应为元组$m_{1j}=(m_{1j}^1,m_{1j}^2)$构成数组$M_1$.

根据计算原理可知,和协议1一样n个参与者集合的交集必包含于P_n选择的数据中,为选择数据$E(m_{i-1,j}^1)=E(1)$对应的$m_{i-1,j}^2$. 若采用十进制,根据密码系统的加法同态性有:

$$E(\mathrm{Sum}(X_1\cap\cdots\cap X_n))=\prod_{j=1}^{l_n}E(m_{nj}^1)^{m_{nj}^2}=E(\sum_{j=1}^{l_n}(m_{nj}^1\times m_{nj}^2)) \quad (2)$$

在式(2)中,若$E(m_{nj}^1)=E(0)$,则m_{nj}^2必为第一个参与

协议3 无全集限制的多方保密计算

输入:$P_i(i=1,\cdots,n)$输入集合$X_i=\{x_{i1},\cdots,x_{il_i}\}$.

输出:Sum$(X_1\cap\cdots\cap X_n)$

准备: 参与者利用Lifted ElGamal门限加密系统, 得到私钥k_i公钥h, 同时P_n计算$2^1,\cdots,2^{l'_n}$生成离散对数表.

(1) P_1根据计算原理2构造数组M_1,加密M_1后得到密文数组$E(M_1)=(E(m_{11}),\cdots,E(m_{1l_1+r_1}))$,并发送给$P_2$.其中,密文$E(m_{1j})$表示为$(E(m^1_{1j}),m^2_{1j})$. 当$m^2_{1j}\in X_1$时,$E(m_{1j})=(E(1),m^2_{1j})$;当$m^2_{1j}\notin X_1$时,$E(m_{1j})=(E(0),m^2_{1j})$.

(2) $P_i(i=2,\cdots,n-1)$修改$E(M_{i-1})$. 当$m^2_{i-1,j}\in X_i$时, 选择$E(m_{i-1,j})$后,只对其重随机化;当$m^2_{i-1,j}\notin X_i$时, P_i将$E(m^1_{i-1,j})$使用$E(0)$替换. P_i将修改的$E(M_i)$发送给P_{i+1}.

(3) P_n根据集合X_n在$E(M_{n-1})$中选择对应元组. 将选择的全部数据记为$E(M_n)=(E(m_{n1}),\cdots,E(m_{nl_n}))$. 其中, 对于

$j\in[1,l_n]$, $E(m_{nj})=(E(m^1_{nj}),m^2_{nj})$且$m^2_{nj}\in X_n$.

(4) 对$j\in[1,l_n]$,P_n根据计算原理2中的步骤(4)计算$E(m^1_{nj}\times m^2_{nj})$的二进制表示.

(5) 对$j\in[1,l_n]$,P_n得到$E(m^1_{nj}\times m^2_{nj})$的二进制,将其表示为$E(m^{d'}_{nj})\cdots E(m^{0'}_{nj})$,然后对于$j\in[1,l_n]$,$k'\in[0',d']$,计算:$E(m^k)=\prod_{j=1}^{l_n}E(m^{k'}_{nj})$.

P_n计算$E(m^k)$得到d个密文,记为

$E(M)=(E(m^d),\cdots,E(m^0))$,然后将$E(M)$公布.

(6) 个参与者联合解密, 得到$M=(m^d,\cdots,m^0)$, 然后计算

Sum$(X_1\cap\cdots\cap X_n)=m^d\times 2^d+\cdots+m^0\times 2^0$

者添加的随机数或$P_i(i=2,\cdots,n-1)$保密替换了不属于自己的元素,此时计算$0\times m^2_{nj}=0$,最后只有P_n选择的数据$E(m_{n-1,j})$中$E(m^1_{n-1,j})=E(1)$的数据参与计算,而$E(m_{n-1,j})$中$E(m^1_{n-1,j})=E(1)$对应的数据$m^2_{n-1,j}$是n个参与者集合交集的元素. 因此,协议4正确计算了集合交集元素的和.

协议3数据都是稀疏集且数据范围较大,需要计算的离散对数非常多. 若将十进制转化为二进制,需要构造的离散对数表长度仅和需要相加数据的个数有关,对于数据本身范围很大的数据可以减小计算复杂度. 因此,在协议3中,将m^2_{nj}转化为二进制后,P_n先计算$E(m^1_{nj}\times m^2_{nj})$的二进制表示,然后利用命题1和密码系统的加法同态性求集合交集元素的和.

安全性: 协议3在半诚实模型下是安全的,能抵抗任意合谋. 其证明与协议1相同,故省略.

无全集限制的两方计算协议和有全集限制一样,是多方参与计算的特例,协议3只有第一个和最后一个参与者,就是两方保密计算协议.

5 协议效率和实验分析

在分析中,有全集限制时设全集有l个元素,无全集限制时,P_i有l_i个元素,加入随机数为r_i个,在文献[9]中,设Alice有l_1个元素以及l_1个关联整数值,Bob有l_2个元素. 在Lifted ElGamal门限加密系统中,联合生成公钥需n次模指数运算,加密一个较小数需2次模指数运算,联合解密一个元素需要n次模指数运算. 在Paillier加密系统中,加密一个数需2次模指数运算,解密一个消息需2次模指数运算. 若将随机数r选做$1+kN$的形式,解密和加密都只需1次模指数运算.

5.1 协议效率

计算复杂性:以模指数运算衡量计算复杂性.

协议1使用了Lifted ElGama门限加密算法,n方参与计算共需$2(n+nl-l)$次模指数运算.

为和文献[9]对比,在协议2中,使用Paillier加密算法. 共需$l+1$次模指数运算. 文献[9]利用Diffie-Hellman判断数据相等,进而求集合交集,集合势都为l时,需$3l+1$次模指数运算.

协议3利用Lifted ElGamal门限加密系统抵抗参与者的合谋,计算复杂性和P_1添加的随机数有关. P_n将数据转化为二进制计算时,只需要根据m^2_{nj}的二进制选择或替换数据,不需要模指数运算,但n个参与者需解密d个数据,所以协议3共需$2(n-1)(l_1+r_1)+n(1+d)$次模指数运算.

通信复杂性:以协议执行过程中交互的次数以及传输密文的数量衡量通信复杂性.

协议1中,通信次数为$3n$次. 总通信密文数量为$(n-1)l+1$.

协议2中,需3次通信. 通信密文数量为$l+1$. 文献[9]双方交互计算,通信次数为5次,全集势均为l时,通信密文量为$4l+1$.

协议3通信次数和协议1相同,需$3n$次通信, 通信密文数量为$(n-1)(l_1+r_1)+1$.

文献[8]和文献[9]方法类似,只是在处理数据时文献[9]使用的是哈希函数,但是两篇文献的复杂性相同,文献[10]是在文献[9]基础上对恶意模型下集合交集元素和的求解,且在文献[9]中作者进行了对比证明使用DDH是计算集合交集和的最优方案. 所以表1仅将本文协议和文献[9]使用DDH方案的计算复杂性和通信复杂性详细列出. 通过表1可知,本文协议的通信复杂性和计算复杂性都有明显的改善.

表1 本文协议与文献[9]协议效率比较

	计算复杂性	通信次数	通信量(比特)	参与人数
协议1	$2(n+nl-l)$	$3n$	$(n-1)l+1$	n
协议2	$l+1$	3	$l+1$	2
文献[9]	$3l+1$	5	$4l+1$	2
协议3	$2(n-1)(l_1+r_1)+n(1+d)$	$3n$	$(n-1)(l_1+r_1)+1$	n

5.2 实验测试

我们使用Java语言对协议的执行时间进行测试,在实验中,加密算法选择的模数p为1 024比特,系统为windows10的64位操作系统,处理器为Intel(R) Core(TM) i5-6600CPU@ 3.30 GHz 3.31 GHz. jdk版本为1.8.0,程序运行平台为Eclipse.

在协议1中,多方参与计算时,执行时间和全集的势以及参与者的数量有关. 图1表示了执行时间随参与者人数和全集势变化的二维图. 通过分析知执行时间分别随着参与者人数和全集势的增加而线性增长.

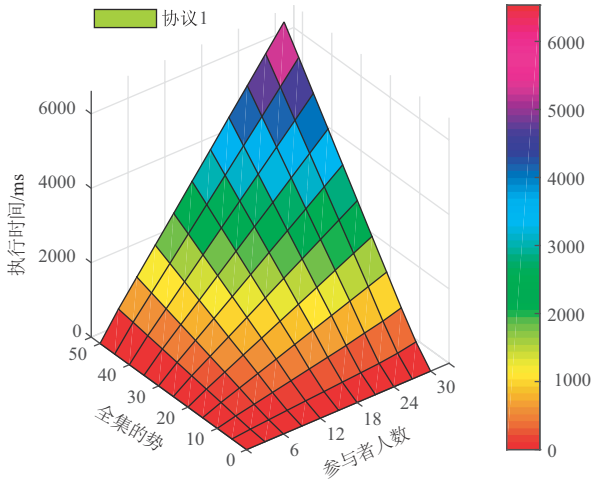

图1　执行时间随参与者人数和全集势变化规律

在协议2和文献[9]中设Bob集合的势和Alice集合的势相同,每次Bob将Alice发送的数据全部选择,然后进行乘法运算,本文和文献[9]协议的执行时间随全集势的增加而线性增长,变化规律如图2所示.

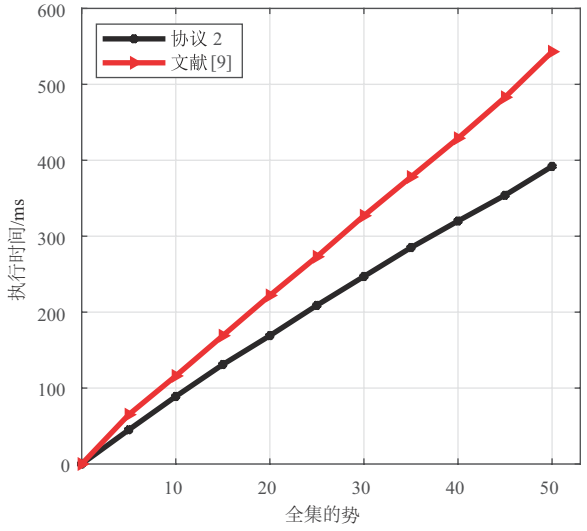

图2　执行时间随全集的势变化规律

协议3执行时间变化规律和协议1类似,因为随机数对实际数据的混淆,其执行时间还受添加随机数个数的影响,增加的时间随添加随机数数量的变化规律如图3所示.

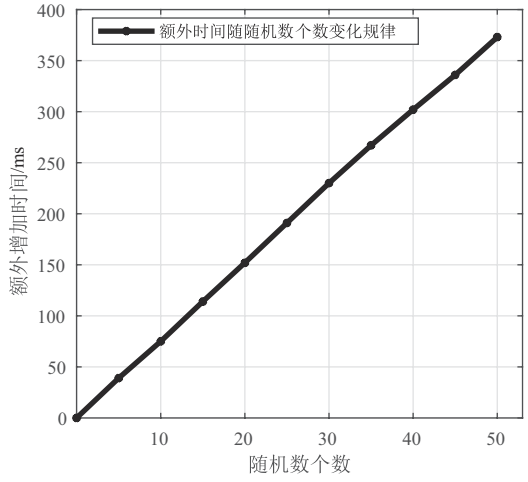

图3　增加时间随随机数数量变化规律

6 总结

随着大数据时代的到来,海量数据交叉计算成为生活中的重要内容,随之带来的隐私泄露也关系千万人的利益. 安全多方计算作为隐私保护的重要技术,其发展也日趋成熟. 由于通用解决方案的计算复杂性和通信复杂性都很高,因此对特定问题的保密计算成为研究的主要问题. 集合的各种运算是保密科学计算中的重要内容,近几年对集合的研究成果很多. 本文主要设计了集合交集元素和的安全计算协议,并通过理论分析和实验分析与文献[9]进行了对比. 通过理论分析和实验证明,本文协议的计算复杂性和通信复杂性都比文献[9]低,同时还保护了集合交集的势. 虽然本文对文献[9]进行了改进,但该问题仍需进一步研究. 本文主要是参与者交互计算,如何将计算协议使用在云环境下,减少用户计算量,这将是我们今后的研究问题. 同时,设计恶意模型下的安全计算协议也会是我们下一步的研究方向.

参考文献

[1] YAO A C. Protocols for secure computations[C]//The 23rd IEEE Annual Symposium on Foundations of Computer Science. Chicago: IEEE Computer Society, 1982: 160-164.

[2] BEN-OR M, GOLDWASSER S, WIGDERSON A. Completeness theorems for non-cryptographic fault-tolerant distributed computation[C]//The 20th Annual ACM Symposium on Theory of Computing. Chicago: ACM, 1988: 1-10.

[3] GOLDREICH O. The Fundamental of Cryptography- Volume: Basic Applications[M]. London: Cambridge Univer-

sity Press, 2004.
[4] BALDI P, BARONIO R, CRISTOFARO E D, et al. Countering gattaca: Efficient and secure testing of fully-sequenced human genomes[C]//The 18th ACM Conference on Computer And Communications Security. New York: ACM, 2011: 691-702.
[5] BLUNDO C, CRISTOFARO E D, GASTI P. EsPRESSO: Efficient privacy-preserving evaluation of sample set similarity[J]. Journal of Computer Security, 2014, 22(3): 355-381.
[6] YAN H, CHAPMAN P, EVANS D. Privacy-preserving applications on smartphones[C]//The 6th USENIX Workshop on Hot Topics in Security. San Francisco: USENIX Association, 2011.
[7] ZHANG E, CHANG J, LI Y. Efficient threshold private set intersection[J]. IEEE Access, 2021, 9: 6560-6570.
[8] ION M, KREUTER B, NERGIZ A E, et al. Private intersection-sum protocol with applications to attributing aggregate ad conversions[J]. IACR Cryptology ePrint Archive, 2017: 738.
[9] ION M, KREUTER B, NERGIZ A E, et al. On deploying secure computing: Private intersection-sum with cardinality [C]//IEEE European Symposium on Security and Privacy. Genova: IEEE, 2020: 370-389.
[10] MIAO P, PATEL S, RAYKOVARM, et al. Two-sided malicious security for private intersection-sum with cardinality[C]//40th Annual International Cryptology Conference. Santa Barbara, Springer, 2020: 3-33
[11] REIMER B, FRIED R, MEHLER B, et al. Brief report: Examining driving behavior in young adults with high functioning autism spectrum disorders: A pilot study using a driving simulation paradigm[J]. Journal of Autism & Developmental Disorders, 2013, 43(9): 2211-2217.
[12] DESMEDT Y, FRANKEL Y. Threshold cryptosystems [C]//The 9th Annual International Cryptology Conference. New York: Springer, 1989: 307-315.
[13] ELGAMAL T. A public key cryptosystem and a signature scheme based on discrete logarithms[J]. IEEE Transactions on Information Theory, 1985, 31(4): 469-47.

作者简介

李顺东 男,1963年生,河南平顶山人. 教授、博士生导师. 现为陕西师范大学计算科学学院博士生导师. 主要研究领域为密码学与信息安全.
E-mail: shundong@snnu.edu.cn

赵雪玲 女,1996年生,陕西西安人. 现为陕西师范大学计算机科学学院硕士研究生. 主要研究领域为密码学、信息安全、有关集合的保密计算.
E-mail: xueling@snnu.edu.cn

家珠亮 女,1992年生,山西运城人. 现为陕西师范大学计算机科学学院硕士研究生. 主要研究领域为密码学、信息安全、有关统计量的保密计算.
E-mail: zhuliang@snnu.edu.cn