# 面向CNN的区块链可信隐私服务计算模型

丁 毅[1],沈 薇[1],李海生[2,3],钟琼慧[1],田明宇[1],李 洁[1]
(1. 北京物资学院信息学院,北京 101149; 2. 北京工商大学农产品质量安全追溯技术及应用国家工程实验室,北京 100048;
3. 北京工商大学计算机学院,北京 100048)

**摘 要:** 在当前移动互联网时代,数据量增长迅速,服务计算能力不断增强,数据隐私保护和服务环境可信成为备受关注的重要问题. 本文研究面向卷积神经网络典型应用场景的可信隐私服务计算模型,探索支持同态加密的数据和模型计算方法,保护数据隐私. 构建基于区块链和智能合约技术服务过程存证及计算权益分配方法,保证服务计算的公开透明、可信可追溯. 探索资源提供者、模型拥有者及用户的新型云环境资源数据服务模式,促进资源有效整合,发展共享经济. 最后,通过实验分析该模型的隐私保护方法.

**关键词:** 同态加密; 隐私保护; 智能合约; 卷积神经网络
**中图分类号:** TP309.2 **文献标识码:** A **文章编号:** 0372-2112(2022)06-1399-11
**电子学报URL:**http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20200731

# Blockchain Trusted Privacy Service Computing Model for CNN

DING Yi[1], SHEN Wei[1], LI Hai-sheng[2,3], ZHONG Qiong-hui[1], TIAN Ming-yu[1], LI Jie[1]
(1. *School of Information, Beijing Wuzi University, Beijing* 101149, *China*;
2. *National Engineering Laboratory for Agri-product Quality Traceability, Beijing Technology and Business University, Beijing* 100048, *China*;
3. *School of Computer Science and Engineering, Beijing Technology and Business University, Beijing* 100048, *China*)

**Abstract:** Data privacy protection and service environment trust have become important issues with the rapidly increasing amount of data and computing power of services in current era of Mobile Internet. This paper studies the trusted privacy service computing model for typical application scenarios of convolutional neural network. It explores data and model computing methods supporting homomorphic encryption to protect data privacy, builds methods of service certificate storage and calculating equity interests distribution based on blockchain and smart contract technology to ensure the openness, transparency, credibility and traceability of service computing. A novel resource and data service paradigm of cloud environment is explored for resource providers, model owners and users to promote the effective integration of resources and develop sharing economy. Finally, the privacy protection method in the model is analyzed through experiments.

**Key words:** homomorphic encryption; privacy protection; smart contract; convolutional neural network

## 1 引言

当前,人类社会已经步入了移动互联网时代,智能计算、移动便捷以及隐私安全成为重要的发展趋势. 如何能在保护用户隐私信息的前提下,加强移动终端的计算能力,提高高智能计算服务体验,是一个亟待解决的问题.

在这种背景下,利用云服务来完成人工智能计算的模式出现了,它既可解决边缘设备计算力不足的问题,又可发挥移动特性. 以基于卷积神经网络(Convolutional Neural Network,CNN)的图像分类为场景,探索新模式应用的关键技术,目前面临着以下两个方面的挑战.

(1)用户隐私保护是应用的重要前提. 2018年欧盟制定的《通用数据保护条例》[1](General Data Protection Regulation,GDPR)提出加强对个人数据在隐私和安全方面的保护. 用户终端数据涉及大量用户信息,将其直接发送到云端缺乏安全保障,具有泄露风险. 云服务商

收稿日期:2020-07-16;修回日期:2021-03-30;责任编辑:孙瑶
基金项目:国家重点研发计划(No.2018YFB1402703);北京市教育委员会科技计划一般项目(No.KM201910037003);北京工商大学农产品质量安全追溯技术及应用国家工程实验室开放课题(No.AQT-2020-YB5);北京市社会科学基金研究基地项目(No.18JDGLB026);北京物资学院2020年度"实培计划"项目

也容易过度使用这些数据或私自销售,谋取利益[2]. 用户数据隐私保护是安全计算外包模式的基本要求[3].

(2)传统云服务是由云供应商控制和维护的,包括服务和权益规则,以及交易和服务数据,缺乏有效的共同参与和管理机制,约束力和透明度不足,出现纠纷难于追责. 同时,容易产生大的云服务商垄断、小的云服务商难以生存的现象,不利于市场良性发展和资源的有效整合.

为了应对上述挑战,本文研究面向CNN的区块链可信隐私服务计算模型. 以典型应用为场景,利用同态加密技术,在保护用户隐私的前提下,有效利用计算资源为边缘设备提供算力服务,智能合约和区块链可加强服务权益管理的公开透明.

## 2 相关工作

本节从同态计算、云计算隐私保护、卷积神经网络隐私保护等方面展开分析.

(1)同态加密技术

1978年Rivest首次提出同态加密的概念,即对密文进行运算的结果与对明文进行相应运算的结果是等效的. 无需解密,通过处理密文即可获得需要的计算结果,这是数据隐私保护的重要手段,具有重要意义[4~6].

2009年Gentry[7]提出了基于理想格的全同态加密方案,复杂度高的限制造成密文数据扩张问题不能有效解决,影响实际应用. Van Dijk等人[8]使用基本的模运算设计了同态加密方案(Dijk Gentry Halevi Vaikutanathan,DGHV),该方案是对文献[7]整数上的全同态加密算法的改进,使计算复杂度降低、效率提高、易于实现,一次加密1 bit的数据,其公钥加密方案的安全性依赖"近似最大公约数"问题.

Coron等人[9]针对DGHV方案公钥尺寸过大的问题,提出一种基于平方公钥压缩的CMNT(Coron Mandal Naccache Tibouchi)公钥优化方案. 其思想是使用公钥集合中非初始元素的$2k$个公钥可生成$k^2$个公钥,压缩公钥尺寸. 首先将$2k$个公钥平均分成两组,然后分别从两组公钥中随机选择一个公钥对应相乘,再乘以随机数,并运算处理生成新的公钥,进而完成加密操作.

另外,文献[10]改进了文献[8]的方案,使其一次可以加密2 bit的数据. 孙霓刚等人[11]进一步改进了DGHV算法,将明文空间由1 bit扩展到$n$ bit,提出一次可以加密$n$ bit数据的方案,降低了加密次数. 为了清晰说明,本文将这种算法称为N-DGHV. N-DGHV算法通用性强,适合服务计算隐私保护场景,但仍存在公钥存储空间过多的问题. 因此,本文在N-DGHV算法的基础上,对公钥进行压缩优化,并加以实现.

(2)云计算隐私保护研究

传统的云计算模式中,终端数据以明文的形式传输到云端进行计算,用户隐私无法得到保障,存在安全隐患[12].

云计算的数据隐私保护解决方案主要有访问控制、数据加密、安全外包、安全多方计算等[13],都是基于数据加密理论展开的. 蒋瀚等人[14]提出了安全多方计算方法来解决云计算隐私保护问题,该方法需要多方参与计算,且通信频繁,不适合本文客户端资源有限的应用场景. 文献[15]使用基于混淆方法的隐私管理器来管理云端和用户终端的数据,保护数据隐私,但重在加密管理,未深入研究密文的智能计算等工作.

在云计算场景中,随着数据量的增大,频繁的加解密操作会造成计算资源的浪费,能够直接对密文进行计算操作显得尤为重要.

同态加密技术以其良好的密文可操作性,成为解决云计算隐私保护问题的重要技术,也成为重要发展和应用方向[16,17]. 而文献[18]也提出同态加密技术在处理大量数据时效率不高. 本文正是致力于探索该技术适合智能计算应用场景的解决方案.

(3)卷积神经网络数据隐私保护相关研究

卷积神经网络是深度学习的重要分支,计算复杂度高,被广泛应用于人脸识别、语音识别等领域[19]. 卷积神经网络的隐私保护工作可在不同阶段进行,分别是训练阶段和预测阶段. 在训练阶段,需要各参与方提供各自的数据来完成模型训练工作,这些数据可能包含隐私信息. 在预测阶段,终端用户待预测数据、服务器端训练好的特征模型都有隐私保护的需求. 本文主要针对卷积神经网络的预测阶段展开工作,使用同态加密技术.

Dowlin等人[20]于2016年提出CryptoNets神经网络模型,使用同态加密算法实现卷积神经网络预测阶段的隐私数据保护. 为了保证同态加密的正确性,此方案化简了预测方法,使用平方函数实现激活层. Chabanne等人[21]使用低次多项式逼近激活函数来加强同态加密计算效率. 文献[22]在CryptoNets基础上优化,同时提出云端双服务器协同模式使加同态算法支持CNN各层模型,提高效率. 文献[23,24]则提出将两方计算技术(混淆电路)与同态加密相结合的解决方案,同态加密处理CNN线性部分,而两方计算则处理非线性部分.

目前,这一领域并不成熟,仍旧存在加密计算开销大、智能算法不适用、管理模式不清晰等问题,还远不能被广泛应用,需要更多实践去探索应用方法和模式.

(4)其他相关研究

区块链具有分布式管理、难以篡改的特点[25],可被广泛应用于医疗、交通、农业等多个领域. 智能合约通常是运行于区块链上的公开透明的计算代码. 本文利用区块链存储服务数据,并设计智能合约权益评估模型,公开透明并自动执行,加强交易的可信度.

联邦学习技术是2016年被提出的,该技术在保证数据隐私安全的前提下,用来完成高效智能计算工作. 联邦学习的过程是将用户数据在本地计算,并将结果传输到服务器参与聚合计算,从而起到保护隐私数据的目的[26]. 然而,联邦学习模式数据提供者在本地进行模型训练,对终端环境的算力要求较高. 因此,本文依据应用场景,选取同态加密方法展开研究工作.

## 3　可信隐私服务计算模型

为了提高云服务环境下卷积神经网络预测服务质量,本文从安全、隐私和可信3个方面考虑,研究可信隐私服务计算模型. 该模型使用非对称的公钥加密、私钥解密机制来加强数据的安全性,避免被恶意截取;同时,通过密文传输,利用同态加密的特点在服务端进行密文计算,将密文结果反馈用户,整个计算过程全密态,保护数据提供者隐私信息;最后使用区块链和智能合约技术完成云服务计算过程的记录,并进行自动权益分配,保证服务过程不可篡改,公开透明,从而增强服务计算模型的可信度.

### 3.1　可信隐私服务计算架构

可信隐私服务计算架构如图1所示,主要可分为用户端、模型提供端、云服务器端3类角色,围绕计算、加密、可信权益等工作运转.

(1)用户端

用户端是服务计算的使用者,拥有数据以及公私钥生成器. 用户端要向云服务端提出需求,请求服务,并取得相应权限(如认证、开通账户),进而开始整个服务流程. 首先,用户端生成公私钥,并将公钥发送给云服务端,如图1中①所示;其次,在本地将数据通过公钥加密,密文上传云服务端,如图1中④所示;再次,用户端得到云服务端提供的密文运算结果及分类标签,如图1中⑤所示,并在本地通过私钥解密进而得到最终结果;最后,用户端收到云服务端的权益分配结果,并提交服务费用,如图1中⑥所示.

(2)模型提供端

首先,模型提供端需要从云服务端获取加密公钥(由用户端提供给云服务端),如图1中②所示;其次,模型提供端将训练好的预测模型使用公钥加密后提供给云服务端,同时需要提供分类标签(无须加密,各个分类在结果向量中的顺序),如图1中③所示;最后,计算服务完成后,模型提供端获得权益分配结果,取得相应费用,如图1中⑦所示.

(3)云服务端

云服务端提供强大的计算资源和模型服务,完成用户端的请求. 首先,云服务端接收用户端公钥,并将其发送给模型提供端加密预测模型,如图1中①与②所示. 其次,若使用云服务端自有的预测模型,则直接加密. 否则,云服务端接收模型提供端的加密模型以及用户端提供的加密数据,如图1中③和④所示,进行密文的卷积神经网络计算,将密文结果返回给用户端. 隐私服务计算过程完毕. 此外,在计算过程中,云服务端计算资源使用及服务提供情况,连同云提供商信息提交区块链存证,并使用区块链智能合约实现权益计算模型并自动执行,分配云服务端、用户端、模型提供端各自的费用和收益. 通常是用户端付费,云服务端和模型提供端获利.

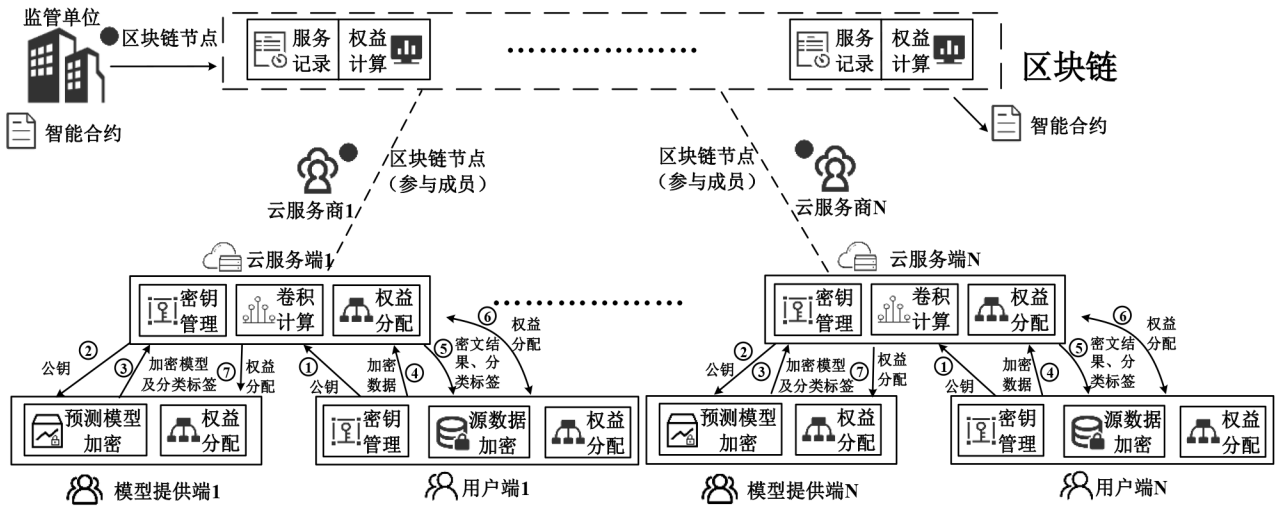



图1　可信隐私服务计算架构图

如图1所示,实际中存在多个云服务商,提供不同的模型和服务.数据拥有者计算资源不足,选择合适的云服务商并借助其算力获得预测结果,但又要保护数据隐私.模型提供者(也可以是云服务商)在保护模型内容的前提下分享模型并获利,同态加密技术在此流程中起到保护数据和模型隐私的作用.另外,这种模式下,可信的运行环境和权益管理机制是破除垄断、提高服务质量的重要保证,区块链和智能合约技术恰能发挥作用.计算资源使用、服务提供情况以及云服务商信息都存证区块链系统,不可篡改,智能合约计算权益分配的规则透明公开、自动执行,并且可查询、追责.另外,模型提供者参与计算过程同样存证区块链系统.这样,可达到模型权属清晰、服务权责透明、权益公平可信的效果.同时,还存在一个监管单位的角色,监管单位可查看全部存证数据和使用规则,有效约束不良行为.

### 3.2 预测服务隐私计算模型

在此应用场景中,系统和用户可根据需求选择不同的同态加密算法.本文同态加密是通过改进N-DGHV算法来实现的,这里称为ON-DGHV(Optimized N-DGHV)算法.DGHV算法的明文空间是$\{0,1\}$(二进制表示).N-DGHV算法通过将加密算法的随机数乘2变换为乘$2^n$,解密算法的模2变成模$2^n$,实现明文空间由1 bit扩大到$n$ bit,减少了加密次数.进一步,ON-DGHV使用平方公钥压缩方法缩减公钥尺寸,减少公钥的存储空间.此算法是面向整数的同态加密算法,将明文加密进行密文计算,会给智能计算的性能、精度乃至正确性带来挑战.该模型努力探索适合该同态算法的CNN预测方法,运行流程如图2所示.

模型数据包括用户端提供的数据矩阵$\boldsymbol{D}$和模型提供端提供的模型$\boldsymbol{M}$(模型$\boldsymbol{M}$包括卷积核$\boldsymbol{K}$、卷积偏移量$\boldsymbol{b}1$、全连接矩阵$\boldsymbol{W}$和全连接偏移量$\boldsymbol{b}2$).

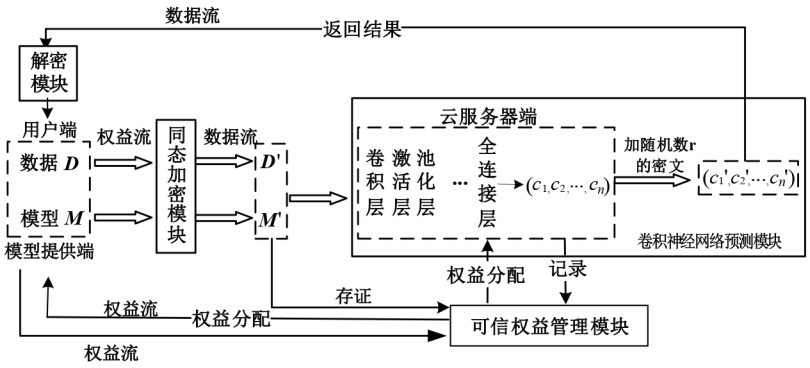



图2 卷积神经网络预测服务计算模型流程图

模型组件可分为4个功能模块,具体如下.

(1)同态加密模块:对用户端的原始数据矩阵$\boldsymbol{D}$使用公钥进行加密得到$\boldsymbol{D}'$,对模型提供端提供的模型$\boldsymbol{M}$使用公钥进行加密得到$\boldsymbol{M}'$(加密后的模型$\boldsymbol{M}'$包括卷积核$\boldsymbol{K}'$、卷积偏移量$\boldsymbol{b}1'$、全连接矩阵$\boldsymbol{W}'$及全连接偏移量$\boldsymbol{b}2'$).

(2)卷积神经网络预测模块:加密后的模型$\boldsymbol{M}'$和同态加密后的数据矩阵$\boldsymbol{D}'$成为卷积神经网络预测模块的输入.进一步,卷积神经网络预测模块各层次的关系和功能操作如图3所示.卷积核$\boldsymbol{K}'$和数据矩阵$\boldsymbol{D}'$作为卷积层的输入,在卷积层利用卷积核$\boldsymbol{K}'$对数据矩阵$\boldsymbol{D}'$进行卷积计算,得到一组线性输出**conv**;**conv**在激活层使用激活函数完成非线性映射操作,为了适应同态密文要求,这里激活函数使用平方函数进行计算,生成密文数据**acti**;然后将**acti**通过池化层进行加和池化[20],完成数据压缩,减少数据量,以简化计算的复杂度,进而输出数据**pool**;最后将数据**pool**和全连接矩阵$\boldsymbol{W}'$放入全连接层进行矩阵乘法,将上层的特征映射到样本空间来实现分类,所有类别中值最大的一类即为CNN的识别结果,表现为密文结果$\boldsymbol{C}$.为了保护数据模型的隐私性,该模型根据需求可增加一个保护机制.那就是将密文$\boldsymbol{C}$中各个元素都加上一个随机数$r$的密文态($r$尽量选取较小的数,使用同一加密算法),随机数的密文可表示为$C_r$,加和之后得到$\boldsymbol{C}'$(也就是Lock($C$)函数),即$\boldsymbol{C}'=C_r+\boldsymbol{C}$,随后将$\boldsymbol{C}'$连同分类标签(各个分类在结果向量中的顺序)发送给用户端.

从图3中可知,卷积层、激活层和池化层、全连接层各层之间存在前后级联关系,前一层的输出作为后一层的输入,是一个有机整体,共同完成密文数据的计算,有效提取数据特征,完成预测功能.

(3)解密模块:用户利用私钥对获得的密文结果$\boldsymbol{C}'$进行解密得到$\boldsymbol{T}'$,根据分类标签获得$\max(\boldsymbol{T}')$对应的分类结果,$\max(\boldsymbol{T}')$是分类标签中的最大值,即预测结果.由于加密算法满足加法同态性,且最终是根据结果向量中元素的数值大小判断分类结果,因此即使卷积神经网络预测模块的结果加了$C_r$随机数,对最终的分类

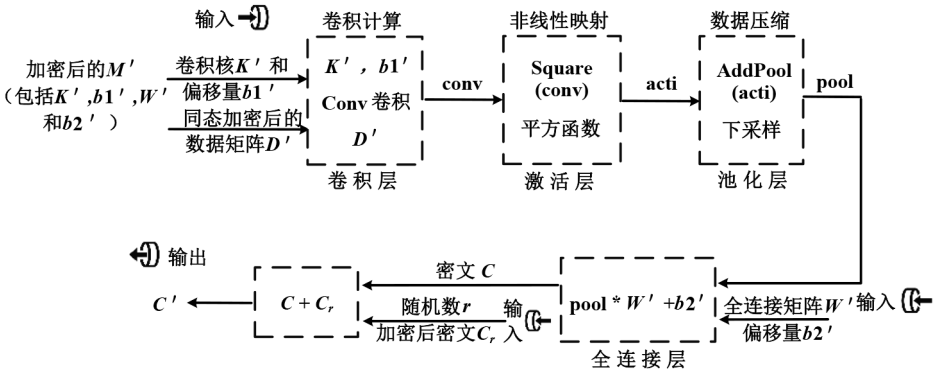



图3　卷积神经网络预测模块各层次关系示意图

结果没有影响. 然而,这样会影响分类概率,云端可根据需求设置,解密模块同样对于密文$\boldsymbol{C}$适用.

另外,如图2所示,该模型还存在一个可信权益管理模型,在模型提供端提供模型后用智能合约进行存证;记录云服务端的服务明细;根据权益评估模型对用户端、模型提供端和云服务端进行权益分配. 整个流程主要由数据和模型的数据流以及权益交易流两个信息流组成,相辅相成,涉及模型各个环节,协同工作.

模型输入输出及算法涉及各个模块,具体如下.

(1)同态加密模块

同态加密模块输入:用户端的原始数据矩阵$\boldsymbol{D}$. 模型提供端提供的模型$\boldsymbol{M}$.

同态加密模块输出:原始数据矩阵$\boldsymbol{D}$加密后的密文矩阵$\boldsymbol{D}'$及加密后的模型$\boldsymbol{M}'$(分别在用户端和模型提供端).

同态加密模块涉及的核心函数描述如下.

(a)GenKey(·):密钥生成函数. 输出用户的私钥SK,公钥集合$\mathrm{PK}=\left\{\mathrm{pk}_f,\mathrm{pk}_0,\mathrm{pk}_1,\cdots,\mathrm{pk}_{2k-1}\right\}$,下标$k$为正整数.

(b)Encrypt(PK,$\boldsymbol{D}$),Encrypt(PK,$\boldsymbol{M}$):加密函数. 输入公钥集合PK,将PK中的非初始的$2k$(即$\mathrm{pk}_0,\mathrm{pk}_1,\cdots,\mathrm{pk}_{2k-1}$)个公钥平均分成两组,然后分别从两组公钥中各随机选择一个公钥相乘,重复这个过程$a$($0<a\leqslant k^2$)次,最后把这$a$次相乘的结果加和得到$S'$,将$S'$作为参数,对$\boldsymbol{D}$和$\boldsymbol{M}$进行加密($S'$,$\boldsymbol{D}$)和($S'$,$\boldsymbol{M}$)操作,输出加密后的密文矩阵$\boldsymbol{D}'$和$\boldsymbol{M}'$.

(2)卷积神经网络预测模块

卷积神经网络预测模块输入:加密后的数据矩阵$\boldsymbol{D}'$和加密后的模型$\boldsymbol{M}'$.

卷积神经网络预测模块输出:返回给用户端的结果矩阵$\boldsymbol{C}'=\left[\boldsymbol{c}_1',\boldsymbol{c}_2',\cdots,\boldsymbol{c}_i'\right]$和分类标签$l=\left\{l_1,l_2,\cdots,l_i\right\}$.

卷积神经网络预测模块涉及的算法描述如下.

(a)卷积层

Convolve($\boldsymbol{K}'$,$\boldsymbol{D}'$):卷积函数. 在卷积神经网络典型应用图像处理的场景下,卷积层的作用是提取图像的特征. 主要的操作是密文图像矩阵$\boldsymbol{D}'$与卷积核$\boldsymbol{K}'$做卷积计算. 设$\boldsymbol{D}'$的高和宽分别为$H_d$和$W_d$,$\boldsymbol{K}'$的高和宽分别为$H_k$和$W_k$,图像$\boldsymbol{D}'$和卷积核$\boldsymbol{K}'$的深度为depth,填充的像素数为$P$,步长为$S$,则可以推出结果矩阵的高和宽分别为$H_c=\left(H_d-H_K+2P\right)\!/S+1$,$W_c=\left(W_d-W_K+2P\right)\!/S+1$,结果矩阵元素卷积计算公式为$\mathbf{conv}_{i,j}=\sum_{d=0}^{\mathrm{depth}-1}\sum_{m=0}^{H_k-1}\sum_{n=0}^{W_k-1}\boldsymbol{D}'_{d,m+i,n+j}\boldsymbol{K}'_{d,m,n}$,其中,$0<i<H_c$,$0<j<W_c$. 此卷积计算公式可以看成多个向量的内积运算,并转换成矩阵乘法:把卷积核$\boldsymbol{K}'$看作高为1、宽为$H_k$*$W_k$*depth的矩阵(若多个卷积核,就会增加高度值),把图像矩阵$\boldsymbol{D}'$的多个卷积窗口组合成向量(不同深度的同一位置组合成一个向量),继而组合成一个高为$H_k$*$W_k$*depth、宽为$H_c$*$W_c$的矩阵,两个矩阵相乘加上偏移量$\boldsymbol{b}1$即得到卷积运算结果.

卷积层涉及的是加和乘的基本运算,满足同态加密的要求,因此加密数据的卷积运算只需要将$\boldsymbol{D}'$和$\boldsymbol{K}'$转换成矩阵进行运算得到结果**conv**. 由于是矩阵运算,在此过程中,可利用并行化技术进行优化,加快运行效率.

(b)激活层

Activate(**conv**):平方函数. 激活层的作用是为CNN提供非线性特征,常用的激活函数(如ReLU,Sigmoid等)需要最大值、除法、指数等运算,不适合使用同态加密方法密文的加和乘法实现. 本文选用CryptoNets[20]使用的平方函数方法加以取代,计算结果为**acti**.

(c)池化层

Pooling(**acti**):池化层主要的作用是下采样,对输入的特征图进行压缩,进一步减少参数数量,简化网络计算复杂度,提取主要特征. 池化的方法很多,为了更好的支持同态加密计算,这里采用加和池化的方法[20],得到结果**pool**.

(d)全连接层

Connect($\mathbf{pool}, \boldsymbol{W}'$):全连接层作用是将上层的特征映射到样本空间,从而实现分类.$\boldsymbol{W}'$为全连接矩阵,将池化层的输出**pool**矩阵转换成向量,即可把全连接层视为矩阵乘法,从而计算$\boldsymbol{C}=\boldsymbol{W}'*\mathbf{pool}+\boldsymbol{b}2$得到结果向量$\boldsymbol{C}$.$\boldsymbol{C}$中的值代表分类标签$l=\{l_1,l_2,\cdots,l_i\}$中对应类别的数值(数值越高,则预测结果是该类别的可能性越大),为密文.

(e)安全处理

Lock($\boldsymbol{C}$):生成随机数$r$(选择尽量小的数),加密得到密文形成$C_r$,利用$C_r$对全连接得到的结果向量$\boldsymbol{C}$进行加密得到$\boldsymbol{C}'$再发送给用户,从而使用户不会得到原始的模型输出结果,减少模型参数泄露的风险.云端可根据实际需求设置输出原始结果,这一部分为可选项.

(3)解密模块

解密模块输入:云服务端返回的密文结果向量$\boldsymbol{C}'$($\boldsymbol{C}$同样适用)和分类标签$l$.

解密模块输出:卷积神经网络预测分类结果$\boldsymbol{T}$.

解密模块涉及的算法描述如下.

(a)Decrypt($\boldsymbol{C}'$,SK):解密函数.$\boldsymbol{C}'=[\boldsymbol{c}_1',\boldsymbol{c}_2',\cdots,\boldsymbol{c}_i']$为云服务端返回的密文结果向量.利用私钥SK进行解密得到明文结果向量$\boldsymbol{T}'=[t_1',t_2',\cdots,t_i']$.

(b)$\boldsymbol{T}=p(\max(\boldsymbol{T}'),l)$:$\max(\boldsymbol{T}')$为集合$\boldsymbol{T}'$中的最大值,即分类结果的数值.分类标签$l$和向量$\boldsymbol{T}'$存在一一映射的关系,$\boldsymbol{T}=p(\max(\boldsymbol{T}'),l)$代表$\max(\boldsymbol{T}')$在$l$中的映射,即为卷积神经网络的预测分类结果.

### 3.3 预测服务权益评估模型

传统服务计算的权益规则由云供应商(云服务端)制定,缺乏透明度和公共约束力,这样,云供应商的权利过大,服务使用者(用户端和模型提供端)的权益得不到有效保证,进而不愿意参与云端服务计算.因此,本模型设计基于智能合约的权益计算模型来进行权益分配,此过程由区块链智能合约自动执行并进行数据的存储.一方面,在模型提供端提供模型时,智能合约对模型的所属权进行记录存证,保证模型提供端的权益.另一方面,智能合约计算权益分配,模型提供端和云服务端根据权益分配的结果获取相应的收益,而用户端通常向云服务端提供相应的费用,收益规则公开透明,保证过程可追溯、权益评估真实可信.

以卷积神经网络的图像分类场景为例,结合传统云服务的计费思路,可设计智能合约的基础权益分配方法,如表1所示.从模型、资源使用、数据角度考虑,具体参数包括模型准确度、使用时长、服务费、存储容量、收益等.

**表1 权益评估计算参数表**

| 参数收益 | 模型准确度$e$ | | | 使用时长$t$ | 时间服务收益$s$ | 时间$i$时存储容量$z_{i_z}$单位容量收益$z$ | 数据量 | |
|---|---|---|---|---|---|---|---|---|
| | <80% | 80%~95% | >95% | | | | 图片大小$w\times h$(默认$r_1\times r_2$) | 通道数$c$(默认$k$) |
| 模型提供端 | $x_1$ | $x_2$ | $x_3$ | 0 | | 0 | 0 | 0 |
| 云服务端 | 0 | 0 | 0 | $t\times s$ | | $z\times\sum_0^t z_i$ | $y_1+(w\times h-r_1\times r_2)\times p_1$ | $y_2+(c-k)\times p_2$ |

(1)模型提供端

模型提供端的收益依赖其提供的模型的准确度,分为3个标准,即小于80%、80%~95%、大于95%,能够获得的收益分别为$x_1,x_2,x_3$.($x_1<x_2<x_3$).这部分由区块链智能合约实现,通常智能合约不支持浮点数,故将模型准确度$e$扩大$N$倍变成整数后传入智能合约计算,这里以100为例,即$e\leftarrow e\times 100$,$e$的范围变成$0\leqslant e\leqslant 100$.当$e<80$时,模型提供方收益为$x_1$;当$80\leqslant e\leqslant 95$时,收益为$x_2$;当$95<e\leqslant 100$时,收益为$x_3$.

(2)云服务端

云服务端的收益根据时间、存储容量、数据量等费用标准来计算.设模型时间为$i$时存储容量为$z_i$(GB)、使用时长为$t$(min)、每分钟服务收益为$s$、每GB存储容量的收益为$z$.图片大小为$w\times h$,默认大小为$r_1\times r_2$,默认收益为$y_1$,每超出一像素加收的费用为$p_1$,即图片大小的收益为$y_1+(w\times h-r_1\times r_2)\times p_1$;通道数为$c$、默认为$k$,默认收益为$y_2$,每多一通道加收的费用为$p_2$,即通道数的收益为$y_2+(c-k)\times p_2$.另外,存储容量的总收益为$z\times\sum_0^t z_i$,总时长服务收益的计算方法是$t\times s$.所以云服务商的收益总和为$z\times\sum_0^t z_i+t\times s+y_1+(w\times h-r_1\times r_2)\times p_1+y_2+(c-k)\times p_2$.模型提供端收益和云服务端收益总和即为用户端的总费用(暂不考虑区块链系统收益,该模型各个指标均取整数以适应智能合约的特点).模型提供端的收益仅与模型本身有关,一次性收取.而对于云服务端,智能合约则以分钟为单位对用户使用的云资源进行权益计算.

### 3.4 实现方法

上述的卷积神经网络预测计算模型是以ONDGHV算法为加解密基础的,这里重点描述该算法的设计与实现.为了表述清晰,部分符号表达和前文概述有

所调整.

(1)同态加密模块

这一模块的工作是生成密钥对、加密明文$m$. 生成私钥SK时,需要保证$|m+2^n r'|<\mathrm{SK}/2$,其中$n$为$m$的比特(位)数,$r'$为加密时的随机正整数,私钥SK为随机大素数. 生成的公钥集合为PK,集合中元素的个数为$2k+1$($k$为正整数). 集合中的第一个元素表示为$\mathrm{pk}_f=q_f\times \mathrm{SK}$,其中$q_f$为随机正奇数,建议取大一些的数. 集合中的后续元素$\mathrm{pk}_{i,j}(0\leqslant i\leqslant 1,0\leqslant j\leqslant k-1)$生成描述如下:随机生成正整数$r_{i,j}$,要求$r_{i,j}>r'$时,生成区间$[0,q_f)$的随机整数$q_{i,j}(0\leqslant i\leqslant 1,0\leqslant j\leqslant(k-1))$,计算$\mathrm{pk}_{i,j}=r_{i,j}+\mathrm{SK}\times q_{i,j}(0\leqslant i\leqslant 1,0\leqslant j\leqslant(k-1))$. 生成密钥对算法如算法1所示.

**算法1　生成密钥对**

输入:明文$m$,$m$的比特(位)数$n$,正整数$k$(公钥包含的元素个数为$2k+1$)
输出:私钥SK,加密公钥PK

```
1. function GenKey(m,n,k)
2.   generate random positive integer numberr'
3.   do
4.     generate random large prime number x
5.   while |m+2^n r'| ≥ x/2
6.   SK ← x
7.   generate random positive odd number q_f
8.   pk_f = q_f × SK
9.   put pk_f into PK
10.   for i from 0 to 1 step 1
11.     for j from 0 to k−1 step 1
12.       do
13.         generate random positive integer number r_{i,j}
14.       while r_{i,j} ≤ r'
15.       do
16.         generate random integer number q_{i,j}
17.       while q_{i,j} < 0 or q_{i,j} ≥ q_f
18.       pk_{i,j} ← r_{i,j} + SK × q_{i,j}
19.       put pk_{i,j} into PK
20.     end for
21.   end for
22.   return SK,PK
23. end function
```

得到公钥后,下面以明文为正整数或零为例对加解密过程进行描述,负整数的处理方法类似,只是需要依据求模操作在不同编程语言中的具体含义来加以处理. 加密过程是首先将PK中的非初始元素的$2k$个公钥平均分成两组,每组元素个数为$k$,然后分别从两组公钥中各随机选择一个公钥相乘,重复进行$a(0<a\leqslant k^2)$次,最后把这$a$次相乘的结果加和得到sum值,即$\mathrm{sum}=\mathrm{sum}+\mathrm{PK}[0][e_1]*\mathrm{PK}[1][e_2]*b$,其中$0\leqslant e_1,e_2\leqslant k-1$,$b$为随机生成的正整数,使用$\mathrm{PK}[X][Y]$这种方法来表示两个分组的公钥元素. 同时生成随机正整数$r$. 计算密文$c=(m+2^n\times r+2^n\times \mathrm{sum})\bmod \mathrm{pk}_f$($n$为明文$m$的比特位数,$\mathrm{pk}_f$为PK的第一个元素). 对应前文描述的模型,sum即为$S'$,$m$即为$D$或者$M$,$c$则是$D'$或者$M'$. 加密算法如算法2所示.

**算法2　数据加密**

输入:明文$m$,位数为$n$,公钥PK,正整数$k$(公钥包含的元素个数为$2k+1$)
输出:加密密文$c$

```
1. function Encrypt(PK,m,k)
2.   do
3.     generate random positive integer number a
4.   while a > k^2
5.     sum ← 0
6.   for i from 0 to a − 1 step 1
7.     do
8.       generate random positive integer number e_1
9.       generate random positive integer number e_2
10.     while e_1 > k or e_2 > k
11.   generate random positive integer number b
12.   sum ← sum + PK[0][e_1]*PK[1][e_2]*b
13.   end for
14.   generate random positive integer number r
15.   c = (m + 2^n × r + 2^n × sum) mod pk_f
16.   return c
17. end function
```

(2)解密模块

对于正整数的密文结果$c$,用户可使用私钥SK对$c$根据$T=(c \bmod \mathrm{SK}) \bmod 2^n$公式进行解密得到明文结果$T$. 负整数密文解密的处理方法类似.

另外,关于基于同态加密的CNN智能计算,不管是加密过程还是密文计算过程,都会涉及大量的循环计算,因此可以使用并行化提高效率,推动密文计算的实际应用.

## 4　实验与分析

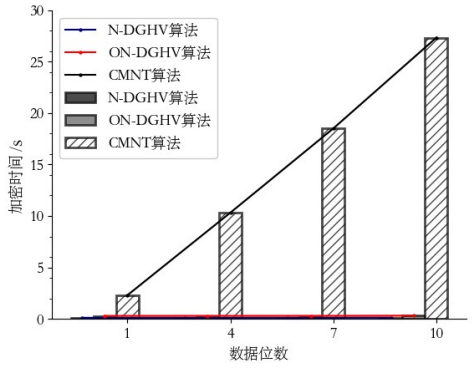
本节对隐私服务计算模型进行实验分析,主要分为算法对比和密文预测分析两部分. 实验环境是使用Dell PowerEdge R740服务器,CPU是Intel Xeon Silver 4116@2,内存64 GB,固态硬盘960 GB,机械硬盘3.6 TB,使用Python语言开发.

### 4.1　算法对比实验

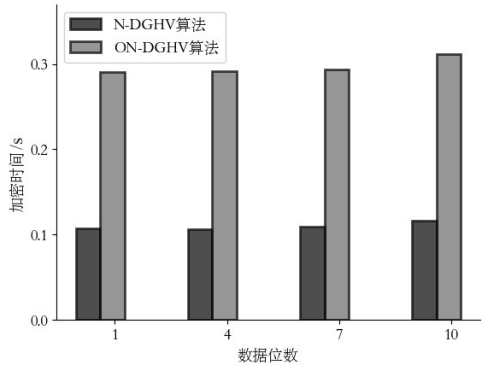
实验主要将本文使用的ON-DGHV算法与N-DGHV算法以及CMNT算法(1 bit)进行比较. 3种算法都使用非对称同态加密,实验从加密数据的个数和位数两个方面展开工作.

(1)加解密数据个数相同,位数不同

实验对象为10 000个数据,设定为常见的十进制数,改变数据位数,分别设计为1,4,7和10位,测试指标是加密和解密时间. 不同位数的10 000个数(数据是随机生成的)要进行50次实验并将测试指标的平均值作为结果,见图4和图5. 为了清晰表述,两个图均分为(*a*)和(*b*)两个子图,子图(*a*)是三种算法的变化对比,而子图(*b*)则是两类DGHV算法的时间对比. 随着数据位数的增多,ON-DGHV和N-DGHV算法的加/解密时间变化幅度都较小,基本保持不变,且当横坐标(位数)大于等于4时,加/解密的时间明显小于CMNT算法. ON-DGHV和N-DGHV算法一次可以加/解密$n$ bit数据(实际应用中需要预处理得到$n$),可直接对十进制数($n$ bit)进行一次加/解密处理,而CMNT算法一次只能加/解密1 bit数据(只支持0与1的加密),需将十进制数据转化为二进制,然后逐位进行加/解密操作,因此当数据位数较多时耗时更明显. 也就是说,CMNT加/解密时间依赖数据位数,呈类线性关系.



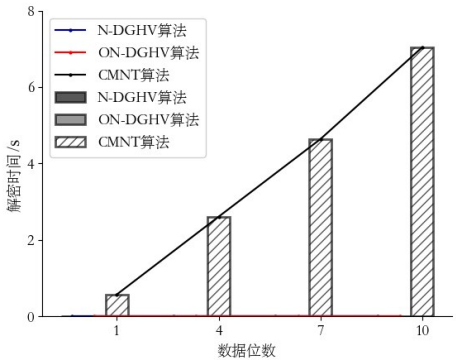

(*a*) 三种算法变化对比图



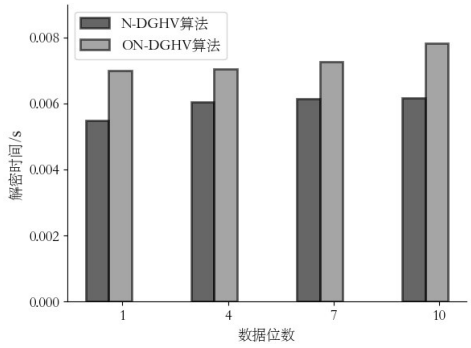

(*b*) 两类DGHV算法对比图

图4 加密时间随数据位数变化对比图

另外,ON-DGHV是在N-DGHV算法的基础上,使用了平方公钥压缩方法对公钥空间进行缩减,因此ON-DGHV比N-DGHV加密算法耗时会增多,但是通过实验




(*a*) 三种算法变化对比图




(*b*) 两类DGHV算法对比图

图5 解密时间随数据位数变化对比图

可得知区别不是很大,是可以接受的.

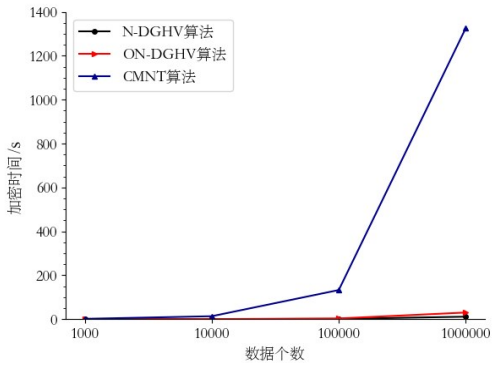
(2)加解密数据位数相同,个数不同

设定数据位数不变,都是5位的十进制数,加密数据个数(随机生成)分别是1000,10 000,100 000和1 000 000,经过50次实验求得测试指标的平均值. 测试指标仍旧是加/解密时间. 实验结果如图6和图7所示,同样,为了清晰展示,两图又分别划分为(*a*)和(*b*)两个子图,用来表示三种算法以及两类DGHV算法的对比. 可以看出,三种算法的加/解密时间都随加密数据个数的增加而增加,但是ON-DGHV和N-DGHV算法增长相对缓慢,CMNT算法增长明显.

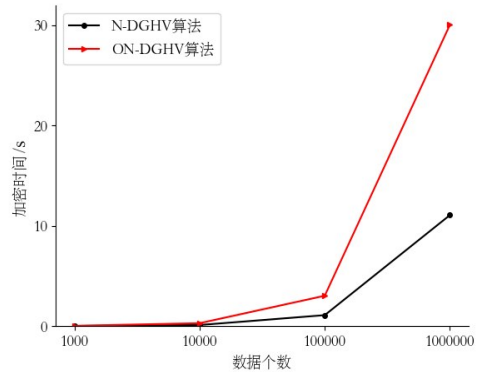
对ON-DGHV与N-DGHV、CMNT算法进行比较,总结如表2所示,可知ON-DGHV算法由于支持多位(bit)加密,加/解密的效率较高,同时使用公钥压缩方法使公钥空间减少,占据存储资源少,符合本文隐私保护服务应用场景.

### 4.2 密文预测分析实验

本节基于numpy库[27]实现前文提到的卷积神经网络计算模型,可分为以下8个部分:图像加密、模型加密、卷积层、激活层、池化层、全连接层、加随机数、解密. 这里使用ON-DGHV加密算法进行实验,并选用常见的MNIST手写数据测试集[28]. 该数据集是用于识别




（a） 三种算法变化对比图

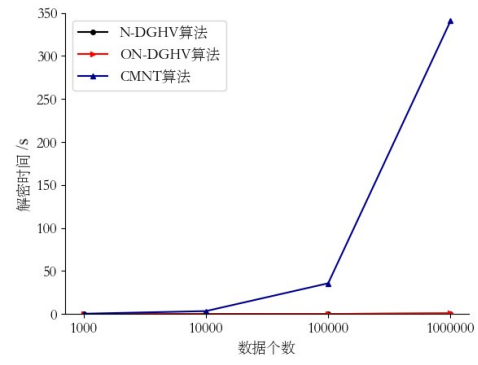



（b） 两类DGHV算法对比图

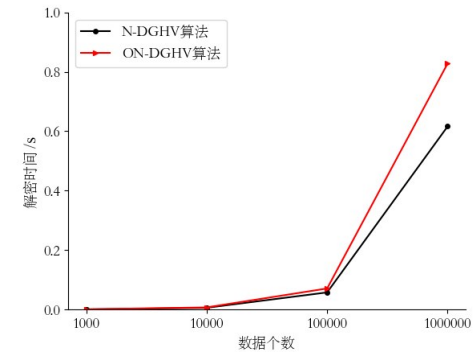
图6 加密时间随数据个数变化对比图

手写数字(图像分类)的,每个图片大小是28×28×1,即尺寸为28×28的灰度图,卷积核个数为6,步长为1.这里首先通过MNIST测试集中的60 000张图片训练模型,然后对10 000张样本图片进行预测操作,分别完成明文及密文预测(设置好支持预测图片数值的加密位数参数$n$)两类计算,再求各个图片预测准确度的平均值,得到明文是98.25%,密文是98.16%,密文对图片分类准确度的影响较小.进而分析明文和密文关于CNN各层的执行时间占比平均值,如表3所示,可以看出,密文计算对卷积层和全连接层的影响较为明显.

进一步,分析密文计算整体各部分的执行情况,如图8所示,最后两步操作,加随机数和解密,由于时间占比过小,在图中合并为一类.分析实验结果,密文计算增加的功能,即图像加密、模型加密、加随机数和解密占比41.9%.其中,模型加密耗时最多,为38.4%,是对卷积层和全连接层参数的加密操作.密文计算各层共占比58.1%,其中卷积层和全连接层则占比达到54.4%.从上述分析可知,卷积层和全连接层对密文计算效率影响大,这是后续应用中性能优化的重点部分.




（a） 三种算法变化对比图




（b） 两类DGHV算法对比图

图7 解密时间随数据个数变化对比图

**表2 同态加密算法对比表**

| 加密算法 | 加密时间 | 解密时间 | 支持多位(bit)加密 | 公钥尺寸 |
|---|---|---|---|---|
| CMNT算法 | 最慢 | 最慢 | 否 | 短 |
| N-DGHV算法 | 最快 | 最快 | 是 | 长 |
| ON-DGHV算法 | 较快 | 较快 | 是 | 短 |

**表3 明密文CNN各层执行时间占比对比表**

| 时间占比 | 卷积层 | 激活层 | 池化层 | 全连接层 |
|---|---|---|---|---|
| 明文 | 39.59% | 2.20% | 54.79% | 3.42% |
| 密文 | 66.49% | 5.84% | 0.49% | 27.18% |

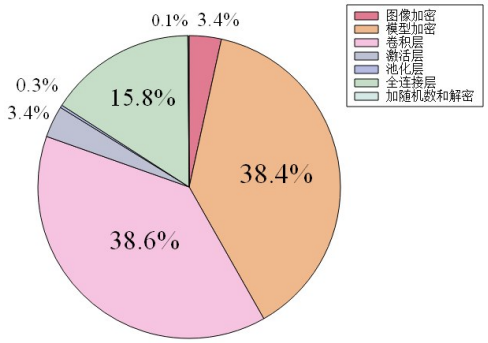



图8 密文卷积预测模型各部分时间占比图

## 5 结论

本文研究面向CNN的区块链可信隐私服务计算模型,使用ON-DGHV同态加密算法和区块链技术,以求加强服务计算中数据的安全、隐私保护和可信性,具有以下特点.

(1)提供了一套服务计算的解决方案,改善了服务计算和数据隐私保护之间的矛盾,在享受云服务计算便捷性的同时,保护用户隐私安全,有助于资源、数据的有效整合,以及推动新技术的应用和发展.

(2)探索同态加密和密文智能计算方法,并进行实践,寻求隐私服务计算的应用可行性.

(3)区块链和智能合约技术贯穿业务全过程,加强服务和交易的可信性. 模型共享、服务等过程上链存证,权益评估智能合约化,可增加规则透明度,权责可追溯,权益可保障,增强隐私计算服务的实用性. 同时,数据隐私保护技术也可增强区块链和智能合约的安全性,扩展其应用范围.

本研究还在初步探索阶段,主要针对基础的同态加密算法、CNN模型及数据集展开工作,后续还需要研究更多的加密和智能计算模型,实践更多的场景,根据场景研究提高准确度和效率的方法,并进一步挖掘区块链技术的融合机制,以求推动更深入、更广泛的应用.

### 参考文献


[1] 王雪乔. 论欧盟GDPR中个人数据保护与"同意"细分[J]. 政法论丛, 2019, (4): 136-146.
WANG X Q. Personal data protection and "consent" segmentationin EU GDPR[J]. Journal of Political Science and Law, 2019, (4): 136-146. (in Chinese)

[2] 陈丹伟, 邵菊, 樊晓唯, 等. 基于MAH-ABE的云计算隐私保护访问控制[J]. 电子学报, 2014, 42(4): 821-827.
CHEN D W, SHAO J, FAN X W, et al. MAH-ABE based privacy access control in cloud computing[J]. Acta Electronica Sinica, 2014, 42(4): 821-827. (in Chinese)

[3] SHAN Z H, REN K, BLANTON M, et al. Practical secure computation outsourcing[J]. ACM Computing Surveys, 2019, 51(2): 1-40.

[4] 蒋林智.(全)同态加密及其在云计算中的应用研究[D]. 成都: 电子科技大学, 2018.
JIANG L Z. Research on (Fully) Homomorphic Encryption and Its Application in Cloud Computing[D]. Chengdu: University of Electronic Science and Technology of China, 2018. (in Chinese)

[5] 谷思竹. 整数上全同态加密及其在云平台的应用研究[D]. 深圳: 深圳大学, 2016.
GU S Z. Research on Fully Homomorphic Encryption Over the Integers and Application in Cloud Platform[D]. Shenzhen: Shenzhen University, 2016. (in Chinese)

[6] 丁毅, 沈薇, 李洁, 等. 卫星通信全代理同态可信传输机制研究[J]. 中国空间科学技术, 2020, 40(4): 84-96.
DING Y, SHEN W, LI J, et al. Research on trusted full-proxy homomorphic transmission mechanism for satellite communication[J]. Chinese Space Science and Technology, 2020, 40(4): 84-96. (in Chinese)

[7] GENTRY C. Fully homomorphic encryption using ideal lattices[C]//STOC´09: Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing. New York: ACM, 2009: 169-178.

[8] VAN DIJK M, GENTRY C, HALEVI S, et al. Fully homomorphic encryption over the integers[C]//29th Annual International Conference on Theory and Applications of Cryptographic Techniques. Berlin: Springer, 2010: 24-43.

[9] CORON J S, MANDAL A, NACCACHE D, et al. Fully homomorphic encryption over the integers with shorter public keys[C]//Proceedings of Conference on Advances in Cryptology. New York: Springer-Verlag, 2011: 487-504.

[10] 林如磊, 王箭, 杜贺. 整数上的全同态加密方案的改进[J]. 计算机应用研究, 2013, 30(5): 1515-1519.
LIN R L, WANG J, DU H. Improved fully homomorphic encryption over integers[J]. Application Research of Computers, 2013, 30(5): 1515-1519. (in Chinese)

[11] 孙霓刚, 朱浩然, 汪伟昕. 一种适用于n bit的整数上全同态加密方案[J]. 计算机应用研究, 2018, 35(4): 1179-1181.
SUN N G, ZHU H R, WANG W X. Fully homomorphic encryption scheme applied to n bit[J]. Application Research of Computers, 2018, 35(4): 1179-1181. (in Chinese)

[12] 拱长青, 肖芸, 李梦飞, 等. 云计算安全研究综述[J]. 沈阳航空航天大学学报, 2017, 34(4): 1-17.
GONG C Q, XIAO Y, LI M F, et al. Summary of cloud computing security research[J]. Journal of Shenyang Aerospace University, 2017, 34(4): 1-17. (in Chinese)

[13] 杨健, 汪海航, 王剑, 等. 云计算安全问题研究综述[J]. 小型微型计算机系统, 2012, 33(3): 472-479.
YANG J, WANG H H, WANG J, et al. Survey on some security issues of cloud computing[J]. Journal of Chinese Computer Systems, 2012, 33(3): 472-479. (in Chinese)

[14] 蒋瀚, 徐秋亮. 基于云计算服务的安全多方计算[J]. 计算机研究与发展, 2016, 53(10): 2152-2162.
JIANG H, XU Q L. Secure multiparty computation in

cloud computing[J]. Journal of Computer Research and Development, 2016, 53(10): 2152-2162. (in Chinese)

[15] PEARSON S, SHEN Y, MOWBRAY M. A privacy manager for cloud computing[C]//IEEE International Conference on Cloud Computing. Berlin, Heidelberg: Springer, 2009: 90-106.

[16] CHAUHAN K K, Sanger A K S, Verma A. Homomorphic encryption for data security in cloud computing[C]// 2015 International Conference on Information Technology(ICIT). Bhubaneswar: IEEE, 2015: 206-209.

[17] MIN Z E, YANG G. Homomorphic encryption technology for cloud computing[J]. Procedia Computer Science, 2019, 154:73-83.

[18] GOEY J Z, LEE W K, GOI B M, et al. Accelerating number theoretic transform in GPU platform for fully homomorphic encryption[J]. The Journal of Supercomputing, 2021, 77(2): 1455-1474.

[19] 袁鹏. 图像卷积算法的隐私保护和应用研究[D]. 西安: 西安电子科技大学, 2017.
YUAN P. Privacy Preserving and Application of Image Convolution Algorithm[D]. Xi´an: Xidian University, 2017. (in Chinese)

[20] DOWLIN N, GILAD-BACHRACH R, LAINE K, et al. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy[C]//International Conference on Machine Learning. New York: ACM, 2016: 201-210.

[21] CHABANNE H, DE-WARGNY A, MILGRAM J, et al. Privacy-preserving classification on deep neural network [EB/OL]. (2017) [2020]. https://eprint. iacr. org/2017/035. pdf.

[22] 许世聪. 隐私保护卷积神经网络前向传播方法研究[D]. 西安: 西安电子科技大学, 2019.
XU S C. Research on Forward Propagation Method of Privacy-Preserving Convolutional Neural Network[D]. Xi' an: Xidian University, 2019. (in Chinese)

[23] 李少华. 云端卷积神经网络算法的安全增强机制研究[D]. 合肥: 中国科学技术大学, 2019.
LI S H. Research on Security Enhancement Mechanism for Convolutional Neural Network Predictions in Cloud [D]. Hefei: University of Science and Technology of China, 2019. (in Chinese)

[24] JUVEKAR C, VAIKUNTANATHAN V, CHANDRAKASAN A. Gazelle: A low latency framework for secure neural network inference[EB/OL]. (2018) [2021]. https://arxiv. org/abs/1801.05507.

[25] 贺海武, 延安, 陈泽华. 基于区块链的智能合约技术与应用综述[J]. 计算机研究与发展, 2018, 55(11): 2452-2466.
HE H W, YAN A, CHEN Z H. Survey of smart contract technology and application based on blockchain[J]. Journal of Computer Research and Development, 2018, 55 (11): 2452-2466. (in Chinese)

[26] 杨强. AI与数据隐私保护:联邦学习的破解之道[J]. 信息安全研究, 2019, 5(11): 961-965.
YANG Q. AI and data privacy protection: The way to federated learning[J]. Journal of Information Security Research, 2019, 5(11): 961-965. (in Chinese)

[27] NumPy Community. NumPy User Guide[EB/OL]. (2020) [2020]. https://numpy.org/doc/1.19/numpy-user.pdf.

[28] LECUN Y, BOTTOU L, BENGIO Y. MNIST[EB/OL]. [2020]. http://yann.lecun.com/exdb/mnist.

## 作者简介

**丁　毅**　男,1981年9月出生于河北省沧州市. 北京物资学院信息学院副教授,硕士生导师. 主要研究方向为区块链和智能合约技术、隐私计算等.

**沈　薇**　女,1999年9月出生于江苏省宿迁市. 北京物资学院本科生. 主要研究方向为同态加密和区块链技术.

**李　洁(通讯作者)**　女,1983年1月出生于北京市. 北京物资学院信息学院助理研究员. 主要研究方向为区块链、隐私计算及形式化验证等.