一类基于安全多方计算的数据挖掘隐私保护模型的可行性分析*

景征骏[1,2], 蒋国平[1], 古春生[2]
(1. 南京邮电大学 计算机学院, 南京 210003; 2. 江苏理工学院 计算机工程学院, 江苏 常州 213001)

摘　要: 阐述安全多方计算(SMC)密码原语在分布式数据挖掘隐私保护中的相关应用后,对方炜炜等人提出的基于 SMC 的隐私保护数据挖掘模型进行分析,论证该类模型所基于的离散对数公钥加密协议不具有全同态的特性,并用简单实例验证。从而得出该类数据挖掘隐私保护模型是不可行的。

关键词: 隐私保护; 数据挖掘; 同态加密; 安全多方计算; 可行性分析
中图分类号: TP309.2　　**文献标志码**: A　　**文章编号**: 1001-3695(2014)02-0543-04
doi:10.3969/j.issn.1001-3695.2014.02.053

Feasibility analysis of privacy-preserving data mining models based on security multiparty computation

JING Zheng-jun[1,2], JIANG Guo-ping[1], GU Chun-sheng[2]
(1. *College of Computer*, *Nanjing University of Posts & Telecommunications*, *Nanjing* 210003, *China*; 2. *School of Computer Engineering*, *Jiangsu University of Technology*, *Changzhou Jiangsu* 213001, *China*)

Abstract: After describing the relevant applications about the security multiparty computation (SMC) in the preserving-privacy data mining, the authors show the infeasibility of SMC-based preserving-privacy data mining models proposed by Fang et al. The reason for this failure is that the cryptology scheme based on discrete logarithm problem (CSBDLP), on which the preserving-privacy models are based, does not satisfy the required fully homomorphic property. Furthermore, this can be seen from a small example.

Key words: privacy-preserving; data mining; homomorphic encryption; secure multiparty computation (SMC); feasibility analysis

0　引言

多元统计分析以及聚类等方法作为常用的数据挖掘技术,在商业、医疗等领域的知识发现中有着重要的应用。例如,多元线性回归分析通过数据驱动建立预报变量与响应变量之间的回归模型,从而实现变量的预测建模[1]。然而,在获得准确的知识挖掘的同时,如何保证敏感的原始数据(如客户资料、病人信息)不被泄露已成为数据挖掘领域的重要课题。隐私保护数据挖掘(privacy preserving data mining,PPDM)旨在保证数据挖掘过程的可靠性,同时避免由于信息泄露给参与挖掘的主体带来损失[2]。近年来,对 PPDM 技术的研究主要通过两种途径:数据扰乱[3~5]和基于密码学的安全多方计算(security multiparty computation,SMC)[6~8]。由于数据扰乱方法易收到各种类型的攻击,且为满足隐私性而添加的扰乱数据会使挖据结果的正确性受到一定的影响[9], 因此该类方法发展较为缓慢。而基于密码学的安全多方计算方法可以对在挖据过程中的敏感数据进行加密保护,且不同种类的隐私保护数据挖据任务可以转换成特定的多方安全计算问题[6]。所以,根据具体的挖据要求设计相应的 SMC 协议成为研究 PPDM 技术的主流方向[10~13]。例如,有 n 个观测样本集(x_1,x_2,x_3,x_4,x_5,y),若样本集是异构垂直分布存储,即 Alice 拥有预报变量(x_1,x_2)的 n 个观测值,而 Bob 拥有预报变量(x_3,x_4,x_5)的 n 个观测值,响应变量 y 的 n 个观测值是公开的。在此情形下,是否能在不公开各数据拥有者原始数据的同时,不可信的第三方可采用多元统计分析来获得相应变量的预测模型,从密码学角度来说,上述情景是一个典型的安全多方计算问题,可通过安全多方计算协议来解决。然而,由于数据挖掘算法通常都极为复杂,且算法的输入值都为庞大的数据集(如多变量线性回归分析),一般的 SMC 协议无法在实际应用中胜任。为此,文献[12]定义了安全的两方多变量统计分析(secure 2-party multivariate statistical analysis,S2-MSA),并设计了点积协议实现矩阵乘法、矩阵逆和转置等运算,从而完成回归系数的求解以及分类统计。但为了有较高执行效率,该方案降低了系统的安全性。它的安全模型从理论上讲是启发式的,如果参数矩阵选择不当则会泄露参与方的原始数据。另外,为了保证运算过程的中间结果的安全性,该方案仍需参与方之间多轮通信。文献[13]提出安全求和协议实现了垂直分布式的隐私保护聚类模型,但对协议

收稿日期: 2013-05-17; **修回日期**: 2013-06-19　**基金项目**: 国家自然科学基金资助项目(60874091,61142007);国家教育部高等学校博士学科点专项科研基金资助项目(20103223110003);国家教育部人文社会科学研究基金资助项目(12YJAZH120);江苏省普通高校研究生科研创新计划资助项目(CXZZ13_0493);江苏省高校自然科学研究面上项目(13KJB520005)

作者简介:景征骏(1978-),男,江苏丹阳人,博士研究生,主要研究方向为网络与信息安全(jzjing@jsut.edu.cn);蒋国平,男,教授,博导,主要研究方向为混沌系统与复杂动态网络;古春生,男,副教授,博士,主要研究方向为密码设计与分析.

的中间结果未作保护,所以某一方的原始数据容易被某两方合谋而泄露。受文献[12,13]的启发,方炜炜等人设计了解决上述情形的一类基于同态公钥加密协议的异构分布式隐私保护数据挖掘模型:多元线性回归隐私保护模型[14]和聚类隐私保护模型[15]。该类模型尝试利用公钥加密协议的同态特性设计密文运算规则,使挖掘算法可以在原始数据的密文空间上进行同态的矩阵加、乘运算。由于不可信的挖掘方获得的只是原始数据的密文,挖掘算法的中间结果也为密文。当然,这也无须参与方多轮通信。若该方案可行,参与方的原始数据在整个数据挖掘过程中都不会被泄露,且能保证挖掘结果的正确性和较高执行效率,从而能彻底解决数据异构分布存储情形下的数据挖掘隐私保护问题。

设计能实际应用的全同态密文计算是密码学领域一直追求的目标之一,也是当前研究的热点。本文论证了文献[14]提出的基于离散对数问题的公钥加密协议设计的运算规则不具有全同态的特性,指出该协议仅能实现密文的同态加法,而不能实现密文的同态乘法,从而得出该类隐私保护数据挖掘模型是不可行的。同时注意到文献[14]对本原根的定义不准确,其公钥加密协议中模数的本原根并不存在,导致其安全性也是较弱的。由于文献[15]中的聚类隐私保护模型是基于文献[14]中提出的同态公钥加密协议,因此本文重点讨论文献[14]中的多元回归隐私保护模型的可行性,对文献[15]的可行性分析与之类似。

本文给出了文献[14]中公钥加密协议 CSBDLP 的相关定义及基于该协议的隐私保护模型的描述,理论推导了隐私保护模型中密文矩阵的乘法运算不具备同态性质,从而得出该类隐私保护数据挖掘模型是不可行的,并用简单实例验证基于该公钥协议设计的运算规则不满足全同态条件。

1 相关定义与模型回顾

为描述方便,本文采用以下的相关符号表示。矩阵 $M_{N\times(n+m+1)}=\begin{bmatrix} x_{1,1} & \cdots & x_{1,n} & x_{1,n+1} & \cdots & x_{1,n+m} & y_1 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ x_{N,1} & \cdots & x_{N,n} & x_{N,n+1} & \cdots & x_{N,n+m} & y_N \end{bmatrix}$ 表示 N 条记录的数据集,其中每列对应于数据集的每个属性($x_1,\cdots x_n,\cdots,x_{n+m},y$)。依照文献[14]所设的情景,可将数据集 $M_{N\times(n+m+1)}$ 垂直划分,并异构分布存储于三个不同的位置,即 Alice 拥有属性($x_1,\cdots,x_n$)的数据子集 V_A,Bob 拥有属性($x_{n+1},\cdots,x_{n+m}$)的数据子集 V_B,Chris 拥有属性(y)的数据子集 V_C。

1.1 相关定义

定义 1 若观测样本数据集 $M=(x_1,\cdots,x_n,\cdots,x_{n+m},y)$,多元线性回归分析就是确定响应变量 y 的值与预报变量 $X=(x_1,\cdots x_n,\cdots,x_{n+m})$ 值的函数关系,即求解参数 β,使得 $y=X\beta$ 最大限度地满足数据集 M。为了使实际的观测值 y 与预测值之间的误差最小(残差),可由最小二乘法求得回归系数 $\beta=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$。当观测样本数据集被垂直划分,即 $M=(V_A:V_B:V_C)$ 时,由分块矩阵运算得

$$\beta=\left[\begin{bmatrix} V_A^{\mathrm{T}} \\ V_B^{\mathrm{T}} \end{bmatrix}\ [V_A V_B]\right]^{-1}\begin{bmatrix} V_A^{\mathrm{T}} \\ V_B^{\mathrm{T}} \end{bmatrix}V_C \tag{1}$$

定义 2 设$\langle R_1,+,\times\rangle$和$\langle R_2,\oplus,\otimes\rangle$是两个环,若映射 $f:R_1\mapsto R_2$,对于任意的 $m_i\in R_1$ 均满足:a)加法同态,即 $f(m_1+m_2)=f(m_1)\oplus f(m_2)$;b)乘法同态,即 $f(m_1\times m_2)=f(m_1)\otimes f(m_2)$,则称 f 为同态映射,环 R_2 是环 R_1 的同态像。

定义 3[16] 如果 a 和 n 是互素的整数且 $n>0$,那么当 $\mathrm{ord}_n^a=\varphi(n)$ 时,称为 r 是模 n 的本原根(以下简称原根)。其中,符号 ord_n^a 表示 a 模 n 的阶,是使 $a^{\mathrm{ord}_n^a}\equiv 1(\mathrm{mod}\ n)$ 成立的最小正整数;$\varphi(n)$ 为欧拉 φ 函数,是 $1\leqslant a\leqslant n$ 的整数中满足 $\gcd(a,n)=1$ 的整数个数($\gcd$ 表示求最大公约数的函数)。

定义 4 给定一个大素数 p 和它的一个原根 b,寻找整数 a 对模 n 的以 b 为底的指数问题称为离散对数问题,即若 $b^x\equiv a\ \mathrm{mod}\ p$,求解指数 x 是困难的。

定义 5 基于离散对数困难的加密协议(cryptology scheme based on discrete logarithm problem, CSBDLP)[14],选择大素数 p 和 q,令 $n=p\times q$,其本原根为整数 b,对于满足 $1<c,r,x<n$ 的整数 c、r 和 x,若 $b^t\equiv c\ \mathrm{mod}\ n$,$g(x)=(x+rp)c^r\mathrm{mod}\ n$,则 $g^{-1}=(b^{tr})^{-1}((x+rp)c^r\mathrm{mod}\ n)\mathrm{mod}\ p$ 成立。

定理 1[16] 如果正整数 n 不是一个素数的幂或者不是一个素数幂的两倍,那么 n 不存在原根。

1.2 模型回顾

对于定义 1 中设定情景的数据集 $M=(V_A:V_B:V_C)$,现将文献[14]根据 CSBDLP 加密协议构造的求多元线性回归系数估计 β 的隐私保护模型概括如下:

a)密钥参数。密钥中心机构生成加密公钥 $pk=\{b,r,p,q,n\}$、加密私钥 $sk=\{k_A,k_B,k_C\}$ 和解密密钥 t'。同时选择随机数 s,并计算 $l=\mathrm{lcm}(k_A^r,k_B^r,k_C^r)$ 及 $K_i=\dfrac{s\cdot l}{k_i^r}$,其中($i=\{A,B,C\}$,lcm 为求最小公倍数函数)。另外,解密密钥 t' 满足离散对数 $b^{t'}\equiv(s\cdot l)\mathrm{mod}\ n$,并由密钥中心保留。

b)数据加密。根据定义 5 描述的加密协议,参与方 Alice、Bob 和 Chris 分别使用私钥 k_A、k_B、k_C 对各自拥有的原始数据加密,得到密文 $E_i=f(V_i)=(V_i+rp)\cdot k_i^r\ \mathrm{mod}\ n$,其中 $i=\{A,B,C\}$。此后,密文发送给不可信的第三方进行数据挖掘。

c)第三方数据挖掘。半诚实的第三方采用密文空间的$\oplus$和$\otimes$的运算规则进行矩阵的加、乘和求逆等运算,实现式(2)的线性回归系数密文的求解。

$$E(\beta)=\left[\begin{bmatrix} E_A^{\mathrm{T}} \\ E_B^{\mathrm{T}} \end{bmatrix}\ [E_A E_B]\right]^{-1}\begin{bmatrix} E_A^{\mathrm{T}} \\ E_B^{\mathrm{T}} \end{bmatrix}E_C \tag{2}$$

而后,将结果发送至密钥中心机构解密获得式(1)的 β 值。其运算规则有两部分。

规则 1 当参与方提供密文数据 E_A、E_B、E_C 时,$E_A\oplus E_B=(K_A\cdot E_A+K_B\cdot E_B)\mathrm{mod}\ n$;$E_A\otimes E_B=(K_A\cdot E_A\cdot K_B\cdot E_B)\mathrm{mod}\ n$。同理,$E_C$ 参与运算时用系数 K_C。

规则 2 当为密文空间数据 x 和 y 时,$x\oplus y=(x+y)\mathrm{mod}\ n$;$x\otimes y=(x\cdot y)\mathrm{mod}\ n$。

d)结果解密。密钥中心机构根据定义 5 的加密协议进行解密,即 $\beta=f^{-1}(E(\beta))=(b^{t'})^{-1}E(\beta)\mathrm{mod}\ p$。

文献[14,15]认为根据定义 5 的加密协议设定的密文运算是全同态的,即满足定义 2 的性质。从而不可信的挖掘方可在密文同态像中计算线性回归估计系数的密文 $E(\beta)$,再解密后可以得到正确的明文 β。

2 模型的可行性分析

虽然 CSBDLP 公钥加密协议能正确地进行明文加密和密文求逆,但基于其协议设定的密文运算规则不满足全同态的条件。

定理 2　基于离散对数困难问题的加密协议(CSBDLP)设定的密文运算不具有全同态特性,在密文矩阵运算中仅加运算⊕同态,而乘运算⊗则不同态。

证明　采用 1.2 节加密协议的密钥参数对密文矩阵的加、乘运算进行推导。

1)密文矩阵加法　现取明文空间的矩阵 $A_{n\times h}$、$C_{n\times h}$,经 CSBDLP 加密协议加密后得密文矩阵 $f(A_{n\times h})=(A+rp)\cdot k_A^r \bmod n$ 及 $f(C_{n\times h})=(C+rp)\cdot k_C^r \bmod n$。令密文矩阵 $f(A_{n\times h})$ 和 $f(C_{n\times h})$ 相加的结果为矩阵 Q,则矩阵的每个元素 $Q_{i,j}$ 按密文运算规则计算如下:

$$Q_{i,j}=f(A_{n\times h})_{i,j}\oplus f(C_{n\times h})_{i,j}=$$
$$(K_A\cdot(A_{i,j}+r\cdot p)\cdot k_A^r \bmod n+K_C\cdot(C_{i,j}+r\cdot p)\cdot k_C^r \bmod n)\bmod n=$$
$$(\frac{s\cdot l}{k_A^r}\cdot(A_{i,j}+r\cdot p)\cdot k_A^r \bmod n+\frac{s\cdot l}{k_C^r}\cdot(C_{i,j}+r\cdot p)\cdot k_C^r \bmod n)\bmod n=$$
$$s\cdot l\cdot(A_{i,j}+C_{i,j}+2rp)\bmod n$$

因此,$f^{-1}(Q_{i,j})=(b^t)^{-1}\cdot s\cdot l\cdot(A_{i,j}+C_{i,j}+2rp)\bmod p\cdot q \bmod p$,因为 $b^t\equiv(s\cdot l)\bmod n$,由定义 5 可得 $f^{-1}(Q_{i,j})=A_{i,j}+C_{i,j}$,即 $f(A_{i,j}+C_{i,j})=f(A_{i,j})\oplus f(C_{i,j})$。从而 $f(A+C)=f(A)\oplus f(C)$。

2)密文矩阵乘法　现取明文空间矩阵 $A_{n\times h}$ 和 $B_{h\times m}$,经 CSBDLP 加密协议加密后得密文矩阵 $f(A)=(A+rp)\cdot k_A^r \bmod n$ 及 $f(B)=(B+rp)\cdot k_B^r \bmod n$。令密文矩阵 $f(A)$ 和 $f(B)$ 相乘的结果为矩阵 $N_{n\times m}$,则矩阵的每个元素 $N_{i,j}$ 按密文运算规则计算为

$$N_{i,j}=(f(A)_{i,1}\otimes f(B)_{1,j})\oplus(f(A)_{i,2}\otimes f(B)_{2,j})\oplus\cdots\oplus(f(A)_{i,h}\otimes f(B)_{h,j})=((K_A\cdot(A_{i,1}+rp)\cdot k_A^r \bmod n)\cdot(K_B\cdot(B_{1,j}+rp)\cdot k_B^r \bmod n))\oplus\cdots\oplus((K_A\cdot(A_{i,h}+rp)\cdot k_A^r \bmod n)\cdot(K_B\cdot(B_{h,j}+rp)\cdot k_B^r \bmod n))=$$
$$\sum_{u=1}^{h}((K_A\cdot(A_{i,u}+rp)\cdot k_A^r \bmod n)\cdot(K_B\cdot(B_{u,j}+rp)\cdot k_B^r \bmod n))\bmod n=\sum_{u=1}^{h}((\frac{s\cdot l}{k_A^r}\cdot(A_{i,u}+rp)\cdot k_A^r \bmod n)\cdot(\frac{s\cdot l}{k_B^r}\cdot(B_{u,j}+rp)\cdot k_B^r \bmod n))\bmod n=(s\cdot l)^2\sum_{u=1}^{h}((A_{i,u}+rp)\cdot(B_{u,j}+rp))\bmod n=(s\cdot l)^2\sum_{u=1}^{h}(A_{i,u}\cdot B_{u,j}+(A_{i,u}+B_{u,j})\cdot rp+r^2p^2)\bmod n$$

因此,根据定义 5 中的加密协议求密文乘法的逆得

$$f^{-1}(N_{i,j})=(b^t)^{-1}\cdot(s\cdot l)^2\cdot\sum_{u=1}^{h}(A_{i,u}\cdot B_{u,j}+(A_{i,u}+B_{u,j})\cdot rp+r^2p^2)\bmod p\cdot q \bmod p=(s\cdot l)\sum_{u=1}^{h}A_{i,u}\cdot B_{u,j}$$

也就是说 $f^{-1}(N_{i,j})\neq\sum_{u=1}^{h}A_{i,u}\cdot B_{u,j}$,从而,$f(A)\otimes f(B)\neq f(A\cdot B)$。

根据定义 2 中同态条件并结合以上的推导 1)和 2),可以得出定理 2。证毕

推论 1　基于 CSBDLP 公钥加密协议的多元线性回归隐私保护模型不可行。

证明　当数据挖掘方在使用最小二乘估计法求解回归系数时,所用的数据都是参与方使用 CSBDLP 公钥加密协议加密后得到密文矩阵,而密文空间的矩阵运算将使用⊕、⊗的运算规则。因此,根据定理 2 易得此推论。证毕

推论 2　基于 CSBDLP 公钥加密协议的隐私保护聚类模型是不可行的

证明　同推论 1。

3　实例验证

在 Elgamal 提出了具有实际应用的 ELGamal 公钥体制和数字签名后[17],有限域上的离散对数问题的难解性就成为很多公钥密码系统安全的理论基础,而本原根是产生离散对数问题的关键。根据定义 3 可知,若 n 为素数,则原根 b 存在($\mathrm{ord}_n^b=\varphi(n)=n-1$),且模 n 是一个乘法循环群 $Z_n\setminus\{0\}$,群的阶 $|Z_n^*|=n-1$。由本原根的性质知道,本原根 b 的乘法幂将遍历群 Z_n^* 的每个元素。因此求解整数 a 对模 n 的以 b 为底的离散对数的概率与在 $n-1$ 个元素集合中随机选取一个元素的概率是不可区分的。也就是说,当 n 为大素数时求解离散对数问题是困难的。虽然素数一定存在本原根[16],可大部分合数是不存在本原根的。根据定理 1 可知,定义 5(文献[14]的定义 6)描述的 CSBDLP 加密协议中的参数 $n(n=p\times q)$ 没有原根。因此,模 n 形成的循环群的阶必定小于 $(p-1)(q-1)$。从而该加密协议的安全性要远低于文献[14]中所证明的安全性。在下面的实例验证中为了获得模 n 的循环群的阶,可分别取素数 p、q 的本原根 (r_1,r_2),其满足 $\mathrm{ord}_n^{r_1\times r_2}=\mathrm{lcm}(p-1,q-1)$,即模 n 的阶为 $(p-1,q-1)$ 的最小公倍数。通过工具 Mathematica 对基于 CSBDLP 加密协议的密文矩阵运算进行实验,本文随机选取如下的原始数据作为此隐私保护数据挖掘模型的输入数据:

$$V_A=\begin{bmatrix}1&2\\1&0\\1&1\\2&3\end{bmatrix},V_B=\begin{bmatrix}2&5\\3&2\\1&4\\1&3\end{bmatrix}$$

密钥参数:pk:$\{r=7,p=17,q=13,n=p\times q=221,b=6\}$,$sk$:$\{k_A=11,k_B=3\}$;随机选择 $s=32$。因为 p、q 的本原根分别是 3 和 2,所以取 $b=6$,其满足 $b^{\mathrm{lcm}(p-1,q-1)}\equiv 1 \bmod n$。再计算 $K_A=148$、$K_B=181$ 和解密密钥 $t'=2$(t' 满足 $b^{t'}\equiv s\cdot\mathrm{lcm}(k_A^r,k_B^r)\bmod n$)。

数据加密:

$$E_A=\mathrm{Mod}[(V_A+r\cdot p)\cdot(k_A\wedge r),n]=\begin{bmatrix}71&125\\71&17\\71&71\\125&179\end{bmatrix}$$

$$E_B=\mathrm{Mod}[(V_B+r\cdot p)\cdot(k_B\wedge r),n]=\begin{bmatrix}90&21\\67&90\\113&44\\113&67\end{bmatrix}$$

密文矩阵运算($E_\oplus$、$E_\otimes$ 分别表示密文矩阵加、乘运算的结果)

$$E_\oplus=E_A\oplus E_B=\mathrm{Mod}[K_A\cdot E_A+K_B\cdot E_B,n]=\begin{bmatrix}57&201\\93&21\\21&129\\57&165\end{bmatrix}$$

$$E_\otimes=E_A^{\mathrm{T}}\otimes E_B=\mathrm{Mod}[\mathrm{Transpose}[\mathrm{Mod}[K_A*E_A,n]]\cdot(\mathrm{Mod}[K_B*E_B,n]),n]=\begin{bmatrix}185&119\\151&126\end{bmatrix}$$

密文解密($V_\oplus$、$V_\otimes$ 分别表示 $E_\oplus$、$E_\otimes$ 解密后的值)

$$V_\oplus=\mathrm{Mod}[\mathrm{Mod}[\mathrm{PowerMod}[b\wedge t',-1,n]*E_\oplus,n],p]=\mathrm{Mod}[\mathrm{Mod}[43\times\begin{bmatrix}57&201\\93&21\\21&129\\57&165\end{bmatrix},221],17]=\begin{bmatrix}3&7\\4&2\\2&5\\3&6\end{bmatrix}$$

$$V_\otimes=\mathrm{Mod}[\mathrm{Mod}[\mathrm{PowerMod}[b\wedge t',-1,n]*E_\otimes,n],p]=\mathrm{Mod}[\mathrm{Mod}[43\times\begin{bmatrix}185&119\\151&126\end{bmatrix},221],17]=\begin{bmatrix}16&0\\16&12\end{bmatrix}$$

对原始数据进行加、乘运算,可以发现 $V_A + V_B = V_{\oplus}$,而 $V_A^{\mathrm{T}} \times V_B \neq V_{\otimes}$。

上面的实例进一步验证了基于 CSBDLP 加密协议的密文矩阵运算不具有全同态特性。从而,文献[14]试图利用此密文运算规则设计的多元线性回归隐私保护模型是不能正确地恢复出回归系数估计值 β。

4 结束语

隐私保护是数据挖掘领域中重要的课题之一,并以完成准确的知识发现的同时不泄露敏感的原始数据为目标。文献[14,15]试图基于全同态思想设计多方安全的密文计算,并以此构建数据挖掘隐私保护模型,确保各个参与方的原始数据在不可信的第三方进行的数据挖掘过程中不被泄露。此类通过设计高效的 SMC 协议来完成数据隐私保护的方法是 PPDM 技术发展的方向之一,是一种有意义的尝试。但文献[14,15]所基于离散对数问题加密协议(CBSDLP)设计的密文运算规则并不具有全同态特性。因此,在此基础上构建的多元线性回归隐私保护模型和聚类隐私保护模型都是不可行的。目前,尽管已经设计出基于格问题的全同态加密方案[18,19],可其效率极低,还不具有实用性。除了不断优化基于格的全同态方案,研究者们也不断尝试基于经典数论难题(如离散对数问题)去设计高效的同态加密方案。然而,是否能够构造出高效且安全的基于离散对数问题的全同态加密方案仍然是一个公开的问题。

参考文献:

[1] YANG Bing-ru, GAO Jing, LI Lin-na. Research overview of regulations in the process of dynamic mining [J]. Journal of Computational Information Systems, 2006,2(3):973-979.

[2] AGRAWAL R, SRIKANT R. Privacy preserving data mining[C]// Proc of ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2000:439-450.

[3] ISALAM M, BRANKOVIC L. Privacy preserving data mining:a noise addition framework using a novel clustering technique[J]. Knowledge-based Systems,2011,2(1):1214 -1223.

[4] 刘峰,薛安荣,王伟. 一种隐私保护关联规则挖掘的混合算法[J]. 计算机应用研究, 2012,29(3):1107-1110.

[5] 张鹏,唐世渭. 一种有效的隐私保护关联规则挖掘方法[J]. 软件学报, 2006,17(8):1764-1774.

[6] PINKAS B. Cryptographic techniques for privacy preserving data mining [J]. ACMSIGKDD Explorations Newsletter,2002,4(2):12-19.

[7] CLIFFTON C, KANTARCIOGLU M, VAIDYA J. Tools for privacy preserving distributed data mining [J]. ACM SIGKDD Explorations Newsletter,2004,4(2):28-34.

[8] DOSHI N. A novel approach for cryptography technique on perturbed data for distributed environment[J]. International Journal on Cryptography and Information Security,2012,2(3):101-106.

[9] AGGARWAL C,YU P S. Privacy preserving data mining models and algorithms[M]. Berlin: Springer-Verlag, 2007:67.

[10] ALI I, YUCEL S. Privacy preserving clustering on horizontally partitioned data[J]. Data & Knowledge Engineering,2007,63(2):646-666.

[11] 刘英华,杨炳儒,马楠,等. 分布式隐私保护数据挖掘研究[J]. 计算机应用研究, 2011, 28(10):3606-3610.

[12] DU W L, HAN Y S. Privacy preserving multivariate statistical analysis: linear regression and classification[C]//Proc of the 4th SIAM International Conference on Data Mining. [S. l.]:IEEE Press,2004:222-233.

[13] VAIDYA J,CLIFTON C. Privacy preserving K-means over vertically partitioned data [C]//Proc of SIGKDD International Conference on Knowledge and Data Mining. New York:ACM Press,2003:490-510.

[14] 方炜炜,仁江,夏红科. 一种异构分布的多元线性回归隐私保护模型[J]. 计算机研究与发展, 2011,48(9):1685-1691.

[15] 方炜炜,杨炳儒,夏红科. 基于 SMC 的隐私保护聚类模型[J]. 系统工程与电子技术, 2012,34(7):1505-1511.

[16] ROSEN K H. Elementary number theory and its applications[M]. 5th ed. New York:Person Education,2005:245.

[17] ELGAMAL T. A public key cryptosystem and a signature scheme based on discrete logarithms [J]. IEEE Trans on Information and Theory,1985,IT-31(4):469-472.

[18] GENTRY C. A fully homomorphic encryption scheme [D]. [S. l.]: Stanford University, 2009.

[19] BRAKERSKI Z, VAIKUNTANATHAN V. Efficient fully homomorphic encryption from (standard) LWE[C]//Proc of the 52nd Annual Symposium on Foundations of Computer Science. 2011:97-106.

(上接第 539 页)

[9] KHADER D. Attribute based group signature with revocation,Report 2007/241[EB/OL]. (2007). http://eprint. iacr. org/2007/241.

[10] LI Jin, KIM K. Attribute-based ring signatures[EB/OL]. [2008-07-12]. http://eprint. iacr. org/2008/394.

[11] HERRANZ J, LAGUILLAUMIE F,LIBERT B,*et al*. Short attribute-based signatures for threshold predicates[C]//LNCS,vol 7178. Berlin:Springer-Verlag,2012:51-67.

[12] BONEH D, BOYEN X, SHACHAM H. Short group signatures[C]// LNCS,vol 3152. Berlin:Springer-Verlag,2004:41-55.

[13] CHASE M. Multi-authority attribute based encryption[C]//Lecture Notes in Computer Science,vol 4392. Berlin:Springer-Verlag,2007:515-534.

[14] LIN Huang, CAO Zhen-fu, LIANG Xiao-hui,*et al*. Secure threshold multi-authority attribute based encryption without a central authority [C]//Lecture Notes in Computer Science, vol 5365. Berlin:Springer-Verlag,2008:426-436.

[15] 孙昌霞,马文平,陈和凤. 多授权中心的基于属性的签名[J]. 四川大学学报:工程科学版,2011,43(1):83-86.

[16] CAO Dan, ZHAO Bao-kang,WANG Xiao-feng,*et al*. Multi-authority attribute-based signature[C]//Proc of the 3rd International Conference on Intelligent Networking and Collaborative Systems. 2011:668-672.

[17] MÄULLER S, KATZENBEISSER S, ECKERT C. Distributed attribute-based encryption[C]//Proc of the 11th International Conference on Information Security and Cryptology. 2008:20-36.

[18] BONEH D, FRANKLIN M. Identity-based encryption from the Weil pairing [C]//LNCS, vol 2139. Berlin: Springer-Verlag, 2001: 213-229.

[19] WATERS B. Efficient identity-based encryption without random oracles[C]//Lecture Notes in Computer Science, vol 3494. Berlin: Springer-Verlag,2005:114-127.

[20] GE A J,MA C G, ZHANG Z F. Attribute-based signature scheme with constant size signature in the standard model[J]. IET Information Security,2012,6(2):47-54.

[21] HU Chun-qiang, ZHANG Nan,LI Hong-juan, *et al*. Body area network security:a fuzzy attribute-based signcryption scheme[J]. IEEE Journal on Selected Areas in Communications,2013,31(9):37-46.

[22] CHEN Yan-li, CHEN Jun-jun, YANG Geng. Provable secure multi-authority attribute based signatures[J]. Journal of Convergence Information Technology(JCIT),2013,8(2).

[23] 张磊,曹珍富. 一个适合分布式网络的属性基加密方案[J]. 上海交通大学学报,2010,44(11):1507-1512.