保护隐私的曼哈顿距离计算及其推广应用

窦家维 葛 雪 王颖因

(陕西师范大学数学与信息科学学院 西安 710062)

摘 要 安全多方计算是信息时代保护隐私和信息安全的一项关键技术.安全多方科学计算是安全多方计算十分重要的组成部分,目前已经有许多安全多方科学计算问题的解决方案,但还有更多的问题值得人们去研究.关于曼哈顿距离的安全多方计算问题目前研究的结果很少,构造曼哈顿距离的安全计算协议在密码学中有着重要的理论意义,作为基础协议能够广泛应用于其他安全多方计算协议的构造,比如保密计算两点间路径问题,保密判定点与区间以及点与点集的关系问题,以及向量相似度的保密计算都可以归约到曼哈顿距离的安全多方计算问题.本文应用加密选择技巧与一种新的编码方法相结合,以 Paillier 加密算法为基础,对于不同的情形(无全集限制或有全集限制)设计两数之差绝对值的高效保密计算协议.并以此为基础,设计出两种不同情形下保密计算曼哈顿距离的协议.本文证明了在半诚实模型下这些协议是安全的,并通过模拟实验来测试协议的具体执行时间,理论分析和仿真结果表明本文方案是简单易行的.最后,文中给出实例阐明本文协议在理论以及实际中的广泛应用.

关键词 安全多方计算;密码学;曼哈顿距离;Paillier 加密算法;编码方法

中图法分类号 TP309 **DOI 号** 10.11897/SP.J.1016.2020.00352

Secure Manhattan Distance Computation and Its Application

DOU Jia-Wei GE Xue WANG Ying-Nan

(*School of Mathematics and Information Science, Shaanxi Normal University, Xi'an* 710062)

Abstract Secure multiparty computation, also called privacy-preserving computation, is a key technology of privacy protecting and information security in the information age. Secure multiparty computation uses many cryptographic primitives, such as public key cryptosystems, secret sharing, oblivious transfer, bit commitment, one way hash function, garbled circuit, zero-knowledge proof to solve various cryptographic problems. Secure multiparty computation protocols are called general cryptographic protocols, and have many practical applications in scientific computation, data-mining, computational geometry, commerce and statistical analysis, etc. It is a focus of the international cryptographic community in recent years. Since it was introduced by Yao, it has been extensively studied. The fruit of secure multiparty computation is rich. Secure multiparty scientific computation is an important part of secure multiparty computation. Though many secure multiparty scientific computation problems have been studied, but the solutions of many addressed problems are not satisfactory and need to be improved or to explore better solutions. There are also various problems that still need to be studied. Secure Manhattan distance computation is such a problem that needs to be studied. Different from Euclidean distance, Manhattan distance between two points in an Euclidean space with fixed Cartesian coordinate system is defined as the sum of the lengths of the projections of the line segment between the points onto the coordinate axes. It is a reasonable distance measure. Private Manhattan distance computation

收稿日期:2018-10-25;在线出版日期:2019-05-13.本课题得到国家自然科学基金(61272435)资助.**窦家维**,博士,副教授,主要研究方向为应用数学与应用密码学.E-mail:jiawei@snnu.edu.cn.**葛 雪**,硕士研究生,主要研究方向为应用数学与应用密码学.**王颖因**,硕士研究生,主要研究方向为应用数学与应用密码学.

has wide applications in practice. It has important theoretical and practical significance in cryptography to design protocol to privately compute the Manhattan distance, because the protocol can be used as a building block to solve many other secure multiparty computation problems such as privately computing the path between two private points, privately determining the relationship between a private point and a private interval, privately determining the relationship between a private point and a private point set and privately computing the similarity between two vectors. At present, there are few research results on Manhattan distance secure computation. We address this problem in this paper. To privately compute the Manhattan distance between two private points, we first design a new encoding scheme. This encoding scheme encodes a private number into a private array which makes private absolute value computation easy. Then we use the Paillier cryptosystem, the encoding scheme, and encrypt-and-choose technique to design a protocol to privately compute the absolute value of the difference of two private numbers. Using the protocol for absolute value problem as a building block, we further design a protocol to privately compute the Manhattan distance between two private points. We design protocols for two different scenarios: one for the scenario where the domain of the private data is known and fairly small; another for the scenario where the domain of the private data is unknown. We prove that these protocols are secure in the semi-honest model. We analyze and test the efficiency of our protocols. Theoretical analysis and simulation show that our protocols are efficient. Finally, we show how to use our protocol to solve other secure multiparty computation problems, such as the secure vector similarity computation and the secure determination of the relationship between a vector and a vector set.

Keywords　secure multiparty computation; cryptography; Manhattan distance; Paillier homomorphic encryption algorithms; encoding scheme

1　引　言

信息技术的飞速发展在给人类带来便捷的同时也给信息安全与隐私保护带来巨大的挑战.为了解决信息安全与隐私保护问题,安全多方计算(Secure Multiparty Computation,SMC)应运而生.

Yao 在文献[1]中首先提出了两方安全计算问题,此后安全多方计算问题便受到学者们的广泛关注及进一步的深入研究.安全多方计算的目的是要解决一组互不信任的参与方之间保护隐私的协同计算问题.要求在计算完成后既要保证各参与方得到规定的输出结果,又要保证各参与方私密数据的隐私性.目前,安全多方计算是密码学界的热点问题[2-6],一些学者已经给出了通用的解决方案[7-9],但在应用这些通用方案解决实际问题时一般来说计算效率都较低,因此需要针对具体问题设计具体的解决方案.目前研究的安全多方计算问题主要包括:保密的科学计算问题[1,10-16]、保密的计算几何问题[17-19]、保密的统计分析问题[20-21]、保密的数据挖掘问题[22-23],以及安全多方计算的实际应用问题[24-25].

两点间距离的安全计算研究具有重要的实际应用价值.比如在生物特征密码研究中一个基本的问题是如何判别两个生物特征模板是否相同或相似,文献[26]指出可以应用一个 16 维向量表示一个 Finger Code,如此即可将两个 Finger Code 之间差异的保密比较问题转化为相应向量之间距离的保密计算问题.在机器学习和数据挖掘中,经常需要计算不同个体间的差异,进而判定个体的相似性和类别,而保密比较两个个体相似度的问题即可归约到保密计算这些个体对应的特征向量之间的距离.不同的场景往往需要采用不同的度量方法,比如常用欧氏距离度量两点间的空间距离,应用曼哈顿距离度量路径长短等,因此在解决实际问题的过程中经常需要根据数据的不同特征以及具体的应用场景采用相适宜的度量方法.Kikuchi 等人在文献[27]中提出了一个计算两个向量欧氏距离的保密计算协议.目前关于曼哈顿距离的保密计算问题研究很少,近期

文献[28]设计了有关曼哈顿距离保密计算协议,但文中仅考虑了二维平面相关问题,且要求点的坐标限制在一个特定全集内取值.本文将对曼哈顿距离保密计算问题进行进一步深入的研究,针对不同的数据特征,设计两点间曼哈顿距离安全高效的保密计算协议.

本文的主要贡献如下:

(1)本文提出了新的转化思路和技巧:设计了一种新的编码方法,在有全集限制条件下,将一个数编码为一个 0-1 数组,巧妙地解决了绝对值问题的保密计算.这种编码方法也可以推广应用于其他相关问题的安全多方计算.

(2)本文首先解决了绝对值的保密计算问题,并以此为基础,进一步研究设计了曼哈顿距离的保密计算协议.针对参与者数据具有或不具有全集限制条件的特性,设计了两套不同的关于绝对值和曼哈顿距离的保密计算方案.

(3)本文的研究思想和得到的结论具有广泛适用性:利用本文提出的安全计算方案,可以高效解决在各种数据特性情形下曼哈顿距离的保密计算问题.我们可以利用这些协议及其设计思想作为基础去解决其他更多的安全多方计算问题(详见第 6 节).

本文第 2 节介绍了预备知识;第 3 节和第 4 节分别针对不同的数据特性(无全集或有全集限制),设计构造关于绝对值问题以及曼哈顿距离问题的保密计算协议;第 5 节对本文协议效率进行了理论分析与仿真测试;第 6 节对协议在其他安全多方计算问题中的应用进行了推广;第 7 节对全文进行了总结.

2 预备知识

2.1 半诚实模型

半诚实模型[9]:在半诚实模型中,参与者不会主动地进行欺骗,他们会按照协议要求履行协议.但在协议结束后,参与者可能会根据其在协议中获得的信息试图进行推算以获得其他参与者隐私数据的额外信息.

模拟范例. 设参与者 P_1,P_2分别输入 x,y 执行协议 π,他们希望保密计算函数 $f(x,y)=(f_1(x,y),f_2(x,y))$.协议结束后,$P_1$(或 P_2)得到输出结果 $f_1(x,y)$(或 $f_2(x,y)$).$P_i(i=1,2)$得到的信息序列记为 $view_i^{\pi}(x,y)$:

$$view_i^{\pi}(x,y)=(x_i,r_i,m_1^i,m_2^i,\cdots,m_{k_i}^i,f_i(x,y)),$$

其中 $x_1=x$,$x_2=y$,r_i表示 P_i产生的随机数,$m_t^i(t=1,\cdots,k_i)$表示 P_i接收到的第 t 个信息.

安全性定义[9]. 如果存在概率多项式时间算法 S_1 和 S_2,使得:

$$\{S_1(x,f_1(x,y))\}_{x,y}\overset{c}{\equiv}\{view_1^{\pi}(x,y)\}_{x,y} \quad (1)$$

$$\{S_2(y,f_2(x,y))\}_{x,y}\overset{c}{\equiv}\{view_2^{\pi}(x,y)\}_{x,y} \quad (2)$$

则称协议 π 保密地计算了函数 $f(x,y)$,其中$\overset{c}{\equiv}$表示计算上不可区分.

2.2 Paillier 同态加密方案

本文主要应用 Paillier 加密方案设计保密计算协议.下面对加密方案进行简要描述[29].

密钥生成. 根据安全参数 τ,密钥生成算法 G 选择两个素数 p、q,使得 $|p|=|q|=\tau$,令 $N=pq$,$\lambda=\mathrm{lcm}(p-1,q-1)$,并定义 $L(x)=(x-1)/N$.随机选择 $g\in Z_{N^2}^*$,满足 $\gcd(L(g^{\lambda}\bmod N^2),N)=1$.算法的公钥为$(g,N)$,私钥为 λ.

加密. 对于消息 $m\in Z_N$,任意选择随机数 $r\in Z_N^*$,加密得到密文 $C=E(m)$为

$$C=g^m r^N \bmod N^2.$$

解密. 对于密文 $C\in Z_{N^2}^*$,解密可得到明文 $m=D(C)$为

$$m=\frac{L(C^{\lambda}\bmod N^2)}{L(g^{\lambda}\bmod N^2)}\bmod N.$$

加法同态性. 对于任意明文消息 $m_1,m_2\in Z_N$,假设$E(m_1)=g^{m_1}r_1^N \bmod N^2$,$E(m_2)=g^{m_2}r_2^N \bmod N^2$,则有

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

上面性质表明Paillier加密方案具有加法同态性.

注解 1. 在 Paillier 加密方案中,如果生成元 g 选用$g=KN+1$ 的形式,加密一次仅需要 1 次模指数运算(如选用一般的 $g\in Z_{N^2}^*$ 作为生成元,加密一次需要 2 次模指数运算).这是由于如果令 $g=KN+1$,则有

$$C=(KN+1)^m r^N \bmod N^2=(mKN+1)r^N \bmod N^2.$$

因此,选用生成元为 $g=KN+1$ 的形式可以使加密运算的复杂性降低一半.本文的加密算法均假设选用生成元为 $g=KN+1$ 的形式.Paillier 加密方案中解密一个密文需要 2 次模指数运算.

3 曼哈顿距离保密计算(无全集限制)

3.1 问题描述

绝对值的保密计算. 假设 Alice 拥有保密数据 x,Bob 拥有保密数据 y.Alice 和 Bob 想要合作计算

出$|x-y|$的值,且保密各自的私有数据.

曼哈顿距离(Manhattan Distance)保密计算. 假设 Alice,Bob 分别拥有点(或向量)$\boldsymbol{X},\boldsymbol{Y}$,Alice 和 Bob 想要计算出 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 之间的曼哈顿距离,而不泄露各自私有数据的任何额外信息.

曼哈顿距离的定义.

(1) 二维平面上两点 $\boldsymbol{X}=(x_1,y_1)$ 与 $\boldsymbol{Y}=(x_2,y_2)$ 的曼哈顿距离定义为

$$d=|x_1-x_2|+|y_1-y_2| \tag{3}$$

显然,当$y_1=y_2$时,距离式(3)成为$d=|x_1-x_2|$,因此绝对值也可看成(一维)曼哈顿距离.

(2) 两个 n 维向量 $\boldsymbol{X}=(x_{11},x_{12},\cdots,x_{1n})$ 与 $\boldsymbol{Y}=(x_{21},x_{22},\cdots,x_{2n})$ 的曼哈顿距离定义为

$$d=|x_{11}-x_{21}|+|x_{12}-x_{22}|+\cdots+|x_{1n}-x_{2n}| \tag{4}$$

3.2 绝对值问题的两方安全计算(无全集限制)

计算原理. 假设 Alice 和 Bob 分别具有数据 x 和 y,由于$|x-y|=\sqrt{(x-y)^2}$,因此,将绝对值的保密计算转化为$(x-y)^2$的保密计算. 由于$(x-y)^2=x^2-2xy+y^2$,应用 Paillier 加密方案进行加密运算,根据加法同态性可以得到

$$C=E((x-y)^2)=E(x^2)E(2x)^{-y}E(y^2) \tag{5}$$

其中 $E(2x)^{-y}=[E(2x)^{-1}]^y$,$E(2x)^{-1}$ 为 $E(2x)$ 在乘法群 $Z_{N^2}^*$ 中的逆元.

由于 Paillier 加密方案的明文空间为 Z_N,因此当 x,y 满足$(x-y)^2\in Z_N$时,可保证对式(5)中密文 C 解密后的结果恰为$(x-y)^2$. 如此,如果 $0\le x,y\le\sqrt{N}$,则可保证$(x-y)^2\in Z_N$,进一步保证对式(5)中的密文 C 解密后获得$(x-y)^2$. 在下面设计协议时,我们总假设 Paillier 加密方案中的参数 $N=pq$ 取得充分大,使得 Alice 和 Bob 的数据 x,y(或向量 $\boldsymbol{X},\boldsymbol{Y}$ 的所有分量 $x_j,y_j,j=1,\cdots,n$)满足 $0\le x,y\le\sqrt{N}$(或 $0\le x_j,y_j\le\sqrt{N},j=1,\cdots,n$).

协议 1. 绝对值保密计算协议(无全集限制).

输入:Alice 和 Bob 各自拥有的保密数据 x,y

输出:$f_1(x,y)=f_2(x,y)=|x-y|$

1. Alice 首先利用密钥生成算法生成私钥 sk 和公钥 $pk=(g,N)$.

2. Alice 加密数据 $2x$,并将 $E(2x)$ 以及公钥发送给 Bob.

3. Bob 计算 $C=E(2x)^{-y}E(y^2)$,并将 C 发送给 Alice.

4. Alice 解密 C 得到 $c=D(C)$,并计算

$$b=\begin{cases}\sqrt{c+x^2}, & \text{if } c+x^2<N\\ \sqrt{c+x^2-N}, & \text{if } c+x^2\ge N\end{cases} \tag{6}$$

并将 b 发送给 Bob.

5. 输出 b.

协议 1 的正确性. 根据加密算法的加法同态性,可得

$$C=E(2x)^{-y}E(y^2)=E(y^2-2xy\bmod N),$$

$$c=D(C)=\begin{cases}y^2-2xy, & \text{if } y^2-2xy\ge 0\\ y^2-2xy+N, & \text{if } y^2-2xy<0\end{cases} \tag{7}$$

由式(6)和式(7)可知

(i) 当 $y^2-2xy\ge 0$ 时,$c+x^2<N$,有

$$b=\sqrt{c+x^2}=\sqrt{y^2-2xy+x^2}=|x-y|.$$

(ii) 当 $y^2-2xy<0$ 时,$c+x^2\ge N$,有

$$b=\sqrt{c+x^2-N}=\sqrt{y^2-2xy+x^2}=|x-y|.$$

故协议 1 是正确的.

协议 1 的安全性.

定理 1. 在半诚实模型下,协议 1 是安全的.

证明. 通过构造满足式(1)和式(2)的模拟器 S_1 以及 S_2 严格证明定理 1.

首先构造模拟器 S_1 使得式(1)成立,模拟过程如下:

(i) 接受输入$(x,f_1(x,y))$后,S_1 随机选择 $y'<\sqrt{N}$,满足 $f_1(x,y')=f_1(x,y)$.

(ii) S_1 加密y'^2,并计算

$$C'=E(2x)^{-y'}E(y'^2).$$

(iii) S_1 解密 C' 得到 c',并计算

$$b'=\begin{cases}\sqrt{c'+x^2}, & \text{if } c'+x^2<N\\ \sqrt{c'+x^2-N}, & \text{if } c'+x^2\ge N\end{cases}.$$

由于在协议执行中

$$view_1^{\pi}(x,y)=(x,C,f_1(x,y)).$$

令 $S_1(x,f_1(x,y))=(x,C',f_1(x,y'))$. 由于Paillier加密方案是语义安全的,因此 $C=E(2x)^{-y}E(y^2)$ 与 $C'=E(2x)^{-y'}E(y'^2)$ 是计算不可区分的. 又 $f_1(x,y)=f_1(x,y')$,因此

$$\{S_1(x,f_1(x,y))\}_{x,y\le\sqrt{N}}\overset{c}{\equiv}\{view_1^{\pi}(x,y)\}_{x,y\le\sqrt{N}}.$$

类似地,可构造模拟器 S_2,模拟过程如下.

(i) 接受输入$(y,f_2(x,y))$后,S_2 任意选择 $x'<\sqrt{N}$,使得 $f_2(x',y)=f_2(x,y)$.

(ii) S_2 加密 $2x'$,得到 $E(2x')$.

(iii) S_2 计算 $C^*=E(2x')^{-y}E(y^2)$.

(iv) S_2 解密 C^* 得到 c^*,并计算

$$b^*=\begin{cases}\sqrt{c^*+x'^2}, & \text{if } c^*+x'^2<N\\ \sqrt{c^*+x'^2-N}, & \text{if } c^*+x'^2\ge N\end{cases}.$$

由于在协议执行中

$$view_2^{\pi}(x,y)=(y,E(2x),f_2(x,y)).$$

令 $S_2(y,f_2(x,y))=(y,E(2x'),f_2(x',y))$. 根据 Paillier 加密方案的语义安全性,任何两个密文都是计算不可区分的,对 Bob 来说,$E(2x)\overset{c}{\equiv}E(2x')$. 又因为 $f_2(x,y)=f_2(x',y)$. 因此

$$\{S_2(y,f_2(x,y))\}_{x,y\leq\sqrt{N}}\overset{c}{\equiv}\{view_2^{\pi}(x,y)\}_{x,y\leq\sqrt{N}}.$$

因此,协议 1 是安全的. 证毕.

3.3 曼哈顿距离的两方安全计算(无全集限制)

以协议 1 为基础,构造曼哈顿距离的保密计算协议.

协议 2. 曼哈顿距离的保密计算协议(无全集限制).

输入:Alice 和 Bob 分别输入向量 $\boldsymbol{X}=(x_1,\cdots,x_n)$,$\boldsymbol{Y}=(y_1,\cdots,y_n)$

输出:$\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的曼哈顿距离 $f_1(\boldsymbol{X},\boldsymbol{Y})=f_2(\boldsymbol{X},\boldsymbol{Y})=\sum_{j=1}^{n}|x_j-y_j|$

1. Alice 首先利用密钥生成算法生成私钥 sk 和公钥 $pk=(g,N)$.

2. 对于每一个 $j\in[1,n]$,Alice 加密数据 x_j^2,$2x_j$,并将 $E(x_j^2)$,$E(2x_j)$ 以及公钥发送给 Bob.

3. 对于每一个 $j\in[1,n]$,Bob 计算

$$C_j=E(x_j^2)E(2x_j)^{-y_j}E(y_j^2),$$

随机置换$(C_1,\cdots,C_n)$,得到$(C_1',\cdots,C_n')$发送给 Alice.

4. Alice 对 C_j',$j\in[1,n]$进行解密得到 $t_j=D(C_j')$,并进一步计算 $t=\sqrt{t_1}+\cdots+\sqrt{t_n}$,并将 t 告诉 Bob.

5. 输出 t.

协议 2 的正确性. 协议 2 本质上是对于每一个 $j\in[1,n]$,Alice 和 Bob 分别以 x_j 和 y_j 为输入,根据计算原理保密计算得到 $|x_j-y_j|$. 由于在协议第 3 步进行了随机置换,我们仍然有 $\{|x_1-y_1|,\cdots,|x_n-y_n|\}=\{\sqrt{t_1},\cdots,\sqrt{t_n}\}$. 因此

$$t=\sqrt{t_1}+\cdots+\sqrt{t_n}=|x_1-y_1|+\cdots+|x_n-y_n|,$$

故所得 t 值即为 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的曼哈顿距离. 协议 2 是正确的.

协议 2 的安全性. 协议 2 与协议 1 非常类似,Alice 所做的本质上是对每一个 $j\in[1,n]$,以 x_j 为输入执行了与协议 1 类似的操作,Alice 仅发送给 Bob 一些密文数组以及最后的解密结果 t,由于 Bob 没有解密密钥,因此从收到的密文数组得不到 Alice 数据的任何信息,最后所收到的 t 值是协议所规定的输出,因此在协议 2 中 Alice 的数据是安全的.

Bob 在协议 2 中所做的基本是对于每一个 $j\in[1,n]$,以 y_j 为输入执行了与协议 1 类似的密文乘积运算获得了密文 $C_j=E[(x_j-y_j)^2]$. 由于在协议中 Bob 将密文随机置换后才发给 Alice,因此 Alice 解密后无法确定 $|x_j-y_j|$ 与集合 $\{\sqrt{t_1},\cdots,\sqrt{t_n}\}$ 中哪个元素对应,也就无法获得 Bob 数据的任何额外信息,因此 Bob 的隐私数据也是安全的(如果 Alice 知道某个 $|x_j-y_j|$ 的具体取值为 $\sqrt{t_l}$,Alice 则可根据 $|x_j-y_j|=\sqrt{t_l}$ 推算出 $y_j=x_j+\sqrt{t_l}$ 或 $y_j=x_j-\sqrt{t_l}$,这是曼哈顿距离保密计算协议不容许泄露的额外信息,因此协议 2 第 3 步中 Bob 通过随机置换保证协议的安全性). 我们有下面的定理,定理 2 的证明与定理 1 类似,在此省略.

定理 2. 在半诚实模型下,协议 2 是安全的.

4 曼哈顿距离保密计算(有全集限制)

在协议 1 和协议 2 中所设计的绝对值及曼哈顿距离的保密计算方案仅要求 Alice 和 Bob 的数据 x,y(或向量 $\boldsymbol{X},\boldsymbol{Y}$ 的所有分量 $x_j,y_j,j\in[1,n]$)满足条件 $0\leq x,y\leq\sqrt{N}$(或 $0\leq x_j,y_j\leq\sqrt{N}$),由于 Paillier 加密算法中的 N 可以取得充分大,上述条件容易满足,因此认为协议 1 和协议 2 对于参与者数据基本没有限制,其适用范围广泛. 我们也注意到,在协议 2 中随着向量维数 n 的增加,协议 2 的在线计算复杂性线性增加,因此我们需要进一步考虑在某些特殊情况下能否对上面两个协议进行优化. 本节中我们考虑如果能事先确定参与者数据在某个全集内取值,则可利用编码方法设计绝对值及曼哈顿距离保密计算方案,可以大大降低协议的在线计算复杂性.

本节中假设所讨论点的坐标或向量的分量都在一个给定的全集 U 中取值,其中 $U=\{a_1,a_2,\cdots,a_m\}$,满足 $a_i=a_{i-1}+1,i=2,\cdots,m$.

4.1 绝对值问题的两方安全计算(有全集限制)

计算原理. 对于任意 $u\in U$,定义 u 在 U 中的序号为 u_{ind},即如果 $u=a_k$,则 $u_{\text{ind}}=k$. 对于 $u\in U$,根据如下方式构造一个数组 T_u 与 u 相对应:

$$T_u=(u_1,\cdots,u_m) \tag{8}$$

其中当 $1\leq i\leq u_{\text{ind}}$ 时,$u_i=0$;当 $u_{\text{ind}}<i\leq m$ 时,$u_i=1$.

首先证明下面命题.

命题 1. 根据式(8)分别构造 u 以及 $w=a_m-u_{\text{ind}}$ 对应的数组:$T_u=(u_1,\cdots,u_m)$ 以及 $T_w=(w_1,\cdots,w_m)$,则有下面结论:

(i) 当 $u>v$ 时,有

$$\sum_{i=1}^{v_{\text{ind}}} u_i = 0 \tag{9}$$

(ii) 当 $u \le v$ 时,有

$$\sum_{i=1}^{v_{\text{ind}}} u_i = |u - v| \tag{10}$$

(iii) 对任意 $u, v \in [a_1, a_m]$,式(11)总成立

$$|u - v| = \sum_{i=1}^{v_{\text{ind}}} u_i + \sum_{i=1}^{(a_m - v_{\text{ind}})_{\text{ind}}} w_i \tag{11}$$

证明. (i) 由于当 $u > v$ 时,有 $u_{\text{ind}} > v_{\text{ind}}$,根据数组 T_u 的构造方式,可知此时 T_u 的前 v_{ind} 个元素全为零,因此式(9)成立.

(ii) 由于当 $u \le v$ 时,有 $u_{\text{ind}} \le v_{\text{ind}}$,并由全集 U 的定义,可知在此情形下,

$$|u - v| = v - u = v_{\text{ind}} - u_{\text{ind}}.$$

根据数组 T_u 的构造方式,T_u 中当 $i \le u_{\text{ind}}$ 时,$u_i = 0$;当 $i \in [u_{\text{ind}} + 1, v_{\text{ind}}]$ 时,$u_i = 1$. 因此 T_u 的前 v_{ind} 个元素之和为 $v_{\text{ind}} - u_{\text{ind}}$,即为 $|u - v|$,因此式(10)成立.

(iii) 根据(i)和(ii)的结果,我们可知:

(a) 假设 $u > v$,那么 $u_{\text{ind}} > v_{\text{ind}}$,这时有 $a_m - u_{\text{ind}} < a_m - v_{\text{ind}}$. 由(i)的结果可知,式(11)右端第一个和式的值为 0;在(ii)中,用 $a_m - u_{\text{ind}}$, $a_m - v_{\text{ind}}$ 分别代替 u, v,式(11)右端第 2 个和式的值为

$$|(a_m - u_{\text{ind}}) - (a_m - v_{\text{ind}})| = |v_{\text{ind}} - u_{\text{ind}}| = |u - v|,$$

因此式(11)成立.

(b) 假设 $u < v$,那么 $u_{\text{ind}} < v_{\text{ind}}$,这时有 $a_m - u_{\text{ind}} > a_m - v_{\text{ind}}$. 由(ii)的结果可知,式(11)右端第一个和式的值为 $|u - v|$;以 $a_m - u_{\text{ind}}$, $a_m - v_{\text{ind}}$ 分别代替(i)中的数据 u, v,由(i)的结果可知,式(11)右端第 2 个和式的值为零. 因此式(11)成立.

(c) 假设 $u = v$,那么 $u_{\text{ind}} = v_{\text{ind}}$,这时也有 $a_m - u_{\text{ind}} = a_m - v_{\text{ind}}$. 由(ii)的结果可知,式(11)右端第 1 个和式的值为 $|u - v| = 0$;以 $a_m - u_{\text{ind}}$, $a_m - v_{\text{ind}}$ 分别代替(ii)中的数据 u, v,由(ii)的结果知,式(11)右端第 2 个和式的值为

$$|(a_m - u_{\text{ind}}) - (a_m - v_{\text{ind}})| = 0.$$

在此情形下,式(11)左右两端都为零,因此式(11)成立. 证毕.

例如,假设 $u = -3$, $v = 2$,并设全集为 $U = \{-5, -4, \cdots, 5\}$,那么有 $u_{\text{ind}} = 3$, $v_{\text{ind}} = 8$,根据式(8)分别构造 u 以及 $w = a_m - u_{\text{ind}}$ 对应的数组:

$$T_x = (u_1, \cdots, u_{11}) = (0,0,0,1,1,1,1,1,1,1,1),$$

$$T_w = (w_1, \cdots, w_{11}) = (0,0,0,0,0,0,0,0,1,1,1).$$

这时有

$$|u - v| = \sum_{i=1}^{v_{\text{ind}}} u_i + \sum_{i=1}^{(a_m - v_{\text{ind}})_{\text{ind}}} w_i = 5 + 0 = 5,$$

因此式(11)成立.

注解 2. 对于任意正整数 s,只要全集 U 中的元素满足 $a_i - a_{i-1} = s$,上述计算原理及命题 1 仍可应用. 只是在此情形下,式(10)和(11)应相应修改为

$$s \sum_{i=1}^{v_{\text{ind}}} u_i = |u - v|,$$

$$|u - v| = s\left[\sum_{i=1}^{v_{\text{ind}}} u_i + \sum_{i=1}^{(a_m - v_{\text{ind}})_{\text{ind}}} w_i\right].$$

协议设计. 命题 1 给出了在有全集限制条件下绝对值问题的计算原理. 下面将以上述计算原理为基础,结合应用 Paillier 同态加密系统设计构造保密计算协议.

协议 3. 绝对值保密计算协议(有全集限制).

输入:Alice 的私密数据 x,Bob 的私密数据 y,以及全集 $U = \{a_1, \cdots, a_m\}$

输出:$f_1(x, y) = f_2(x, y) = |x - y|$

1. Alice 运行 $G(\tau)$ 生成 Paillier 加密算法的私钥 sk 和公钥 $pk = (g, N)$.

2. (a) Alice 按照式(8)将数据 x 和 $z = a_m - x_{\text{ind}}$ 分别转化为 m 维数组 T_x 和 T_z,得到

$$T_x = (x_1, \cdots, x_m),\ T_z = (z_1, \cdots, z_m).$$

(b) Alice 加密 T_x, T_z,得到数组:

$$H_x = (E(x_1), \cdots, E(x_m)) := (X_1, \cdots, X_m),$$

$$H_z = (E(z_1), \cdots, E(z_m)) := (Z_1, \cdots, Z_m),$$

并将 H_x, H_z 和公钥发送给 Bob.

3. Bob 加密 0,得到 $E(0)$,并计算

$$H = \left[\prod_{i=1}^{y_{\text{ind}}} X_i \prod_{k=1}^{(a_m - y_{\text{ind}})_{\text{ind}}} Z_k\right] E(0) \bmod N^2.$$

Bob 将 H 发送给 Alice.

4. Alice 解密 H,得到 $h = D(H)$,并将 h 告诉 Bob.

5. 输出 h.

协议 3 的正确性. 根据加密算法的加法同态性,

$$\begin{aligned} H &= \left[\prod_{i=1}^{y_{\text{ind}}} X_i \prod_{k=1}^{(a_m - y_{\text{ind}})_{\text{ind}}} Z_k\right] E(0) \bmod N^2 \\ &= \left[\prod_{i=1}^{y_{\text{ind}}} E(x_i) \prod_{k=1}^{(a_m - y_{\text{ind}})_{\text{ind}}} E(z_i)\right] E(0) \bmod N^2 \\ &= E\left(\sum_{i=1}^{y_{\text{ind}}} x_i + \sum_{k=1}^{(a_m - y_{\text{ind}})_{\text{ind}}} z_k\right) \bmod N^2. \end{aligned}$$

因此解密后 $h = D(H) = \sum_{i=1}^{y_{\text{ind}}} x_i + \sum_{k=1}^{(a_m - y_{\text{ind}})_{\text{ind}}} z_k$. 由命题 1(iii),可知 $f(x, y) = h$. 因此协议 3 是正确的.

协议 3 的安全性.

定理 3. 在半诚实模型下,协议 3 是安全的.

分析. 在协议 3 中,只有 Alice 有私钥可以解密,Bob 在协议中仅获得 Alice 发送的密文 H_x, H_z,因此 Bob 从协议中无法得到 Alice 隐私数据的任何额外信息.

另一方面,Bob 在计算密文 H 时乘了一个 $E(0)$,由于 $E(0)$ 是 Bob 加密的,解密前,对 Alice 来说 H 与随机数不可区分. Alice 在解密后仅得到结果 $h=f(x,y)$. 因此 Alice 也无法获得任何额外信息.

证明. 下面通过构造满足式(1)和式(2)的模拟器 S_1 和 S_2 严格证明定理 3.

首先构造模拟器 S_1,模拟过程如下:

(i) 接受输入 $(x, f_1(x,y))$ 后,S_1 任意选择 $y' \in U$,满足 $f_1(x,y')=f_1(x,y)$.

(ii) S_1 加密 0 得到 $E'(0)$,进一步计算

$$H' = \Big[\prod_{i=1}^{y'_{\text{ind}}} X_i \prod_{k=1}^{(a_m-y'_{\text{ind}})_{\text{ind}}} Z_i\Big]E'(0) \bmod N^2.$$

(iii) S_1 解密 H',得到 h'.

由于在协议执行中

$$view_1^{\pi}(x,y) = (x, H, f_1(x,y)),$$

令 $S_1(x, f_1(x,y)) = (x, H', f_1(x,y'))$. 由 Paillier 加密系统的语义安全性,任何两个密文都是计算不可区分的,由于 $E(0)$ 是 Bob 加密的密文,虽然 Alice 能够解密,对 Alice 来说,$E(0) \overset{c}{\equiv} E'(0)$,进一步可知 H 与 H' 是计算上不可区分的. 又因为 $f_1(x,y')=f_1(x,y)$. 因此

$$\{S_1(x, f_1(x,y))\}_{x,y\in U} \overset{c}{\equiv} \{view_1^{\pi}(x,y)\}_{x,y\in U}.$$

类似地,可构造模拟器 S_2,模拟过程如下:

(i) 接受输入 $(y, f_2(x,y))$ 后,S_2 任意选择 $x' \in U$,满足 $f_2(x',y)=f_2(x,y)$.

(ii) S_2 根据方式(4)构造 $x', z'=a_m-x'_{\text{ind}}$ 对应的数组 $T_{x'}=(x'_1,\cdots,x'_m)$, $T_{z'}=(z'_1,\cdots,z'_m)$,并对其进行加密,得到

$$H_{x'}=(E(x'_1),\cdots,E(x'_m)):=(X'_1,\cdots,X'_m),$$

$$H_{z'}=(E(z'_1),\cdots,E(z'_m)):=(Z'_1,\cdots,Z'_m).$$

(iii) S_2 计算

$$H^* = \Big[\prod_{i=1}^{y_{\text{ind}}} X'_i \prod_{k=1}^{(a_m-y_{\text{ind}})_{\text{ind}}} Z'_k\Big]E(0) \bmod N^2.$$

(iv) S_2 解密 H^*,得到 h^*.

由于在协议执行中

$$view_2^{\pi}(x,y) = (x_2, H_x, H_z, f_2(x,y)),$$

令 $S_2(y, f_2(x,y)) = (y, H_{x'}, H_{z'}, f_2(x',y))$,根据 Paillier 加密系统的语义安全性,任何两个密文都是计算不可区分的,对 Bob 来说,$H_x \overset{c}{\equiv} H_{x'}$, $H_z \overset{c}{\equiv} H_{z'}$,又因为 $f_2(x',y)=f_2(x,y)$. 因此

$$\{S_2(y, f_2(x,y))\}_{x,y\in U} \overset{c}{\equiv} \{view_2^{\pi}(x,y)\}_{x,y\in U}.$$

综上所述,协议 3 在半诚实模型下是安全的. 证毕.

4.2 曼哈顿距离的两方安全计算(有全集限制)

以绝对值保密计算协议 3 为基础,本小节研究设计曼哈顿距离的保密计算协议. 计算的基本思想是首先保密计算对应坐标之差的绝对值的密文,其次利用加密算法的加法同态性,将这些密文相乘,最后进行解密即可得到两点的曼哈顿距离. 具体协议如下:

协议 4. 曼哈顿距离的保密计算协议(有全集限制).

输入:Alice 和 Bob 的向量 $\boldsymbol{X}=(x_1,\cdots,x_n)$, $\boldsymbol{Y}=(y_1,\cdots,y_n)$,其中 $x_j, y_j \in U, j=1,\cdots,n$

输出:两点 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 之间的曼哈顿距离 $f_1(\boldsymbol{X},\boldsymbol{Y})=f_2(\boldsymbol{X},\boldsymbol{Y})=\sum_{j=1}^{n}|x_j-y_j|$

1. Alice 运行 $G(\tau)$ 生成 Paillier 加密算法的私钥 sk 和公钥 $pk=(g,N)$.

2. 对于每一个 $j\in[1,n]$,Alice 按照式(8)构造 x_j 以及 $z_j=a_m-(x_j)_{\text{ind}}$ 对应的数组

$$T_{x_j}=(x_{j1},\cdots,x_{jm}), T_{z_j}=(z_{j1},\cdots,z_{jm}),$$

并对其进行加密,得到数组:

$$H_{x_j}=(E(x_{j1}),\cdots,E(x_{jm})):=(X_{j1},\cdots,X_{jm}),$$

$$H_{z_j}=(E(z_{j1}),\cdots,E(z_{jm})):=(Z_{j1},\cdots,Z_{jm}),$$

将 (H_{x_j}, H_{z_j}), $j\in[1,n]$ 和公钥发送给 Bob.

3. Bob 加密 0,得到 $E(0)$,并计算

$$\hat{H} = \prod_{j=1}^{n}\Big[\prod_{i=1}^{(y_j)_{\text{ind}}} X_{ji} \prod_{k=1}^{(a_m-(y_j)_{\text{ind}})_{\text{ind}}} Z_{jk}\Big]E(0) \bmod N^2.$$

Bob 将 $\hat{H}$ 发送给 Alice.

4. Alice 解密 $\hat{H}$,得到 $\hat{h}=D(\hat{H})$,并将 $\hat{h}$ 告诉 Bob.

5. 输出 $\hat{h}$.

协议 4 的正确性. 由协议 3 与加密算法的加法同态性,可得

$$\begin{aligned}\hat{H} &= E(|x_1-y_1|)\cdot E(|x_2-y_2|)\cdots E(|x_n-y_n|)\\ &= E(|x_1-y_1|+|x_2-y_2|+\cdots+|x_n-y_n|),\end{aligned}$$

因此 $\hat{h}=D(\hat{H})=\sum_{j=1}^{n}|x_j-y_j|$,故协议 4 是正确的.

协议 4 的安全性. 协议 4 与协议 3 非常类似,Alice 所做的实际是对每一个 $j\in[1,n]$,以 x_j 为输入执行了与协议 3 完全相同的操作,因此 Alice 的数据是安全的. Bob 在协议 4 中所做的是对于每一个 $j\in[1,n]$,以 y_j 为输入执行了与协议 3 类似的密文乘积运算,以得到绝对值的密文. 把这些(关于所有 $j\in[1,n]$)乘积密文再相乘得到 $\hat{H}$,最后由 Alice 解密 $\hat{H}$ 得出所需结果,这样做的目的是为避免因分别解密而造成各个分量差的绝对值 $|x_j-y_j|$ 泄露

(在曼哈顿距离保密计算中各个分量差的绝对值不允许泄露).由于 Bob 在计算乘积 $\hat{H}$ 时,乘了一个自己加密的密文 $E(0)$,虽然 Alice 有密钥,但对于 Alice 来说,解密前 $\hat{H}$ 与随机数不可区分,解密后也仅得到了协议规定获得的输出结果,因此 Bob 的隐私数据也是安全的.我们有下面的定理,定理 4 的严格证明在此省略.

定理 4.　在半诚实模型下,协议 4 是安全的.

5　协议的效率分析和实验测试

5.1　协议效率分析

本文都是应用 Paillier 公钥加密系统解决问题的,因此计算复杂性以开销较大的模指数运算次数作为衡量标准,而忽略各协议中所需要的模乘运算.我们应用协议中需要的总通信次数来衡量通信复杂性.需要全集时,假设全集的势为 m,Alice 和 Bob 具有的向量维数为 n.应用 Paillier 加密方案加密及解密一次分别需要 1 次及 2 次模指数运算(参看注解 1).

在协议 1 中,Alice 需要加密 1 次,解密 1 次.其次,Bob 进行了 1 次模指数运算.因此协议 1 共需要进行 4 次模指数运算,3 次通信.

在协议 2 中,Alice 需要加密 $2n$ 次,解密 n 次.其次,Bob 需要进行 n 次模指数运算,因此协议 2 共需要进行模指数运算 $5n$ 次,3 次通信.

在协议 3 中,Alice 首先加密两个数组,需要 $2m$ 次加密.其次,Bob 需要加密 1 次,Alice 需要解密 1 次,因此协议 3 共需要 $2m+3$ 次模指数运算,Alice 所做的 $2m$ 次加密运算均是加密 0 与 1,可以借助云服务器事先完成.这样的话,在线复杂性就可以大大减少了,因此协议 3 仅需要 3 次模指数运算,3 次通信.

在协议 4 中,Alice 首先加密 $2n$ 个 m 维数组,这需要 $2mn$ 次加密,其次,在协议 4 中 Bob 需要加密 1 次,Alice 需要解密 1 次,因此协议 4 共需要进行 $2mn+3$ 次模指数运算,Alice 加密 $2n$ 个 m 维数组均为加密 0 与 1,可以借助云服务器事先完成.因此协议 4 仅需要 3 次模指数运算,3 次通信.具体分析结果参看表 1.

表 1　本文协议效率分析结果

参数	协议 1	协议 2	协议 3(在线)	协议 4(在线)
计算复杂性	4	$5n$	$2m+3(3)$	$2mn+3(3)$
通信复杂性	3	3	3	3

由表 1 可知,本文 4 个协议的通信复杂性都相同,协议 3 和协议 4 的计算复杂性与全集的势 m 有关,但协议 3 和协议 4 中 Alice 加密的数组元素都是 0 与 1,这部分运算可以利用云服务器事先进行计算,如此可知协议 3 和协议 4 的在线计算复杂性均为 3 次模指数运算(即为表中括号内所列数据).

5.2　实验数据分析

为了进一步分析协议的效率和方案的实际可行性,本部分通过模拟实验来测试本文协议在具体执行中所用的时间.

实验测试环境. 本文实验应用的是 Windows 7 旗舰版 32 位操作系统,处理器是 Intel(R) Core(TM) i3-2100 CPU@3.10 GHz,内存是 4.00 GB,并应用 java 语言在 MyEclipse 上运行实现.

实验方法. 实验设定全集 $m=10$,向量维数分别为 $n=1,2,\cdots,15$(当 $n=1$ 时,即向量维数为 1 时为绝对值协议的执行时间;当 $n>1$ 时,为曼哈顿距离协议的执行时间).实验中没有考虑预处理需要的时间.在模拟协议 3 与协议 4 时,我们借助云服务器事先生成了若干个 0 与 1 的密文,以减少这两个协议的在线复杂性.为获得较准确的实验结果,我们对向量维数 n 的每个值分别进行了 1000 次模拟测试,计算出所需时间的平均值.

本文协议的具体执行时间见表 2(这里忽略了云服务器加密所需的时间,执行时间以毫秒为单位).以表 2 为基础,绘制出的图 1 直观地表明了向量维数增长对执行时间产生的影响.其中,将协议 1 与协议 3 分别看作是协议 2 与协议 4 在向量维数为 1 时的特殊情况.由图 1 可知,在利用云服务器的情况下,保密计算绝对值的协议 1 和协议 3 的执行时间相差不大.保密计算曼哈顿距离的协议 2 和协议 4 的执行时间随向量维数增加均呈线性增长,但协议 4 的曲线增长缓慢,并且容易看到协议 4 的执行效率要明显高于协议 2 的执行效率.即如果能够事先确定所有参与计算的数据属于某个全集,并可

表 2　本文协议具体实验测试执行时间

向量维数	执行时间/ms	
	协议 1 与协议 2	协议 3 与协议 4
1	29	13
2	72	13
4	145	14
6	217	14
8	292	15
10	371	15
12	430	15
14	503	16

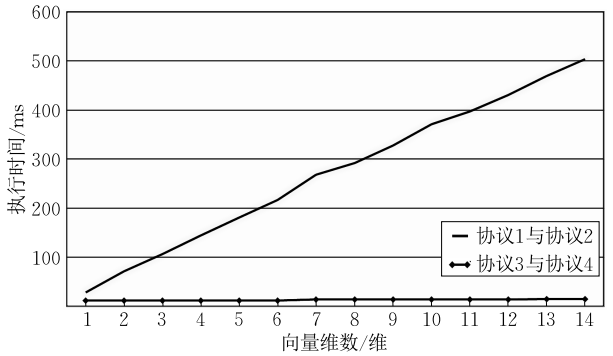

图 1 协议的执行时间随向量维数增长的变化规律

以利用云服务器进行预备运算,应用协议 4 进行计算可大大降低协议的在线计算复杂性.这和我们的效率分析结果也是一致的.

6 本文协议的推广应用

第 6 节将给出一些应用实例以说明绝对值以及曼哈顿距离的保密计算在解决实际问题中的应用.

6.1 点与区间关系的保密判定

假设 Alice 拥有一个保密整数 x,Bob 拥有一个保密区间$[y_1,y_2]$(y_1,y_2均为整数),Alice 和 Bob 希望合作保密判定数 x 是否属于区间$[y_1,y_2]$.参看图 2.

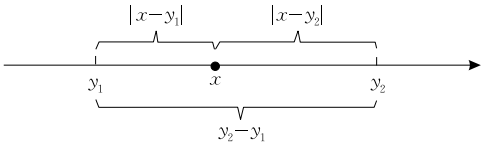

图 2 点与区间的关系

保密判断 x 是否属于区间$[y_1,y_2]$当且仅当式(12)成立:

$$|x-y_1|+|x-y_2|=y_2-y_1 \tag{12}$$

根据式(12)即可将保密判定 x 是否属于区间$[y_1,y_2]$的问题转化为保密计算点 $X=(x,x)$与 $Y=(y_1,y_2)$的曼哈顿距离是否等于区间长度 y_2-y_1 的问题.

下面分别在无全集限制以及具有全集限制条件下设计保密判定协议.

(I)假设 Alice 的点 x 和 Bob 的区间端点 y_1,y_2 均为正整数(不需要全集).以协议 2 为基础设计协议如协议 5.

协议 5. 点与区间关系保密判定协议(无全集限制).

输入:Alice 和 Bob 分别输入向量 $\boldsymbol{X}=(x_1,x_2)=(x,x)$,$\boldsymbol{Y}=(y_1,y_2)$

输出:x 是否属于区间$[y_1,y_2]$

1. Alice 和 Bob 分别以 $\boldsymbol{X}=(x_1,x_2)=(x,x)$,$\boldsymbol{Y}=(y_1,y_2)$为输入执行协议 2 的前 2 步,Bob得到密文$E(x_1^2)$,$E(2x_1)$,$E(x_2^2)$,$E(2x_2)$.

2. 对于 $j=1,2$,Bob 选择随机数 r,计算

$$C_j=[E(x_j^2)E(2x_j)^{-y_j}E(y_j^2)]^{r^2},$$

Bob 随机置换(C_1,C_2),得到(C_1',C_2')发送给 Alice.

3. Alice 对 C_j',$j=1,2$ 进行解密得到 $t_j=D(C_j')$,任意选择数 s,计算 $E(t)=E(s(\sqrt{t_1}+\sqrt{t_2}))$,$E(s)$,并发送给 Bob.

4. Bob 计算 $G=E(t)(E(s)^{-1})^{r(y_2-y_1)}$,并将 G 发送给 Alice.

5. Alice 解密 G,如果 $D(G)=0$,则输出 $x\in[y_1,y_2]$;否则,输出 $x\notin[y_1,y_2]$.

协议 5 的正确性. Alice 和 Bob 进行类似于协议 2 的计算:

$$C_j=(E(x_j^2)E(2x_j)^{-y_j}E(y_j^2))^{r^2}=E(r^2(x_j-y_j)^2).$$

由于在协议第 2 步进行了随机置换,故有$\{r|x_1-y_1|,r|x_2-y_2|\}=\{\sqrt{t_1},\sqrt{t_2}\}$.因此

$$t=s(\sqrt{t_1}+\sqrt{t_2})=sr(|x_1-y_1|+|x_2-y_2|),$$

又因为

$$\begin{aligned}G&=E(t)(E(s)^{-1})^{r(y_2-y_1)}\\&=E(rs(|x_1-y_1|+|x_2-y_2|-(y_2-y_1))),\end{aligned}$$

因此 $D(G)=rs(|x_1-y_1|+|x_2-y_2|-(y_2-y_1))$.根据式(12),如果 $D(G)=0$,则有 $x\in[y_1,y_2]$,否则 $x\notin[y_1,y_2]$.因此协议 5 是正确的.

协议 5 的安全性. 协议 5 与协议 2 非常类似,Alice 的行为本质上是以 x_1,x_2 为输入执行与协议 2 类似的操作,Alice 仅发送给 Bob 一些密文数组以及最后的输出结果.由于 Bob 没有解密密钥,因此在协议 5 中 Alice 的数据是安全的.

Bob 在协议 5 中所做的基本是以 y_1,y_2 为输入执行了与协议 1 类似的密文乘积运算获得了密文 C_1,C_2,由于密文 C_1,C_2 中包含有 Bob 的随机数 r,对于 Alice 来说,C_1,C_2 与随机数是计算上不可区分的,因此并不会泄露$|x_1-y_1|+|x_2-y_2|$.

在协议第 2 步 Bob 对C_1,C_2 置换后发送给 Alice,解密后 Alice 只获得明文值 t_1,t_2,无法确定 $r|x_1-y_1|$(或 $r|x_2-y_2|$)与集合$\{\sqrt{t_1},\sqrt{t_2}\}$中哪个元素相对应,因此也无法获得 Bob 数据的任何额外信息.我

们有下面的定理 5,定理证明在此省略.

定理 5.　在半诚实模型下,协议 5 是安全的.

(II) 假设 x,y_1,y_2 均属于给定全集 $U=\{a_1,\cdots,a_m\}$. 以协议 4 为基础设计协议如协议 6.

协议 6.　点与区间关系保密判定协议(有全集限制).

输入:Alice 和Bob分别输入向量 $\boldsymbol{X}=(x_1,x_2)=(x,x)$, $\boldsymbol{Y}=(y_1,y_2)$

输出:x 是否属于区间 $[y_1,y_2]$

1. Alice 和 Bob 执行协议 4 的前 3 步,Bob 得到密文 $H_{x_1},H_{z_1},H_{x_2},H_{z_2}$.

2. Bob 选择随机数 r,计算密文:

$$C_1=\prod_{j=1}^{2}\Big[\prod_{i=1}^{(y_j)_{\text{ind}}}X_{ji}\prod_{k=1}^{(a_m-y_j)_{\text{ind}}}Z_{jk}\Big]\bmod N^2,$$

$$C_2=E(y_2-y_1),C=(C_1C_2^{-1})^r\bmod N^2.$$

Bob 将 C 发送给 Alice.

3. Alice 解密 C,如果 $D(C)=0$,则输出 $x\in[y_1,y_2]$;否则,输出 $x\notin[y_1,y_2]$.

协议 6 的正确性. 由协议 4 与算法的加法同态性,可得

$$\begin{aligned}C&=(C_1C_2^{-1})^r\bmod N^2\\&=(E(|x-y_1|+|x-y_2|)E(y_2-y_1)^{-1})^r\\&=E(r(|x-y_1|+|x-y_2|-(y_2-y_1))),\end{aligned}$$

因此 $D(C)=r(|x-y_1|+|x-y_2|-(y_2-y_1))$. 根据式(12),如果 $D(C)=0$,则有 $x\in[y_1,y_2]$,否则 $x\notin[y_1,y_2]$. 故协议 6 是正确的.

协议 6 的安全性. 协议 6 与协议 4 非常类似,Alice 所做的实际是以 $X=(x_1,x_2)=(x,x)$ 为输入执行了与协议 4 几乎相同的操作,因此 Alice 的数据是安全的. Bob 是根据 Alice 发送的密文直接计算得到密文 C. 由于 C 中包含密文 C_2 以及随机数 r,对于 Alice 来说,解密前 C 与随机数不可区分. 解密后 Alice 仅可得到协议规定获得的输出结果,而无法获得 Bob 数据的任何其他额外信息,因此 Bob 的隐私数据也是安全的. 我们有下面的定理 6,定理证明在此省略.

定理 6.　在半诚实模型下,协议 6 是安全的.

6.2　点与矩形及长方体的关系问题

6.1 节中,点与区间关系保密判定问题以及计算方案还可以进一步推广到平面(或空间)上的点是否属于矩形(或长方体)区域的保密判定问题. 即假设 Alice 拥有一个保密点 $P(x,y)$(或 $P(x,y,z)$),Bob 拥有一个保密的矩形 $ABCD$(或长方体 $ABCD-A'B'C'D'$),Alice 和 Bob 希望合作保密判定点 P 是否属于矩形(或长方体). 如图 3 所示.

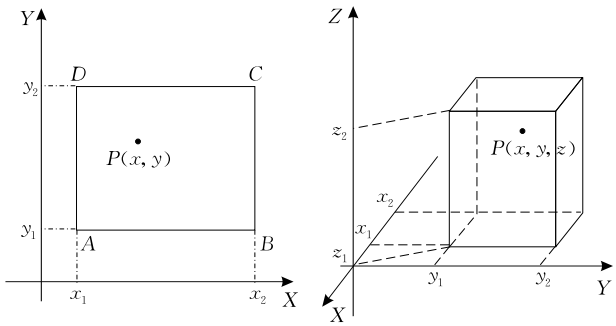

图 3　点与矩形以及长方体的关系

首先容易证明下面结论.

结论 1.　$P(x,y)$ 属于矩形 $ABCD$:$[x_1,x_2]\times[y_1,y_2]$ 当且仅当下式成立:

$$\begin{aligned}&|x-x_1|+|x-x_2|+|y-y_1|+|y-y_2|=\\&\qquad|x_1-x_2|+|y_1-y_2|\end{aligned}\tag{13}$$

结论 2.　$P(x,y,z)$ 属于长方体 $ABCD-A'B'C'D'$:$[x_1,x_2]\times[y_1,y_2]\times[z_1,z_2]$ 当且仅当式(14)成立:

$$\begin{aligned}&|x-x_1|+|x-x_2|+|y-y_1|+|y-y_2|+\\&|z-z_1|+|z-z_2|=\\&\qquad|x_1-x_2|+|y_1-y_2|+|z_1-z_2|\end{aligned}\tag{14}$$

证明.　(a) 首先易知 $P(x,y)$ 属于矩形 $ABCD$:$[x_1,x_2]\times[y_1,y_2]$ 当且仅当 $x\in[x_1,x_2]$ 以及 $y\in[y_1,y_2]$ 同时成立;

(b) 由图 2 可知,$x\in[x_1,x_2]$ 当且仅当 $|x-x_1|+|x-x_2|=|x_1-x_2|$ 成立;$y\in[y_1,y_2]$ 当且仅当 $|y-y_1|+|y-y_2|=|y_1-y_2|$ 成立.

因此,$x\in[x_1,x_2]$ 以及 $y\in[y_1,y_2]$ 同时成立,当且仅当 $|x-x_1|+|x-x_2|=|x_1-x_2|$,$|y-y_1|+|y-y_2|=|y_1-y_2|$ 同时成立.

(c) 显然如果 $|x-x_1|+|x-x_2|=|x_1-x_2|$,$|y-y_1|+|y-y_2|=|y_1-y_2|$ 同时成立,则一定有式(13)成立.

反之,如果 $|x-x_1|+|x-x_2|=|x_1-x_2|$,$|y-y_1|+|y-y_2|=|y_1-y_2|$ 不同时成立,这时由于

$$|x-x_1|+|x-x_2|\geq|x_1-x_2|,$$
$$|y-y_1|+|y-y_2|\geq|y_1-y_2|,$$

即有

$$\begin{aligned}&|x-x_1|+|x-x_2|+|y-y_1|+|y-y_2|>\\&\qquad|x_1-x_2|+|y_1-y_2|,\end{aligned}$$

因此式(13)不成立.

联合(a)~(c),即证明了结论 1 成立. 同理可以证明结论 2 也成立,在此省略.

于是,上述关于点是否属于矩形(或长方体)的保密判定问题可以转化为保密判定 $\boldsymbol{X}=(x,x,y,y)$ 与 $\boldsymbol{Y}=(x_1,x_2,y_1,y_2)$($\boldsymbol{X}=(x,x,y,y,z,z)$ 与 $\boldsymbol{Y}=(x_1,x_2,y_1,y_2,z_1,z_2)$)的曼哈顿距离与 $|x_1-x_2|+|y_1-y_2|$($|x_1-x_2|+|y_1-y_2|+|z_1-z_2|$)是否相等的问题.

应用类似于保密判定 x 是否属于区间 $[y_1,y_2]$ 的解决方案即可解决上面推广问题.下面简要叙述保密判定 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的曼哈顿距离与 $|x_1-x_2|+|y_1-y_2|$ 是否相等问题的解决方案.保密判定 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的曼哈顿距离与 $|x_1-x_2|+|y_1-y_2|+|z_1-z_2|$ 是否相等问题的解决方案与其完全类似,在此不再赘述.

(III) 假设 Alice 和 Bob 的数据 x,y 以及 x_1,x_2,y_1,y_2 为正整数,但无全集条件限制.在此情形下以协议 2 为基础协议设计方案如下:

(i) Alice 和 Bob 分别以 $\boldsymbol{X}=(u_1,u_2,u_3,u_4)=(x,x,y,y)$,$\boldsymbol{Y}=(v_1,v_2,v_3,v_4)=(x_1,x_2,y_1,y_2)$ 为输入执行协议 2 的前 2 步,Bob 得到密文 $E(u_j^2)$,$E(2u_j)$,$j\in[1,4]$ 及公钥.

(ii) 对于每一个 $j\in[1,4]$,Bob 选择随机数 r_1,计算

$$C_j=(E(u_j^2)E(2u_j)^{-v_j}E(v_j^2))^{r_1^2},$$

并随机置换 (C_1,C_2,C_3,C_4),得到 (C_1',C_2',C_3',C_4') 发送给 Alice.

(iii) Alice 对 C_j',$j\in[1,4]$ 进行解密得到 $t_j=D(C_j')$,并选择随机数 r_2,进一步计算 $E(t)=E(r_2(\sqrt{t_1}+\cdots+\sqrt{t_4}))$,$E(r_2)$,并将 $E(t)$ 和 $E(r_2)$ 告诉 Bob.

(iv) Bob 计算 $G=E(t)(E(r_2)^{-1})^{r_1(v_2-v_1+v_4-v_3)}$,发送给 Alice.

(v) Alice 解密 C,如果 $D(C)=0$,则输出点属于矩形,否则,输出点不属于矩形.

(IV) 假设 x,y,x_1,x_2,y_1,y_2 均属于给定全集 $U=\{a_1,\cdots,a_m\}$,此种情形下以协议 4 为基础设计方案如下:

(i) Alice 和 Bob 分别输入 $\boldsymbol{X}=(u_1,u_2,u_3,u_4)=(x,x,y,y)$,$\boldsymbol{Y}=(v_1,v_2,v_3,v_4)=(x_1,x_2,y_1,y_2)$ 执行协议 4 的前 3 步,Bob 得到密文 (H_{u_j},H_{z_j}),$j\in[1,4]$.

(ii) Bob 选择随机数 r,计算密文:

$$C_1=\prod_{j=1}^{4}\Big[\prod_{k=1}^{(v_j)_{\text{ind}}}X_{ji}\prod_{k=1}^{(a_m-v_j)_{\text{ind}}}Z_{jk}\Big]\bmod N^2,$$

$$C_2=E(v_2-v_1+v_4-v_3),\ C=(C_1C_2^{-1})^r\bmod N^2.$$

Bob 将 C 发送给 Alice.

(iii) Alice 解密 C,如果 $D(C)=0$,则输出点属于矩形,否则,输出点不属于矩形.

6.3 其他应用问题

向量相似度保密计算. 向量的相似度计算在实际生活中应用非常广泛,比如在生物特征密码中,判断生物特征模板是否相似.正如引言中所说,这类问题可以转化为保密计算描述这些生物特征的有关向量的相似度问题.在电子商务中也常常需要保密地计算消费者的相似度或者产品的相似度.而向量之间的相似程度往往是通过向量之间的"距离"来衡量的,因此很多情形下需要以曼哈顿距离保密计算协议为基础设计构造向量相似度的保密计算协议.

向量与向量集合的关系问题. 假设 Alice 有一个 n 维向量 $\boldsymbol{u}$,Bob 有一个 n 维向量构成的集合 $V=\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m\}$,那么如何保密判定 $\boldsymbol{u}$ 是否属于集合 V?对于这个问题的研究在保密的信息搜索、信息匹配等方面有广泛的应用.这个问题可以转化为曼哈顿距离问题获得解决.实际上,对 $i\in[1,m]$,如果以 $d(\boldsymbol{u},v_i)$ 表示向量 $\boldsymbol{u}$ 与 v_i 的曼哈顿距离,那么易知 $\boldsymbol{u}\in V$ 当且仅当存在 $i\in[1,m]$,使得 $d(\boldsymbol{u},v_i)=0$ 成立,应用前面的协议 5 或协议 6 的设计思想,再结合应用 Paillier 加密算法的加法同态性,容易求得 $d(\boldsymbol{u},v_i)$ 的一个密文,记为 C_i,判断 C_i 是否等于 $E(0)$,即保密判断 $d(\boldsymbol{u},v_i)=0$ 是否成立,就可使问题得到解决.

上述关于向量与向量集合的关系的保密判定问题可以转化为保密判定向量 $\boldsymbol{u}$ 与 v_i 的曼哈顿距离是否等于 0 的问题.

应用类似于保密判定点与区间关系问题的解决方案即可解决上面推广问题.

注解 3. 对于如何应用本文所述方法我们给出了一些实际的、技术方面的建议,具体如下:

(1) 如果要应用本文所述方法进行曼哈顿距离有关的安全计算,首先根据实际问题判定所涉及的隐私数据能否在某个全集内取值,如果可以确定适当的全集,则考虑应用协议 3 或 4 进行计算,使得在线复杂性较低.如果无法在保护隐私的情况下确定适当的全集,或不愿意借助云服务器帮助执行协议,则可应用协议 1 或 2 计算.比如考虑下面问题:

某城市设计新的开发区(为方型建筑区块).公司 A 与公司 B 都有自己理想的公司选址,在未正式确定选址位置时,他们的拟选位置属于公司的商业秘密,双方都不愿意泄露给对方.但由于公司 A 与

B之间有业务往来,他们希望两个公司之间有较合适的车程距离,如果距离不合适,还可以进行适当调整.这个问题即可转化为保密地计算平面上两个点的曼哈顿距离.在这例子中,由于开发区的大小范围确定,因此可以选择适当的坐标系和全集,使得两公司的选址坐标在给定的全集内取值.对于这个问题即可应用协议4进行保密计算.

而对于6.3节所考虑的保密信息搜索、保密信息匹配等问题,由于在保护隐私的情况下很难确定信息的范围,这类问题则需应用无全集限制下的协议2进行保密计算.

(2)在曼哈顿距离的计算中认为各个分量对于距离的贡献权重是一样的,这就决定了曼哈顿距离适用的场景.在实际生活的三维空间中长,宽,高确实可以放在一个体系下来度量,曼哈顿的行车路线也是如此,这些场景都适合直接应用曼哈顿距离.但在有些场景中需要对数据进行预处理后应用曼哈顿距离计算才有意义.例如,对于一个二维样本(身高,体重),其中身高取值范围为150～190,体重取值范围为50～60.若存在样本 $A(180,50)$,$B(190,50)$,$C(180,60)$,那么A,B之间的曼哈顿距离与A,C之间的曼哈顿距离相等.但显然身高10 cm与体重10 kg是不等价的.产生上面问题的原因是将各个分量的量纲当作相同的看待了.若要计算这类问题的相似度,应该首先对数据进行预处理,将各个分量都进行标准化后,便可应用本文协议去进行保密地计算曼哈顿距离.

7　总　结

本文基于Paillier同态加密算法设计了曼哈顿距离保密计算协议2与协议4.协议2与协议4是针对不同数据条件设计构造的:如果事先能够确定所有参与计算的数据属于某个全集内,那么应用协议4进行计算时,可以利用云生成若干个0与1的密文,这样可以大大降低协议的在线计算复杂性,协议4的效率很高.在无法确定全集U的情况下,则可以使用协议2进行运算,根据协议2的计算原理,只要在生成Paillier加密算法时参数N取得足够大,应用协议2计算曼哈顿距离对于数据再没有其他限制.因此本文设计的协议简单易行,拥有广泛的应用前景.关于设计恶意模型下曼哈顿距离安全高效的计算协议是我们进一步要研究的课题.

参　考　文　献

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

[2] Gennaro R, Robshaw M. Advances in Cryptology—CRYPTO 2015, Part II. Lecture Notes in Computer Science, 2015: 1-787

[3] Robshaw M, Katz J. Advances in Cryptology—CRYPTO 2016, LNCS 9815. Heidelberg, Germany: Springer, 2016: 1-685

[4] Fischlin M, Coron J S. Advances in Cryptology—EUROCRYPT 2016, Part II, LNCS 9666. Heidelberg, Germany: Springer, 2016: 1-703

[5] Goldwasser S. Multi-party computations: Past and present//Proceedings of the 16th Annual ACM Symposium on Principles of Distributed Computing. New York, USA, 1997: 1-6

[6] Cramer R, Damgård I B, Nielsen J B. Secure Multiparty Computation and Secret Sharing. London, UK: Cambridge University Press, 2015: 6-14

[7] Yao A C. How to generate and exchange secrets//Proceedings of the 27th Annual Symposium on Foundations of Computer Science. Toronto, Canada, 1986: 162-167

[8] Goldreich O, Micali S, Wigderson A. How to play any mental game//Proceedings of the 19th Annual ACM Conference on Theory of Computing. New York, USA, 1987: 218-229

[9] Goldreich O. The Fundamental of Cryptography: Basic Applications. London, UK: Cambridge University Press, 2004: 599-729

[10] Fagin R, Naor M, Winklery P. Comparing information without leaking it. Communications of the ACM, 1996, 39(5): 77-85

[11] Samanthula B K, Jiang W, et al. Secure intersection cardinality and its application to Jaccard coefficient. IEEE Transactions on Dependable and Secure Computing, 2016, 13(5): 591-604

[12] Freedman M J, Nissim K, Pinkas B. Efficient private matching and set intersection//Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques. Interlaken, Switzerland, 2004: 1-19

[13] Song D X. Privacy-preserving set operations//Proceedings of the 25th Annual International Cryptology Conference. California, USA, 2005: 241-257

[14] Yang Xiao-Yi, Li Shun-Dong, Kang Jia. Private replacement and its applications in scientific computation. Chinese Journal of Computers, 2018, 41(5): 1132-1142(in Chinese)
(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. 计算机学报, 2018, 41(5): 1132-1142)

[15] Dou J W, Gong L M, Li S D, et al. Efficient private subset computation. Security and Communication Networks, 2016, 9(18): 5965-5976

[16] Chen Zhen-Hua, Li Shun-Dong, Chen Li-Chao, et al. Fully privacy-preserving determination of point-range relationship. Science China Information Science, 2018, 48(2): 187-204(in Chinese)
(陈振华, 李顺东, 陈立朝等. 点和区间关系的全隐私保密判定. 中国科学: 信息科学, 2018, 48(2): 187-204)

[17] Atallah M J, Du W L. Secure Multi-Party Computational Geometry. Algorithms and Data Structures. Heidelberg, Germany: Springer, 2001: 165-179

[18] Li S D, Wu C S, Wang D Y, et al. Secure multiparty computation of solid geometric problems and their applications. Information Sciences, 2014, 282: 401-413

[19] Qin J, Duan H, Zhao H, et al. A new lagrange solution to the privacy-preserving general geometric intersection problem. Journal of Network and Computer Applications, 2014, 46: 94-99

[20] Du W L, Atallah M J. Privacy-preserving cooperative statistical analysis//Proceedings of the 17th Annual Conference of Computer Security Applications. New Orleans, USA, 2001

[21] Jawurek M, Kerschbaum F. Fault-Tolerant Privacy-Preserving Statistics. Privacy Enhancing Technologies. Heidelberg, Germany: Springer, 2012: 221-238

[22] Mehmed K. Data mining concepts, models, methods, and algorithms. IEEE Transactions, 2005, 36(5): 495-496

[23] Aggarwal C C. Privacy preserving data mining. Application Research of Computers, 2008, 9(8): 616-621

[24] Du W L, Atallah M J. Protocols for Secure Remote Database Access with Approximate Matching. E-Commerce Security and Privacy. New York, USA: Springer, 2001: 87-111

[25] Cachin C. Efficient private bidding and auctions with an oblivious third party//Proceedings of the 6th ACM Conference on Computer and Communications Security. New York, USA, 1999: 120-127

[26] Jain A K, Prabhakar S, Hong L, et al. FingerCode: A filterbank for fingerprint representation and matching//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Fort Collins, USA, 1999: 187-193

[27] Kikuchi H, Nagai K, Ogata W, et al. Privacy-preserving similarity evaluation and application to remote biometrics authentication. Soft Computing, 2010, 14(5): 529-536

[28] Fang Le-Di, Li Shun-Dong, Dou Jia-Wei. Secure Manhattan distance computation. Journal of Cryptologic Research, 2019, 6(4): 512-525(in Chinese)
(方乐笛, 李顺东, 窦家维. 曼哈顿距离的保密计算. 密码学报, 2019, 6(4): 512-525)

[29] Paillier P. Public-key cryptosystems based on composite degree residuosity classes. Advances in Cryptology-Eurocrypt, 1999, 547(1): 223-238

DOU Jia-Wei, Ph. D., associate professor. Her main research interests include applied mathematics and applied cryptography.

GE Xue, M. S. candidate. Her main research interests include applied mathematics and applied cryptography.

WANG Ying-Nan, M. S. candidate. Her main research interests include applied mathematics and applied cryptography.

Background

Secure multiparty computation (SMC), a crucial field of cryptography, was first introduced by Yao in the 1980s, and has become one of the most active research fields of modern cryptography and a focus in the international cryptographic community. Since SMC was introduced, cryptographic scholars have studied numerous SMC problems arising in various fields such as secure scientific computation, secure computational geometry, privacy-preserving data mining, secure statistical analysis and social management (e. g., private voting); there are also many new problems have not been studied; many previously addressed problems also need further effort to improve the protocol efficiency or to explore more efficient solutions.

Our paper studies secure Manhattan distance computation. This problem has not been investigated in the literatures. Designing protocol for this problem has important theoretical significance in cryptography, and the protocol has very important practical significance and broad application prospect in constructing other secure multiparty computation protocols to privately compute the path between two private points, to privately determine the relationship between a private point and a private interval, between a private point and a private set, and to privately compute the similarity between two vectors, etc.

By introducing new encoding schemes and using the Paillier additively homomorphic encryption, combining the

encrypt-and-choose technology, we first design a protocol to privately compute the absolute value of the difference of two private numbers. And then based on the protocol for absolute value, we have to keep the absolute value secret and add up privately obtain the Manhattan distance.

We design two new protocols for computing the Manhattan distance for two different scenarios: either the domain of the private data is known and not very large, or the domain of the private data is unknown. We prove that these protocols are secure in the semi-honest model. We analyze the efficiency of these protocols and simulate these protocols in our PC. Theoretical analysis and simulation result show that our protocols are efficient. Finally, we show how to use our protocol as a building block to solve other secure multiparty computation problems.

This study is sponsored by the National Natural Science Foundation of China under Grant No. 61272435. The aim of these projects is to address various privacy preserving problems arising in cyberspace. Our team has been engaged in the design and analysis of cryptographic protocols, such as SMC, secret sharing, bit commitment, zero-knowledge proof, digital signature, secure scientific computation, and secure computational geometry over 10 years. We have published over 50 papers, of which near 30 have been indexed by SCI. And we have proposed some effective methods such as encrypt-and-choose, private substituting, encoding method to efficiently solve various secure multiparty computation problems.