计算机科学
COMPUTER SCIENCE

集合交集与并集的安全多方计算

谢琼, 王维琼, 许豪杰

引用本文

谢琼, 王维琼, 许豪杰. 集合交集与并集的安全多方计算[J]. 计算机科学, 2024, 51(2): 371-377.

XIE Qiong, WANG Weiqiong, XU Haojie. Secure Multiparty Computation of Set Intersection and Union [J]. Computer Science, 2024, 51(2): 371-377.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221000235

集合交集与并集的安全多方计算

谢 琼 王维琼 许豪杰
长安大学理学院 西安 710064
(xieqiongxq@163.com)

摘 要 集合的安全多方计算问题是保密科学计算研究的重要问题之一,在电子选举、门限签名、保密拍卖等场景中有着重要的应用。文中主要研究多个集合的保密计算问题,首先针对不同的集合运算提出了对应的转化方式将集合转化为向量,然后基于哥德尔编码提出了新的编码方式,再结合 ElGamal 门限加密算法设计了半诚实模型下可输出多个集合交集或并集,以及同时输出交集与并集的保密计算协议,最后应用模拟范例证明了协议的安全性,协议可以抵抗任意的合谋攻击。实验测试了协议的执行效率,当集合的势满足一定条件时,与现有协议相比,所提协议的计算效率更高。

关键词:安全多方计算;集合交集与并集;ElGamal 加密算法;半诚实模型;模拟范例
中图分类号 TP309.7

Secure Multiparty Computation of Set Intersection and Union

XIE Qiong, WANG Weiqiong and XU Haojie
School of Science, Chang'an University, Xi'an 710064, China

Abstract Secure multiparty computation of sets is one of the most important problems in confidential scientific computing research, which has significant applications in electronic election, threshold signature, and confidential auction. This paper mainly studies secure set operations for multiple parties. Corresponding coding methods are proposed for different set operations to transform sets into vectors, and then these vectors are divided in pairs and encoded by Gödel coding. Combined with the ElGamal threshold encryption algorithm with homomorphism, several secure computing protocols for set intersection and union operations are designed in the semi-honest model. These protocols can resist any collusive attack of arbitrary parties and the simulation paradigm is used to prove that these proposed protocols are secure in the semi-honest model. The protocols' efficiency is verified by experiments. When the cardinality of set meets certain conditions, the proposed protocols have higher computational efficiency compared with the existing schemes.

Keywords Secure multiparty computation, Set intersection and union, ElGamal encryption algorithm, Semi-honest model, Simulation paradigm

1 引言

安全多方计算[1]指多个参与者将保密数据作为计算输入,参与者协同进行保密计算,但无法获得计算结果外的其他额外信息。集合的安全多方计算是保密科学计算[2-4]研究的重要内容之一,在自然科学、工程技术、社会科学等领域有着重要的现实意义和应用价值。近年来,研究学者对保密判断元素与集合的关系[5]、保密计算集合的交集或并集[6-8]以及保密判断集合包含关系[9-10]等问题进行了大量研究。

针对多方集合的保密计算问题,文献[4]中应用多项式表示集合,结合多项式不经意计算的方法,设计了 n 个集合的交集或并集及其势的计算协议。若每个集合的势确定,则交集或并集计算协议具有二次计算复杂性和线性通信复杂性。文献[11]提出用多项式的点表示集合,设计了多方集合交集的保密计算协议,其计算复杂性与集合的势为拟线性关系。文献[12]应用 Bloom 过滤器设计了集合交集或并集势的保密计算协议,具体协议没有应用公钥加密算法,计算复杂度较低,但参与者之间需要建立安全信道,并且最终结果仅为近似值。文献[13]通过不经意排序和数据比较,研究了集合和多重集的多种保密运算,如果所有集合的势和为 m,那么协议具有 $O(m\log m)$ 的计算复杂度和通信复杂度,但文中协议要求各参与者之间具有安全认证信道。在给定全集时,文献[14]和文献[15]应用不同的编码方式,结合具有同态性质的加密算法,将集合的安全多方计算问题转化为向量的安全计算

到稿日期:2022-10-28 返修日期:2023-02-20
基金项目:陕西省自然科学基础研究计划(2020JQ-343);陕西省高校科协青年人才托举计划(20200505)
This work was supported by the Natural Science Basis Research Plan in Shaanxi Province of China(2020JQ-343) and Young Talent Fund of University Association for Science and Technology in Shaanxi, China(20200505).
通信作者:王维琼(wqwang@chd.edu.cn)

问题，如果有 n 个集合参与保密计算且全集的势为 m，那么协议的计算复杂度和通信复杂度分别为 $O(nm)$ 和 $O(n)$。

已有的集合保密计算协议大多集中于两方集合的计算问题，对多个集合的保密计算研究较少，且在现有的研究成果中，协议的计算开销仍较高。此外，具体协议只能输出交集或并集一种结果，但实际问题中存在需要同时计算交集和并集的情况。例如，n 位零件制造商 $P_1, P_2, \cdots, P_n$，每位制造商能够制造的零件型号构成其私密集合 $A_1, A_2, \cdots, A_n$，他们要进行一个行业交流，希望在不泄露自己私密信息的同时，得知大家都能够制造的零件型号和行业内能够制造的所有零件型号，以便安排后续的生产和研发工作。这就需要保密计算多方集合的交与并。本文在确定集合元素取值范围的情况下，针对不同的集合运算提出了新的编码方式，结合具有乘法同态性的门限加密算法，设计了可以输出多方集合交集或并集，以及同时输出交集与并集的保密计算协议。

2 预备知识

2.1 安全性定义

安全多方计算中常用的模型有半诚实模型和恶意模型两种。半诚实模型下的参与者是诚实且好奇的，它们会严格履行协议内容，但会收集和保留执行协议中的相关信息，并在完成协议后试图根据所获得的信息推导其他参与者的私密信息。如果所有参与者都是半诚实的，则称这样的计算模型为半诚实模型[16]。

模拟范例 1[17] 设 $P_1, P_2, \cdots, P_n$ 分别拥有保密数据 $x_1, x_2, \cdots, x_n$，记 $\boldsymbol{X}=(x_1, x_2, \cdots, x_n)$，它们利用协议 Π 计算函数 $f(\boldsymbol{X})=(f_1(\boldsymbol{X}), f_2(\boldsymbol{X}), \cdots, f_n(\boldsymbol{X}))$，其中 $f_i(\boldsymbol{X})$ 为参与者 $P_i(i=1,2,\cdots,n)$ 得到的输出结果。在协议执行过程中，P_i 得到的信息序列记为：

$$view_i^{\Pi}(\boldsymbol{X})=(x_i, r_i, M_i^1, M_i^2, \cdots, M_i^t)$$

其中，r_i 表示 P_i 在协议中产生的随机数，$M_i^j(j=1,2,\cdots,t)$ 表示 P_i 收到的第 j 条信息。对集合 $I=\{P_{i_1}, P_{i_2}, \cdots, P_{i_a}\}\subseteq\{P_1, P_2\cdots, P_n\}$，记

$$view_I^{\Pi}(\boldsymbol{X})=(I, view_{i_1}^{\Pi}(\boldsymbol{X}), view_{i_2}^{\Pi}(\boldsymbol{X}), \cdots, view_{i_a}^{\Pi}(\boldsymbol{X}))$$

定理 1(半诚实模型下协议的安全性[16]) 在参与者都是半诚实的情况下，如果存在概率多项式时间算法 S，使得对任意部分参与者集合 $I=\{P_{i_1}, P_{i_2}, \cdots, P_{i_a}\}\subseteq\{P_1, P_2, \cdots, P_n\}$，式(1)均成立：

$$\{S(I,(x_{i_1}, x_{i_2}, \cdots, x_{i_a}), f_I(\boldsymbol{X}))\}_{\boldsymbol{X}\in(\{0,1\}^*)^n}\overset{c}{\equiv}\{view_I^{\Pi}(\boldsymbol{X})\}_{\boldsymbol{X}\in(\{0,1\}^*)^n} \tag{1}$$

其中 $\overset{c}{\equiv}$ 表示计算上不可区分，则称协议 Π 保密地计算了函数 $f(\boldsymbol{X})$。

2.2 ElGamal 门限加密算法

在安全多方计算中，门限密码体制[18-19]是用来抵抗合谋攻击的重要工具。在门限密码体制中，n 个参与者选取各自的私钥，并协作生成公钥，参与者可以使用公钥加密明文信息，但需要一定数量的参与者共同解密密文。如果至少需要 n 个参与者中的 t 个人合作才能解密密文，那么称这样的密码体制为 (t,n) 门限密码体制。结合具有乘法同态性的 ElGamal 加密算法[20]构造 (n,n) 门限加密算法，具体构造如下。

给定安全参数 k，生成大素数 p 和生成元 $g\in Z_p^*$，各参与者 $P_i(i=1,2,\cdots,n)$ 随机选取一个私钥 $sk_i\in Z_p^*$，计算 $h_i=g^{sk_i} \bmod p$，并共同生成公钥 h，计算式如下：

$$h=\prod_{i=1}^{n}h_i \bmod p=g^{\sum_{i=1}^{n}sk_i} \bmod p$$

对明文 M 加密，选取随机数 r，加密过程如下：

$$E(M)=(c_1, c_2)=(g^r \bmod p, Mh^r \bmod p)$$

对所得密文 $E(M)=(c_1, c_2)$ 解密，解密过程如下：

$$M=c_2\prod_{i=1}^{n}c_1^{-sk_i} \bmod p$$

ElGamal 门限加密算法具有乘法同态性质：

$$E(M_1)\times E(M_2)=(g^{r_1+r_2} \bmod p, M_1M_2h^{r_1+r_2} \bmod p)=E(M_1M_2)$$

3 集合交集的保密计算

3.1 集合交集的保密计算协议

假设 $n(n>2)$ 个参与者 $P_1, P_2, \cdots, P_n$ 分别拥有私密集合 $A_i(i=1,2,\cdots,n)$，他们想保密计算 $\bigcap_{i=1}^{n}A_i$，且不泄露其他任何私密信息。

设 $a_1, a_2, \cdots, a_n$ 是一有限序列，令

$$x=\prod_{i=1}^{n}p_i^{a_i}$$

称 x 为有限序列 $a_1, a_2, \cdots, a_n$ 的哥德尔数，其中 a_i 为非负整数，p_i 表示从小到大排列的第 i 个素数，即 $p_1=2, p_2=3, p_3=5, \cdots$

给定全集 $U=\{u_1, u_2\cdots, u_m\}$，$m$ 为偶数，满足 $A_i\subseteq U(i=1,2,\cdots,n)$。参与者 P_i 根据私密集合 A_i 构造向量 $\boldsymbol{T}_i=(t_{i1}, t_{i2}, \cdots, t_{im})$。其中，

$$t_{ij}=\begin{cases}0, & u_j\in A_i\\ r_{ij}, & u_j\notin A_i\end{cases}, i\in\{1,2,\cdots,n\},\ j\in\{1,2,\cdots,m\} \tag{2}$$

其中，r_{ij} 为 P_i 随机选取的正整数。若给定全集 U 的势 m 为奇数，则随机增加一个元素，使得全集的势为偶数。因此在下文的协议设计中给定全集的势为偶数。

对向量 $\boldsymbol{T}_i$ 的分量按顺序两两分组，得到 $\boldsymbol{T}_i=[(t_{i1}, t_{i2}), (t_{i3}, t_{i4}), \cdots, (t_{i,m-1}, t_{im})]$。按照哥德尔编码，将第 $k(k=1,2,\cdots,m/2)$ 个分组 $(t_{i,2k-1}, t_{i,2k})$ 编码为自然数 $v_{ik}=2^{t_{i,2k-1}}\cdot 3^{t_{i,2k}}$，得到 $m/2$ 维的向量 $\boldsymbol{V}_i=(v_{i1}, v_{i2}, \cdots, v_{i,m/2})$。将向量 $\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_n$ 的各分量对应相乘，得到向量

$$\boldsymbol{D}=(2^{\sum_{i=1}^{n}t_{i1}}\cdot 3^{\sum_{i=1}^{n}t_{i2}}, 2^{\sum_{i=1}^{n}t_{i3}}\cdot 3^{\sum_{i=1}^{n}t_{i4}}, \cdots, 2^{\sum_{i=1}^{n}t_{i(m-1)}}\cdot 3^{\sum_{i=1}^{n}t_{im}})$$

用算数基本定理将向量 $\boldsymbol{D}$ 的每个分量展开得到向量 $\boldsymbol{W}=(\sum_{i=1}^{n}t_{i1}, \sum_{i=1}^{n}t_{i2}, \cdots, \sum_{i=1}^{n}t_{im})$。

命题 1 对任意 $j\in\{1,2,\cdots,m\}$，$u_j\in\bigcap_{i=1}^{n}A_i$ 当且仅当 $\sum_{i=1}^{n}t_{ij}=0$。

证明：如果 $u_j\in\bigcap_{i=1}^{n}A_i$，则对所有 $i\in\{1,2,\cdots,n\}$，都有 $u_j\in A_i$，因此 $t_{ij}=0$，从而 $\sum_{i=1}^{n}t_{ij}=0$。

如果 $u_j\notin\bigcap_{i=1}^{n}A_i$，则至少存在一个 $t_{ik}=r_{ik}$，使得 $\sum_{i=1}^{n}t_{ij}\neq 0$。

这就说明如果$\sum_{i=1}^{n} t_{ij}=0$,那么对所有的 $i\in\{1,2,\cdots,n\}$,都有 $t_{ij}=0$,即 $u_j\in A_i$,因此 $u_j\in\bigcap_{i=1}^{n}A_i$。

命题 1 是多方集合交集保密计算的基本原理,为了保密各个参与者的集合元素,我们通过 ElGamal 门限加密算法设计多方集合交集的保密计算协议,具体协议设计如协议 1 所示。

协议 1　集合交集计算协议

输入:$P_1,P_2,\cdots,P_n$ 的私密集合 $A_1,A_2,\cdots,A_n$

输出:交集 $f(A_1,A_2,\cdots,A_n)=\bigcap_{i=1}^{n}A_i$

准备工作:给定全集 $U=\{u_1,u_2\cdots,u_m\}$,满足 $A_i\subseteq U(i=1,2,\cdots,n)$。运行 ElGamal 门限加密算法,各参与者 P_i 随机选取私钥 sk_i,并共同计算公钥 $h=g^{\sum_{i=1}^{n} sk_i} \bmod p$。

1. 参与者 $P_i(i=1,2,\cdots,n)$根据式(2)将私密集合 A_i 转化为向量 $\mathbf{T}_i=(t_{i1},t_{i2},\cdots,t_{im})$。

2. P_i 将向量 $\mathbf{T}_i$ 的分量按顺序两两分组,根据哥德尔编码原理将每个分组编码自然数,得到 m/2 维向量 $\mathbf{V}_i$:

$$\mathbf{V}_i=(v_{i1},v_{i2},\cdots,v_{i,m/2})=(2^{t_{i1}}\cdot 3^{t_{i2}},2^{t_{i3}}\cdot 3^{t_{i4}},\cdots,2^{t_{i,m-1}}\cdot 3^{t_{im}})$$

使用 ElGamal 门限加密算法加密 $\mathbf{V}_i$ 的每个分量,得到 $\mathbf{C}_i=(E(v_{i1}),E(v_{i2}),\cdots,E(v_{i,m/2}))$。

3. P_1 将 $\mathbf{C}_1$ 发送给 P_2。

4. $P_i(i=2,3,\cdots,n-1)$计算 $\mathbf{C}_1\cdots\mathbf{C}_{i-1}$ 与 $\mathbf{C}_i$ 对应分量的乘积,并将其发送给 P_{i+1}。

5. P_n 计算 $\mathbf{C}$,即

$$\mathbf{C}=(\prod_{i=1}^{n}c_{i1},\prod_{i=1}^{n}c_{i2},\cdots,\prod_{i=1}^{n}c_{i,m/2}),$$并公开 $\mathbf{C}$。

6. 所有参与者共同解密 $\mathbf{C}$,得到向量 $\mathbf{D}=(d_1,d_2,\cdots,d_{m/2})$,用算数基本定理将向量 $\mathbf{D}$ 的各分量展开,得到向量 $\mathbf{W}$:

$$\mathbf{W}=(w_1,w_2,\cdots,w_m)$$

其中 $w_j=\sum_{i=1}^{n}t_{ij}(j=1,2,\cdots,m)$。

7. 对任意 $j\in\{1,2,\cdots,m\}$,如果 $w_j=0$,则 $u_j\in\bigcap_{i=1}^{n}A_i$。

8. 输出交集 $\bigcap_{i=1}^{n}A_i$。

3.2　协议的正确性和安全性

协议 1 的正确性意味着对 n 名参与者任意输入的集合 $A_i\subseteq U(i=1,2,\cdots,n)$,协议 1 都能正确地输出集合的交集 $\bigcap_{i=1}^{n}A_i$。参与者根据式(2)将集合转化为向量,该向量通过分量顺序分组和哥德尔编码得到了一个维数较低的向量。结合 ElGamal 门限加密算法的乘法同态性,协议 1 的正确性可由命题 1 保证。

协议 1 的安全性是基于 ElGamal 门限加密算法的安全性。门限加密算法的公钥由所有参与者共同生成,因此密文解密需要所有参与者合作。只要有任意一个参与者不参与解密,那么每个参与者产生的密文对其他参与者来说都是计算不可区分的。在半诚实模型下,我们通过模拟范例证明协议的安全性。

定理 1　在半诚实模型下协议 1 是安全的,能够抵抗任意的合谋攻击。

证明:对于任意 n 个参与者构成的合谋者集合 I,构造模拟器 S,满足式(1)。由于各个参与者计算地位相等,不妨取 $I=\{P_2,P_3\cdots,P_n\}$,他们想要合谋得知 P_1 的私密集合 A_1 中的元素,S 按如下方式运行:

(1)S 运行 ElGamal 门限加密算法生成私钥 sk_i,并计算公钥 pk。

(2)对输入$(A_2,A_3,\cdots,A_n,f(A_1,A_2,\cdots,A_n))$,$S$ 构造集合 A_1',使得 $f(A_1',A_2,\cdots,A_n)=f(A_1,A_2,\cdots,A_n)$。

根据式(2)构造 A_1'对应的向量 $\boldsymbol{T}_1'$。

(3)根据协议 1 的步骤 2 对 $\boldsymbol{T}_1'$的分量按顺序两两分组,将每个分组根据哥德尔编码原理编码,然后加密,得到向量 $\boldsymbol{C}_1'=(c_{11}',c_{12}',\cdots,c_{1,m/2}')$。

(4)S 计算密文乘积

$$\boldsymbol{C}'=(c_{11}'\prod_{i=2}^{n}c_{i1},c_{12}'\prod_{i=2}^{n}c_{i2},\cdots,c_{1,m/2}'\prod_{i=2}^{n}c_{i,m/2})$$

(5)S 解密得到 $\boldsymbol{D}'=(d_1',d_2',\cdots,d_{m/2}')$,用算术基本定理展开后得到 $\boldsymbol{W}'=(w_1',w_2',\cdots,w_m')$。

在协议执行过程中,

$view_1^{\Pi}(A_1,A_2,\cdots,A_n)=\{A_2,A_3,\cdots,A_n,C,D,W\}$

令

$S(A_2,A_3,\cdots,A_n,f(A_1,A_2,\cdots,A_n))=\{A_2,A_3,\cdots,A_n,\boldsymbol{C}',\boldsymbol{D}',\boldsymbol{W}'\}$

ElGamal 门限加密算法是语义安全的[20]乘法同态加密算法,需要所有参与者共同参与才能解密密文。在 P_1 不参与的情况下,即使集合 I 中的所有参与者合谋也无法解密 $\boldsymbol{C}'$。因此对合谋集合 I 中的参与者来说,$\boldsymbol{C}$ 与 $\boldsymbol{C}'$、$\boldsymbol{D}$ 与 $\boldsymbol{D}'$在计算上不可区分,即 $\boldsymbol{C}\overset{c}{\equiv}\boldsymbol{C}'$,$\boldsymbol{D}\overset{c}{\equiv}\boldsymbol{D}'$,从而有

$$(w_1,w_2,\cdots,w_m)\overset{c}{\equiv}(w_1',w_2',\cdots,w_m')$$

故,

$$\{view_1^{\Pi}(A_1,A_2,\cdots,A_n)\}_{A_i\subseteq U,i=1,2,\cdots,n}\overset{c}{\equiv}\{S(A_2,A_3,\cdots,A_n,f(A_1,A_2,\cdots,A_n))\}_{A_i\subseteq U,i=1,2,\cdots,n}$$

定理得证。

4　集合并集的保密计算

假设 $n(n>2)$个参与者 $P_1,P_2,\cdots,P_n$ 分别拥有私密集合 $A_i(i=1,2,\cdots,n)$,他们想保密计算 $\bigcup_{i=1}^{n}A_i$,且不泄露其他任何私密信息。

给定全集 $U=\{u_1,u_2\cdots,u_m\}$,m 为偶数,满足 $A_i\subseteq U(i=1,2,\cdots,n)$。参与者 P_i 根据私密集合 A_i 构造向量 $\boldsymbol{T}_i=(t_{i1},t_{i2},\cdots,t_{im})$,其中

$$t_{ij}=\begin{cases}r_{ij}, & u_j\in A_i\\ 0, & u_j\notin A_i\end{cases},i\in\{1,2\cdots,n\},j\in\{1,2,\cdots,m\}\quad(3)$$

其中,r_{ij} 为 P_i 随机选取的正整数。

对向量 $\boldsymbol{T}_i$ 的分量按顺序两两分组,得到 $\boldsymbol{T}_i=[(t_{i1},t_{i2}),(t_{i3},t_{i4}),\cdots,(t_{i,m-1},t_{im})]$。按照哥德尔编码,将第 $k(k=1,2,\cdots,m/2)$个分组 $(t_{i,2k-1},t_{i,2k})$编码为自然数 $v_{ik}=2^{t_{i,2k-1}}\cdot 3^{t_{i,2k}}$,得到 $m/2$ 维的向量 $\boldsymbol{V}_i=(v_{i1},v_{i2},\cdots,v_{i,m/2})$。将向量 $\boldsymbol{V}_1$,$\boldsymbol{V}_2,\cdots,\boldsymbol{V}_n$ 的各分量对应相乘,得到向量

$$\boldsymbol{D}=(2^{\sum_{i=1}^{n}t_{i1}}\cdot 3^{\sum_{i=1}^{n}t_{i2}},2^{\sum_{i=1}^{n}t_{i3}}\cdot 3^{\sum_{i=1}^{n}t_{i4}},\cdots,2^{\sum_{i=1}^{n}t_{i(m-1)}}\cdot 3^{\sum_{i=1}^{n}t_{im}}$$

用算数基本定理将向量 $\boldsymbol{D}$ 的每个分量展开得到向量 $\boldsymbol{W}=(\sum_{i=1}^{n}t_{i1},\sum_{i=1}^{n}t_{i2},\cdots,\sum_{i=1}^{n}t_{im})$。

命题 2 对任意 $j\in\{1,2,\cdots,m\}$，$u_j\in\bigcup_{i=1}^{n}A_i$ 当且仅当 $\sum_{i=1}^{n}t_{ij}\neq0$。

证明：如果 $u_j\in\bigcup_{i=1}^{n}A_i$，那么存在 $i\in\{1,2,\cdots,n\}$，使得 $u_j\in A_i$，就有 $t_{ij}=r_{ij}$，从而 $\sum_{i=1}^{n}t_{ij}\neq0$。

如果 $\sum_{i=1}^{n}t_{ij}\neq0$，那么存在 $i\in\{1,2,\cdots,n\}$，使得 $t_{ij}=r_{ij}$，即 $u_j\in A_i$。这意味着 $u_j\in\bigcup_{i=1}^{n}A_i$。

根据命题 2，我们可设计多方集合并集保密计算协议。具体协议设计只需对协议 1 中的第 1 步和第 7 步进行部分修改，其余步骤保持不变。最后输出集合并集 $\bigcup_{i=1}^{n}A_i$。具体设计如协议 2 所示。

协议 2 集合并集计算协议

1. 参与者 $P_i(i=1,2,\cdots,n)$ 根据式(3)将私密集合 A_i 转化为向量 $\mathbf{T}_i=(t_{i1},t_{i2},\cdots,t_{im})$。

步骤 2－步骤 6 同协议 1

7. 对任意 $j\in\{1,2,\cdots,m\}$，如果 $w_j\neq0$，则 $u_j\in\bigcup_{i=1}^{n}A_i$。

8. 输出并集 $\bigcup_{i=1}^{n}A_i$。

协议 2 的正确性和安全性可分别由命题 2 和定理 2 保证。定理 2 的证明过程与定理 1 类似，故省略定理的证明过程，仅叙述下面定理。

定理 2 在半诚实模型下协议 2 是安全的，能够抵抗任意的合谋攻击。

5 集合交集与并集的保密计算

5.1 集合交集与并集的保密计算协议

假设 $n(n>2)$ 个参与者 $P_1,P_2,\cdots,P_n$ 分别拥有私密集合 $A_i(i=1,2,\cdots,n)$，他们想同时保密计算 $\bigcap_{i=1}^{n}A_i$ 和 $\bigcup_{i=1}^{n}A_i$，且不泄露其他任何私密信息。

给定全集 $U=\{u_1,u_2\cdots,u_m\}$，m 为偶数，满足 $A_i\subseteq U(i=1,2,\cdots,n)$。参与者 P_i 根据私密集合 A_i 构造向量 $\boldsymbol{T}_i=(t_{i1},t_{i2},\cdots,t_{im})$，其中

$$t_{ij}=\begin{cases}1, & u_j\in A_i\\ 0, & u_j\notin A_i\end{cases},i\in\{1,2\cdots,n\},j\in\{1,2,\cdots,m\}\qquad(4)$$

对向量 $\boldsymbol{T}_i$ 的分量按顺序两两分组，得到 $\boldsymbol{T}_i=[(t_{i1},t_{i2}),(t_{i3},t_{i4}),\cdots,(t_{i,m-1},t_{im})]$。按照哥德尔编码，将第 $k(k=1,2,\cdots,m/2)$ 个分组 $(t_{i,2k-1},t_{i,2k})$ 编码为自然数 $v_{ik}=2^{t_{i,2k-1}}\cdot3^{t_{i,2k}}$，得到 $m/2$ 维的向量 $\boldsymbol{V}_i=(v_{i1},v_{i2},\cdots,v_{i,m/2})$。将向量 $\boldsymbol{V}_1$，$\boldsymbol{V}_2$，$\cdots$，$\boldsymbol{V}_n$ 的各分量对应相乘，得到向量

$$\boldsymbol{D}=(2^{\sum_{i=1}^{n}t_{i1}}\cdot3^{\sum_{i=1}^{n}t_{i2}},2^{\sum_{i=1}^{n}t_{i3}}\cdot3^{\sum_{i=1}^{n}t_{i4}},\cdots,2^{\sum_{i=1}^{n}t_{i(m-1)}}\cdot3^{\sum_{i=1}^{n}t_{im}})$$

用算数基本定理将向量 $\boldsymbol{D}$ 的每个分量展开得到向量 $\boldsymbol{W}=(\sum_{i=1}^{n}t_{i1},\sum_{i=1}^{n}t_{i2},\cdots,\sum_{i=1}^{n}t_{im})$。

命题 3 对任意 $j\in\{1,2,\cdots,m\}$，$u_j\in\bigcap_{i=1}^{n}A_i$ 当且仅当 $\sum_{i=1}^{n}t_{ij}=n$。

证明：如果 $u_j\in\bigcap_{i=1}^{n}A_i$，则对所有 $i\in\{1,2,\cdots,n\}$，都有 $u_j\in A_i$，故 $t_{ij}=1$，从而 $\sum_{i=1}^{n}t_{ij}=n$。

如果 $u_j\notin\bigcap_{i=1}^{n}A_i$，则至少存在一个 $t_{ik}=0$，使得 $\sum_{i=1}^{n}t_{ij}\neq n$。这就说明如果 $\sum_{i=1}^{n}t_{ij}=n$，那么对所有的 $i\in\{1,2,\cdots,n\}$，都有 $t_{ij}=1$，即 $u_j\in A_i$，因此 $u_j\in\bigcap_{i=1}^{n}A_i$。

命题 4 对任意 $j\in\{1,2,\cdots,m\}$，$u_j\in\bigcup_{i=1}^{n}A_i$ 当且仅当 $\sum_{i=1}^{n}t_{ij}\neq0$。

证明：如果 $u_j\in\bigcup_{i=1}^{n}A_i$，那么存在 $i\in\{1,2,\cdots,n\}$，使得 $u_j\in A_i$，则有 $t_{ij}=1$，从而 $\sum_{i=1}^{n}t_{ij}\neq0$。

如果 $\sum_{i=1}^{n}t_{ij}\neq0$，那么存在 $i\in\{1,2,\cdots,n\}$，使得 $t_{ij}=1$，即 $u_j\in A_i$。这意味着 $u_j\in\bigcup_{i=1}^{n}A_i$。

命题 3 和命题 4 是同时计算若干个集合交集与并集的基本原理，结合具有乘法同态性的 ElGamal 门限加密算法，设计同时保密计算集合交集和并集的协议。具体协议设计只需将协议 1 中的第 1 步和第 7 步进行简单修改，其余保持不变，最后协议同时输出集合交集 $\bigcap_{i=1}^{n}A_i$ 和并集 $\bigcup_{i=1}^{n}A_i$。具体协议设计如协议 3 所示。

协议 3 集合交集与并集计算协议

1. 参与者 $P_i(i=1,2,\cdots,n)$ 根据式(4)将私密集合 A_i 转化为向量 $\mathbf{T}_i=(t_{i1},t_{i2},\cdots,t_{im})$。

步骤 2－步骤 6 同协议 1

7. 对任意 $j\in\{1,2,\cdots,m\}$，如果 $w_j=n$，则 $u_j\in\bigcap_{i=n}^{n}A_i$；如果 $w_j\neq0$，则 $u_j\in\bigcup_{i=1}^{n}A_i$。

8. 输出交集 $\bigcap_{i=1}^{n}A_i$ 和并集 $\bigcup_{i=1}^{n}A_i$。

协议 3 的正确性由命题 3 和命题 4 共同保证。协议 3 的安全性由定理 3 来证明，由于证明过程与定理 1 完全相同，故省略。

定理 3 在半诚实模型下协议 3 是安全的，能够抵抗任意的合谋攻击。

5.2 实例计算

以同时输出多方集合交集与并集的保密计算为例，简单说明上述协议的实际应用性。为减少加解密过程中的运算量，输入简单的集合。

选取 ElGamal 门限加密算法的参数：生成元 $g=2$，素数 $p=269$。

假设 3 个半诚实参与者 P_1,P_2,P_3 分别拥有私密集合 $A_1=\{1,2,3\}$，$A_2=\{2,4\}$，$A_3=\{2,4,5\}$。令全集 $U=\{1,2,\cdots,6\}$，同时保密计算 3 个集合的交集与并集。

(1) P_1,P_2,P_3 分别选取私钥 $sk_1=4$，$sk_2=12$，$sk_3=17$，并分别计算

$$h_1\equiv g^{sk_1}\bmod 269\equiv16\bmod 269$$

$$h_2\equiv g^{sk_2}\bmod 269\equiv61\bmod 269$$

$h_3 \equiv g^{sk_3} \bmod 269 \equiv 69 \bmod 269$

公开 h_1, h_2, h_3。

(2) P_1, P_2, P_3 共同计算公钥 h：

$h \equiv 16 \cdot 61 \cdot 69 \equiv 94 \bmod 269$

(3) P_1, P_2, P_3 根据式(4)将各自的私密集合转化为向量：

$\boldsymbol{T}_1 = (1,1,1,0,0,0)$

$\boldsymbol{T}_2 = (0,1,0,1,0,0)$

$\boldsymbol{T}_3 = (0,1,0,1,1,0)$

(4) P_1, P_2, P_3 分别将 $\boldsymbol{T}_1, \boldsymbol{T}_2, \boldsymbol{T}_3$ 中的元素两两划分，并用哥德尔编码方式将其编码为 3 维向量：

$\boldsymbol{U}_1 = (6,2,1)$

$\boldsymbol{U}_2 = (3,3,1)$

$\boldsymbol{U}_3 = (3,3,2)$

(5)加密 $\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3$ 得到加密后的向量为 $\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3$：

$\boldsymbol{C}_1 = [(8,10),(32,222),(243,174)]$

$\boldsymbol{C}_2 = [(64,98),(16,201),(128,22)]$

$\boldsymbol{C}_3 = [(243,253),(165,118),(256,101)]$

其中，P_1 加密 $\boldsymbol{U}_1$ 各分量选取的随机数为 $r=3,5,9$，P_2 加密 $\boldsymbol{U}_2$ 各分量选取的随机数为 $r=6,4,7$，P_3 加密 $\boldsymbol{U}_3$ 各分量选取的随机数为 $r=9,11,8$。

(6) P_1, P_2, P_3 将加密后向量的对应分量相乘，得到乘积 $\boldsymbol{C}$：

$\boldsymbol{C} = [(138,191),(14,259),(224,75)]$

(7) P_1, P_2, P_3 联合解密 $\boldsymbol{C}$，得到解密后的向量 $\boldsymbol{D} = (54,18,2)$。用算数基本定理将 $\boldsymbol{D}$ 展开得到向量 $\boldsymbol{W}$：

$\boldsymbol{W} = (1,3,1,2,1,0)$

因此，求得的交集为 $A_1 \cap A_2 \cap A_3 = \{2\}$，并集为 $A_1 \cup A_2 \cup A_3 = \{1,2,3,4,5\}$。

6　协议的效率分析

6.1　复杂性分析

6.1.1　计算复杂性分析

本文设计的 3 个协议均使用了 ElGamal 门限加密方案，在计算复杂度分析中忽略了协议执行过程中所需要的数乘运算，只考虑了耗时较多的模指数运算。应用 ElGamal 门限加密方案加密一次需要进行两次模指数运算，n 个参与者共同解密一个密文需要进行 $n+1$ 次模指数运算。

在 3 个协议中，n 个参与者首先将私密集合转化为向量，对向量的分量按顺序分组，分组后根据哥德尔编码将 m 维向量转化为 $m/2$ 维向量。每个参与者需要用 ElGamal 加密算法加密 $m/2$ 个自然数，因此需要 $2n \cdot m/2$ 次模指数运算；然后，除去第一个参与者外，每个参与者都需要把上一个参与者发送来的密文和自己的密文做乘法运算；最后，所有参与者合作对获得的 $m/2$ 维的密文乘积向量解密，需要 $(n+1) \cdot m/2$ 次模指数运算。所以 3 个协议都需要 $m(3n+1)/2$ 次模指数运算。

6.1.2　通信复杂性分析

在 3 个协议中，n 个参与者应用 ElGamal 门限加密算法进行计算，首先需要 $n-1$ 次通信来构造公钥；然后每个参与者将密文乘积运算后得到的新密文发送给下一个参与者，得到所有参与者密文的乘积，这个过程需要 $n-1$ 次通信；最后所有参与者合作解密需要 $n-1$ 次通信。所以 3 个协议都需要 $3(n-1)$ 次通信。

本文结果与现有方案的结果对比如表 1 所列。其中 n 为参与者个数，m 为给定的全集的势，k_i 为参与者集合的势，$k_{\min} = \min\{k_1, k_2, \cdots, k_n\}$，$k_{\max} = \max\{k_1, k_2 \cdots, k_n\}$，则当 $k_{\min} > m(n-1)/4n$ 时，本文交集计算协议所需要的模指数运算次数小于文献[14]，当 $k_{\max} < 3m(n+1)/4n$ 时，本文并集计算协议所需要的模指数运算次数小于文献[14]。本文的交集或并集计算协议与文献[15]相比，均减少了 $m(3n+1)/2$ 次模指数运算。

表 1　本文方法与现有方法的对比

Table 1　Comparison between the proposed method and existing methods

文献	计算功能	模指数运算次数	通信次数
文献[14]协议 1	交集	$(n+1)m+2\sum_{i=1}^{n} k_i$	$3(n-1)$
文献[15]协议 2		$m(3n+1)$	
本文协议 1		$m(3n+1)/2$	
文献[14]协议 2	并集	$(n+1)m+2\sum_{i=1}^{n}(m-k_i)$	$3(n-1)$
文献[15]协议 1		$m(3n+1)$	
本文协议 2			
本文协议 3	交集与并集	$\boldsymbol{m(3n+1)/2}$	$3(n-1)$

6.2　实验测试

本文对协议进行实验测试的实验环境为：Window11 家庭中文版，AMD Ryzen 7 5800H with Radeon Graphics @ 3.20GHz，内存 16.0GB，64 位操作系统，python3.10.4 编程语言。

固定参与者人数 $n=5$，交集计算协议的执行时间随全集的势 m 的变化规律如图 1 所示。并集计算协议的执行时间随全集的势 m 的变化规律如图 2 所示。交集与并集计算协议的执行时间随全集的势 m 的变化规律如图 3 所示。

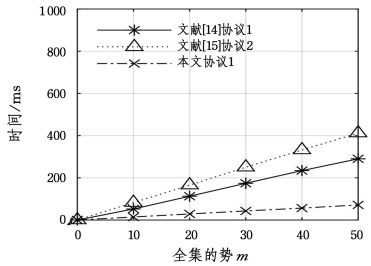

图 1　交集计算协议执行时间随全集势的变化规律

Fig. 1　PSI execution time changes with cardinality of set

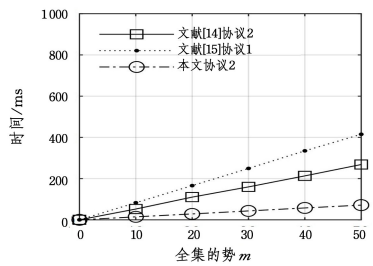

图 2　并集计算协议执行时间随全集势的变化规律

Fig. 2　PSU execution time changes with cardinality of set

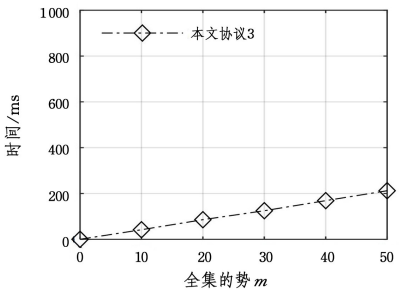

图 3 交集与并集计算协议执行时间随全集势的变化规律

Fig. 3 PSI and PSU execution time changes with cardinality of set

固定全集的势 m，交集计算协议的执行时间随参与者人数 n 的变化规律如图 4 所示。并集计算协议的执行时间随参与者人数 n 的变化规律如图 5 所示。交集与并集计算协议的执行时间随参与者人数 n 的变化规律如图 6 所示。

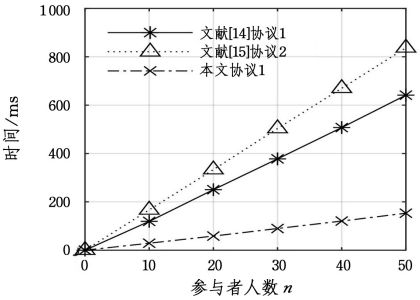

图 4 交集计算协议执行时间随参与者人数的变化规律

Fig. 4 PSI execution time changes with number of participants

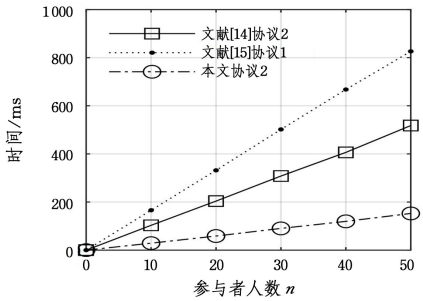

图 5 并集计算协议执行时间随参与者人数的变化规律

Fig. 5 PSU execution time changes with number of participants

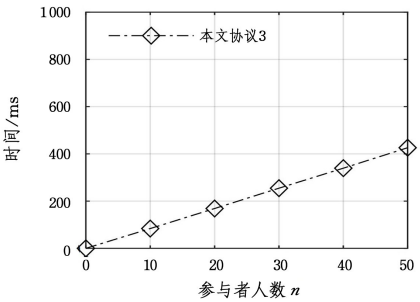

图 6 交集与并集计算协议执行时间随参与者人数的变化规律

Fig. 6 PSI and PSU execution time changes with number of participants

由图 1—图 6 可知，协议的执行时间与全集的势和参与者人数均成线性关系。与文献[14]和文献[15]设计的交集、并集的保密计算协议相比，本文的协议效率明显更高，说明了本文协议的高效性。

结束语 针对集合交集与并集运算，本文借助哥德尔编码设计了一种新的编码方式，并结合 ElGamal 门限加密算法构造了多方集合交集或并集，以及可同时输出多方集合交集与并集的保密计算协议。协议适用于参与者的私密集合包含于某个确定集合的安全多方计算。本文通过模拟范例证明了 3 个协议在半诚实模型下的安全性，并且其可以抵抗任意的合谋攻击。效率分析表明本文提出的协议是高效的。今后我们将进一步研究更高效的整数集合交集、并集和混合运算问题的解决方案，尝试开展有理数域上集合的安全多方计算研究。

参 考 文 献

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

[2] FAGIN R, NAOR M, WINKLERY P. Comparing information without leaking it [J]. Communications of the ACM, 1996, 39(5): 77-85.

[3] BEIMEL A, GABIZON A, ISHAI Y, et al. Non-interactive secure multiparty computation [C] // Advances in Cryptology (CRYPTO 2014). Berlin: Springer, 2014: 387-404.

[4] KISSNER L, SONG D. Privacy-preserving set operations[C]// Advances in Cryptology(CRYPTO 2005). Berlin: Springer, 2005: 241-257.

[5] LI S D, WANG D S, DAI Y Q. Symmetric cryptographic protocols for extended millionaires' problem[J]. Science in China Series F: Information Science, 2009, 52(6): 974-982.

[6] FREEDMAN M J, HAZAY C, NISSIM K, et al. Efficient set intersection with simulation-based security[J]. Journal of Cryptology, 2016, 29(1): 115-155.

[7] CRISTOFARO E D, GASTI P, TSUDIK G. Fast and private computation of cardinality of set intersection and union[C]// Cryptology and Network Security. Berlin: Springer, 2012: 218-231.

[8] SEO J H, CHEON J H, KATZ J. Constant-Round Multi-Party Private Set Union Using Reversed Laurent Series[C]// Public Key Cryptography(PKC 2012). Berlin: Springer, 2012: 398-412.

[9] CHEN Z H, LI S D, HUANG Q, et al. Protocols for secure computation of two set-relationships with the unencrypted method [J]. Journal of Software, 2018, 29(2): 473-482.

[10] ZHOU S F, LI S D, DOU J W, et al. Efficient secure multiparty subset computation[J]. Security and Communication Networks, 2017, 2017(3): 1-11.

[11] CHEON J H, JARECKI S, SEO J H. Multi-party privacy-preserving set intersection with quasi-linear complexity[J]. IEICE Transactions on Fundamentals of Electronics Communications & Computer Sciences, 2010, 95(8): 1366-1378.

[12] EGERT R, FISCHLIN M, GENS D, et al. Privately computing set-union and set-intersection cardinality via bloom filters[C]//

Information Security and Privacy. Switzerland: Springer, 2015: 413-430.

[13] BLANTON M. AGUIAR E. Private and oblivious set and multiset operations[J]. International Journal of Information Security, 2016, 15(5): 493-518.

[14] DOU J W, LIU X H, ZHOU S F. Efficient secure multiparty set operations protocols and their application[J]. Chinese Journal of Computers, 2018, 41(8): 1844-1860.

[15] WANG W L, LI S D, DOU J W, et al. Privacy-preserving mixed set operation[J]. Information Sciences, 2020, 525: 67-81.

[16] GOLDREICH O. The Fundamental of Cryptography: Basic Applications[M]. London: Cambridge University Press, 2004.

[17] 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 and Developmental Disorders, 2013, 43(9): 2211-2217.

[18] DESMEDT Y, FRANKEL Y. Threshold cryptosystems[C]// Advances in Cryptology—CRYPTO' 89 Proceedings Advances in Cryptology. New York: Springer, 1989: 307-315.

[19] LONG Y, CHEN K F, MAO X P. New constructions of dynamic threshold cryptosystem[J]. Journal of Shanghai Jiaotong University(Science), 2014, 19(4): 431-435.

[20] ElGAMAL T. A public key cryptosystem and a signature scheme based on discrete logarithms[J]. IEEE Transactions on Information Theory, 1985, 31(3): 469-472.

XIE Qiong, born in 1998, postgraduate. Her main research interest is cryptography.

WANG Weiqiong, born in 1979, Ph.D, professor. Her main research interests include coding theory and cryptography.

（责任编辑：何杨）

定了！5月16—18日，YEF 2024落地宁波

中国计算机学会青年精英大会(CCF YEF)是由中国计算机学会(CCF)主办的面向计算领域青年精英的年度综合性会议，创办于2011年，每年一届，至今已成功举办十三届。YEF旨在为计算领域的学术界和企业界人士提供深入交流和提升的机会，促进青年精英人才的成长，提升他们的领导力，促进相互之间的合作。YEF现已成为计算领域学术界、企业界最富有影响力的以青年人为主的年度跨界学术、技术交流大会。大会影响力逐年提升，会议逐年扩大，从初期100多人，2022年采取以线上线下结合的方式召开，线上会议的人气值接近41万，到2023年现场参会人数约1 100人。

宁波市海曙区政府将作为大会承办单位，为大会的顺利举行提供服务保障。

海曙区位于宁波市的中心区域，东临奉化江，北濒余姚江，西与余姚市接壤，南与奉化区连接。距今近1 200年前，明州刺史韩察在三江口建立子城，标志着宁波建城的开端。境内有始建于唐长庆元年的鼓楼海曙楼，海曙区因此而得名。

海曙历史文化底蕴深厚，遗存丰富。散发着四百余载翰墨书香的天一阁，是中国现存最早的私家藏书楼，是宁波的城市印记和文化之脉；千年月湖，是宁波市中心开放式的江南山水园林，十洲胜景，人文荟萃，素有“浙东邹鲁”之美誉；始建于唐代的它山堰，是世界灌溉工程遗产，与四川都江堰等并称中国四大古代水利工程；梁祝文化园，是中国古代四大爱情传说梁祝故事的发祥地……海曙，就像一座没有围墙的博物馆，处处镌刻着历史的印迹，时时散发着文化的气息。

海曙的西南，是“物色甲东南”的四明山麓，这里有中国单体面积最大的竹海、中国最美的赏樱基地、中国最大的赏桂佳处，有以碧水龙潭而著称的五龙潭风景名胜区、自然清新的天然氧吧绿谷龙观。龙观禅那、法曼庄园、向阳舍、枕溪山房以及南塘阅居、隐逸月湖、书房、正房客栈……这些或深藏在山乡村野，或隐匿于市井街巷的民宿客栈，无不成为当下放空心灵、寄托情怀的时尚，是现代版的“小隐于野、大隐于市”。

作为宁波中心商贸商务区、历史文化名城核心区，海曙还拥有浙江省首家五星级酒店南苑饭店、宁波的城市客厅天一广场、宁波的“香榭丽舍”和义大道、“浙东第一街”中山路商业特色街，以及历史文化特色街区南塘老街、月湖盛园、鼓楼沿……均以不同的姿态绽放着时尚的魅力。

据CCF微信公众号