2023年全同态加密研究热点回眸

范瑞琦[1],陈铭志[1],牛鑫丽[2],董文阔[1],李晓霖[1],刘硕[1],刘静[1],赵明[3],蔡嘉跃[2],闫闱[1],朱树永[1],郑珂威[2],徐鹏[4],郝沁汾[1*],孙凝晖[1]

1. 中国科学院计算技术研究所,北京 100086
2. 无锡芯光同态科技有限公司,无锡 214104
3. 无锡芯光互连技术研究院有限公司,无锡 214104
4. 华中科技大学网络空间安全学院,武汉 430074

摘要 在大数据和人工智能时代,全同态加密方法被公认为是解决数据安全与隐私泄露问题的理想技术,但目前存在计算效率差、密文膨胀等问题,严重影响了全同态加密技术的应用和推广。从针对全同态加密算法的硬件加速和围绕全同态加密算法的优化2个方面,回顾了2023年计算机体系结构与密码学相关研究进展:以专用集成电路技术路线为代表的硬件加速效果明显;从算法角度进行优化,进展显著。可以预测,未来几年内,同态加密将与人工智能技术相结合,在跨行业、行业总分机构数据协作和利用中发挥更多价值。

关键词 全同态加密;数据安全;隐私泄露防护;计算机体系结构;密码学

完全同态加密(FHE)技术在不同领域展现出了强大的应用潜力。在区块链领域,FHE被应用于提升共识协议的安全性,Freitas等[1]提出了一种名为“Homomorphic Sortition Single Secret Leader Election”(SSLE)的协议,保护共识协议免受拒绝服务攻击,通过加密份额的排序进行领导节点选举,提高了系统的安全性和可靠性;在存储内容外包方面,Cong等[2]提出了Panacea ORAM设计,基于FHE的Oblivious RAM(ORAM)技术允许在不信任的服务器上安全存储和检索数据,保证了数据的隐私性;在通用应用程序执行方面,基于环上全同态加密(CGGI)方案的GPU加速框架ArctyrEX[3]被开发出来,以加速在加密状态下执行通用应用程序,使得性能得到显著提升,这对云计算环境中的数据安全具有重要意义,也使未来有可能出现一种新型的、在加密状态下运行的全同态计算机。在人工智

收稿日期:2023-12-31;修回日期:2024-01-08

作者简介:范瑞琦,博士研究生,研究方向为计算机体系结构与安全,电子信箱:fanruiqi20g@ict.ac.cn;郝沁汾(通信作者),研究员,研究方向为计算机系统结构,电子信箱:haoqinfen@ict.ac.cn

引用格式:范瑞琦, 陈铭志, 牛鑫丽, 等. 2023年全同态加密研究热点回眸[J]. 科技导报, 2024, 42(1): 286-295; doi:10.3981/j.issn.1000-7857.2024.01.018

能领域,FHE被应用于密态机器学习,Lam等[4]提出了一个基于2位查找表的单隐藏层算法,减少整数转换的调用次数,同时优化了神经网络结构,提高了面部识别和语音识别的效率和性能。Cheon等[5]提出了一种基于CKKS完全同态加密算法的深度卷积神经网络的评估方法,该方法适用于批量推理,引入了一种名为"Channel-by-Channel Packing"的打包方法,显著降低了ResNet推理的计算成本。在数据库管理方面,基于FHE的加密数据库管理系统HE3DB[6]被提出,它支持多种SQL查询和服务器端分析处理,有效解决了高计算延迟和查询处理能力的问题。这些应用共同证明了FHE技术在保护隐私和增强数据安全性方面的应用价值。

1　同态加密算法的原理

同态加密算法,是指在明文空间中的代数运算,在密文空间依然成立,简单来说,即存在让A+(或X)B=C,E(A)+(或X)E(B)=E(C)同时成立的E,E是指一种加解密算法,又叫同态加密算法。根据同态加密算法所支持的同态操作种类和次数,可以将现有的同态加密算法分为部分同态加密、近似同态加密、全同态加密3个层级。部分同态加密只支持单一同态加密操作,即加法或乘法;近似同态加密则可以同时支持加法和乘法同态操作,但执行次数有限,而完全同态加密可以支持无限制次数的加法和乘法同态操作。除了算术同态操作以外,还有搜索同态等操作,即可以支持在密态下的搜索,其核心操作为大小对比等非算数操作。由于部分同态、近似同态算法已经在某些领域获得了少量应用,本文主要针对全同态加密实现算术操作的研究进行梳理,后续如无特殊指明,均指全同态加密算法。

同态加密是一种具有特殊性质的加密方案。经过同态加密技术加密的密文,可以在不需要密钥方参与的情况下支持和明文等价的各类代数运算,同时由于同态加密本质上属于非对称加密,因此具有较强抵御传统密码破解技术的能力,是一种理想的用于保护数据安全和隐私的加密算法。

2　全同态加密算法的研究现状

2.1　全同态加密算法存在的问题

与部分同态、近似同态加密算法相比,全同态加密算法具有更广阔的应用前景,但是存在一些问题,影响了其应用和推广,主要包括高计算复杂度、频繁的降噪操作和密文膨胀3个方面。

全同态加密需要在密文上做运算,其计算过程往往比较复杂,即便是最基本的同态乘法操作,也需要大量的时间,举例来说,在不做任何优化的前提下,一次明文同态乘法和一次密文同态乘法性能相差10^8倍[7],即使利用中央处理器(central processing unit,CPU)中的单指令多数据流(single-instruction multiple-data stream,SIMD)单元充分把操作并行起来,性能差距也可以达到1万倍左右,这是因为在同态乘法过程中需要进行大量的多项式乘法计算,而多项式乘法又涉及数论变换、向量乘等计算操作,构成了限制同态乘法速度的核心制约因素;在重缩放等过程中,还会进行非常难以并行化的自同态操作,增加了重缩放等操作的计算时间。

同时,由于密文运算所基于的格困难性问题使计算的过程中会产生噪声,从而影响计算的精度,例如,对于环维度在2^{15}次方量级的CKKS算法,在13次密文乘法以后,必须进行自举(auto strapping)操作,以继续全同态计算[8]。除了自举外,用于处理计算过程中产生的噪声手段还有重线性化、重缩放等操作,以帮助实现无限次的同态计算过程,这些操作也消耗了大量的时间。

另外,同态计算过程中产生的密文通常要长于对应的明文,这导致密文无论是在计算还是在存储过程中,其膨胀的数据量带来的传输带宽和存储空间消耗都要大于明文。举例来说,实现一个明文20 bit的同态乘法,其结果密文将膨胀10^5倍[7],远远超过明文的长度,给数据存储和传输带来负担。

2.2　全同态加密算法的研究进展

由于原始的全同态加密算法的性能较差,围绕其进行硬件加速是一条受到广泛关注的技术路线。

目前,围绕全同态加密算法的硬件加速工作进展很快,从2021年学术界开始围绕全同态加密加速展开研究,有相当多的优秀工作问世,不断将全同态加密的加速性能刷新,2022年之前的研究成果如表1所示,值得关注的是,学术界围绕全同态加密研究所面向的应用场景主要是围绕着人工智能领域,即在加密状态下进行机器学习或者深度学习。

表1 2022年前全同态加密研究成果性能

工作实现	芯片架构	年份	实现ResNet-20神经网络同态推理性能相对CPU倍数	实现同态逻辑回归性能相对CPU倍数
Lattigo[9]	CPU	2020	1	1
F1+[10]	ASIC	2021	511	558
BTS[11]		2022	928	1594
CraterLake[12]		2022	5519	2978
ARK[13]		2022	14173	6100
100x[14]	GPGPU	2021	146	456

可以看到,近年来全同态加密加速技术发展迅猛,在硬件加速技术的帮助下,密文下的最好计算性能已经可以达到CPU性能的14173倍,如ARK[13];根据BTS[11]研究工作中同态加密加速方案的性能比明文的性能慢141倍,结合表1中ARK和BTS的性能对比,可以得出ARK研究成果的性能比明文还差1个数量级;同时,可以看到大量的同态加密加速工作围绕ASIC技术路线展开,并取得了较好的性能,作为一种加速芯片,用于计算目的的图形处理单元(general purpose graphics processing unit,GPGPU)并不能满足要求。

3 同态加密算法2023年研究热点回顾

全同态加密算法如果要进入应用阶段,必须解决其性能问题。目前针对同态加密的性能加速主要有2条技术路线:一条是通过硬件对CKKS等经典主流的全同态加密方案实现中的核心基本操作如NTT(Number Theoretic Transform)等硬件实现加速;另外一条是对全同态加密算法进行优化,在不影响其安全应用的前提下,通过调整全同态加密算法中的参数,或使用近似计算方法,或引入新的数学技术、数学工具、数学理论、降噪技术等技术手段进行优化,以提升其计算性能。

目前针对全同态加密算法的研究工作,主要发表在计算机体系结构领域的4个顶级学术会议:国际计算机体系结构会议(International Symposium on Computer Architecture,ISCA)、国际高性能计算架构会议(International Symposium on High Performance Computer Architecture,HPCA)、国际微架构研讨会议(International Symposium on Microarchitecture,MICRO)、国际编程语言和操作系统体系结构支持会议(International Symposium on Architectural Support for Programming Languages and Operating Systems,ASPLOS),以及安全领域的顶级学术会议,如ACM计算机和通信安全会议(The ACM Conference on Computer and Communications Security,CCS)上。本文主要针对这些学术会议中值得关注的热点研究工作进行分析和讨论。

3.1 全同态加密算法的硬件加速工作

3.1.1 基于ASIC的全同态加密算法加速

SHARP是Kim等[15]在ISCA 2023会议上提出的一个基于ASIC的全同态加密加速器,面向机器学习领域,以解决ASIC全同态加密加速器芯片面积过大和功耗过高的问题。SHARP的整体架构如图1所示。作者指出了在机器学习领域内,缩短字长可以有效地降低对片上存储和带宽的需求。为此,论文中经过实验分析得出36位较短机器字长的FHE加速器以维持应用鲁棒性和效率的平衡。在此基础上,作者设计了一种分层架构,配合文中给出的数据排布方式,可以使数据尽可能地只在组内和相邻组间进行交换,减小对片上存储带宽的需

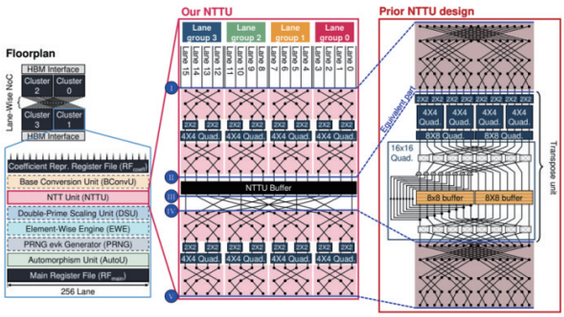

图1　SHARP整体架构

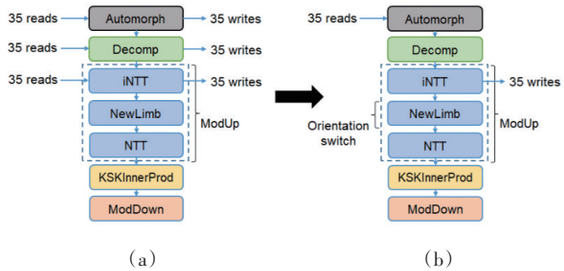

图2　MAD工作中将105次DRAM读取和写入(a)优化为35次读取和写入(b)

求。对比表1中基于ASIC的加速器ARK[13],SHARP在ResNet-20神经网络推理上相对CPU取得了17895倍的加速效果,在逻辑回归上达到17890倍的加速效果,是目前已知最快的全同态加密加速器成果,越来越接近明文的性能。根据表1中的数据,可以得出SHARP比明文性能还慢7倍,考虑有些对安全比较关注,并且愿意牺牲一定性能的应用场景,已经可以达到实用水平。

全同态加密算法中的自举操作对内存带宽需求较大。Agrawal等[16]在MICRO 2023会议中提出了一种适用于全同态加密算法的内存感知设计技术MAD来加速自举操作。该工作提出了缓存优化策略,通过重新排序操作以实现最大数据重用,从而降低自举操作的内存带宽要求。如图2所示,以单次同态旋转(HROTATE)操作为例,MAD工作将105次DRAM读写优化为35次DRAM读写。同时,该工作还提出了几种算法优化,通过减少密文乘法、旋转等各种操作中模降(ModDown)子操作的数量,减少了数据访问模式切换和昂贵的快速数论变换NTT操作的数量,从而减少自举操作中的内存访问次数。

根据使用SimFHE模拟器的基准测试结果,F1+使用MAD-64在逻辑回归训练上可以获得最高27倍的性能提升,BTS使用MAD-512在ResNet-20推理上可以获得最高57倍的性能提升,同时BTS和ARK使用MAD,能将片上缓存要求降低1/16。

3.1.2　基于GPU的全同态加密算法加速

Fan等[17]在HPCA 2023会议上提出的TensorFHE是一个使用GPU的全同态加密加速方案。TensorFHF的NTT执行流程如图3所示。其主要贡献是充分利用了Nvidia GPU中的Tensor Core Units(TCUs)加速NTT的计算过程,与以往的工作不同,该研究放弃了使用蝶形变换来计算NTT,直接使用矩阵乘法来计算NTT操作,通过进一步将矩阵乘法分块来使用GPU中的TCU单元实现加速。相较于表1中效率最高的CPU实现Lattigo[9],TensorFHE在ResNet-20和LSTM两种神经网络推理上分别取得了278倍和223倍的加速效果,在同态逻辑回归任务上取得了1605倍的加速效果,但参考表1数据,其性能依然不如ASIC技术路线好。

GME是由Shivdikar等[18]在MICRO 2023会议上提出的另外一个基于GPU加速全同态加密计算

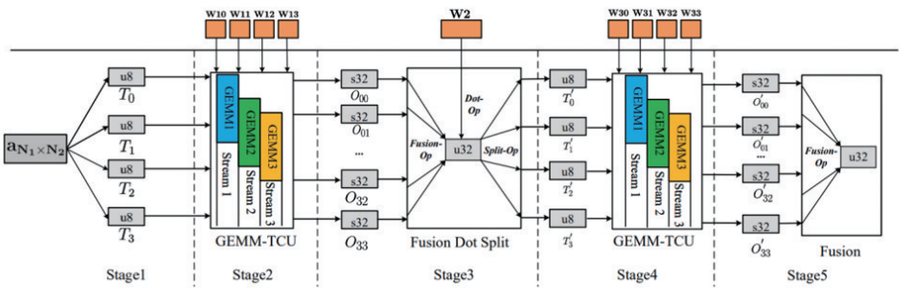

图3　TensorFHF的NTT执行流程

的工作,与其他在GPU上加速FHE的工作不同,其侧重于扩展GPU的微架构使其适用于FHE操作。GME的整体架构如图4所示。作者基于AMD GPU的CDNA微架构,提出了4种微架构扩展,分别是计算单元侧片上互连网络(cNoC)、基于图的局部感知块调度器(LABS)、计算模约减的模块(MOD)和宽乘法累加单元(WMAC)。其中cNoC和LABS结合使用可以尽可能地减少FHE操作中的内存瓶颈,MOD和WMAC则提高了FHE中算数流水线的吞吐量。相较于表1中效率最高的CPU实现工作Lattigo[9],GME在ResNet-20神经网络推理和同态逻辑回归任务上分别取得了1804倍和831倍的加速效果。

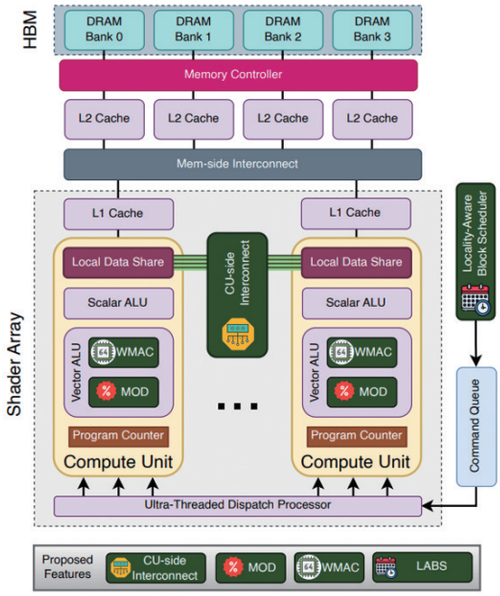

图4 GME的整体架构

3.1.3 基于FPGA同态加密算法加速

Yang等[19]在HPCA 2023会议上提出了一个基于FPGA的全同态加密加速器Poseidon,该工作主要解决加速器的资源占用和带宽问题。Poseidon的整体结构如图5所示。作者将CKKS基本算法拆解为模加(MA)、模乘(MM)、快速数论变换(NTT)、自同构(automorphsim)和共享Barrett约减(SBT)共5个基本算子,通过结合和复用构成计算同态乘法等基本全同态加密操作。在同态逻辑回归(HELR)任务中,Poseidon与表1中基于ASIC的

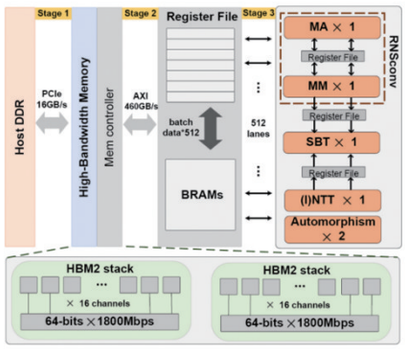

图5 Poseidon的整体结构

CraterLake[12]相比取得了1.64倍的加速比;而在ResNet-20神经网络推理任务中,Poseidon的效率已经与表1中基于ASIC的F1+[10]相近,尽管其效率对比CraterLake仍有11倍左右的差距,但相较于CraterLake需要的256 MB暂存器容量和29 TB/s的带宽要求,Poseidon仅需要8.6 MB大小的暂存器和3.4 TB/s的带宽。

Agrawal等[20]在HPCA 2023会议上提出的FAB是一个基于FPGA的FHE加速器,其工作重点是加速自举过程,并完成了FAB在8卡FPGA集群上的实现。FAB的整体结构如图6所示。作者修改了NTT和密钥切换(KeySwitch)过程中的数据路径,设计了智能操作调度和片上内存管理技术,以此改善自举过程中的内存瓶颈。尽管在同态逻辑回归(HELR)任务上相较表1中基于ASIC的ARK[13]仍有约10倍的差距,其功耗和对于资源的要求也都小于基于ASIC的FHE加速器。

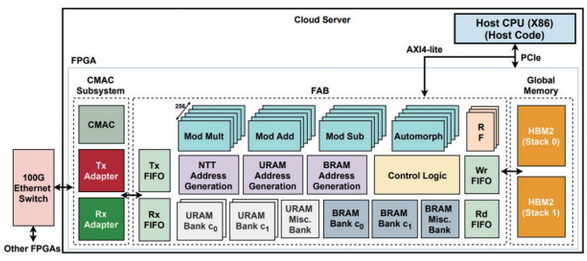

图6 FAB的整体结构

3.2 同态加密算法优化

除了利用硬件,如ASIC、GPU、FPGA等对全同态加密算法进行加速,还有一些工作针对同态加密算法本身进行优化以提升性能。

3.2.1　针对自举操作的优化

Antonio等[21]利用分摊自举操作,减少每条消息所需的同态乘法次数,并减少噪声开销。图7描述了分摊自举操作的主要构建模块。首先,将高噪声的密文打包成一个单一的密文,该密文加密一个多项式,原始消息作为多项式的系数。接下来,在加密状态下通过iNTT(反向数论变换)完成对密文多项式的噪声规约,从而获得表达为X幂的密文,其指数中包含了原始消息及低噪声。最后执行一个消息提取过程,去除噪声项,并对消息应用任何所需的函数集。

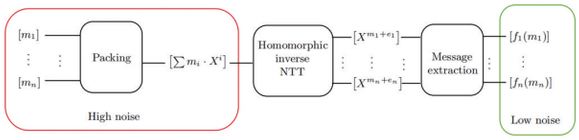

图7　分摊自举操作的构建模块

文献还提出了一种双CRT(Chinese remainder theorem,中国剩余定理),也称为RNS(residue number system,余数系统)版本的GSW方案,包括新的收缩操作,用于降低维度和密文模数来加速同态操作。此外,该方案更加通用,可以处理非二进制消息,并支持可编程的自举。与非摊销版本的TFHE[22]等类似方案的最新启动方法相比,该算法版本快了3.4倍。

Liu等[23]针对分摊自举操作采用并行技术,同时进行n个LWE(learning with errors)密文的引导,将计算复杂度从$O(n^2)$降低到约$O(n)$次多项式乘法,支持任意函数的批量引导,提供更为高效的密文计算。该文章实现了一个C++库,并显示对于二进制门,每个LWE密文的引导时间少于5 ms,这比OpenFHE中现有LWE密文引导的C++实现快了1个数量级。此外,该方法还支持任意函数的批量引导。对于9位消息空间,该方法在评估任意函数时的引导时间约为6.7 ms/LWE密文,这比所有实现类似功能和消息空间的现有方案,如LMP22[24]、GPL23[21]、FDFB[25]、GBA21[26]、LXDX23[27]、PEGASUS[28],都快了2~3个数量级。

Kim等[29]针对管理噪声的分解操作,使用整数环上更高效的分解运算,替代大模数下的离散傅里叶变换(DFT),从而降低了运算成本。这种方法将DFT的计算复杂度从平方降低到线性,而不会增加额外的噪声。此外,该方法还被应用于密钥切换过程,实验结果表明新方法在基础环的维度为2^{15}和2^{16}时比文献[30]中转换为在多精度模数上的NTT形式的方法分别快1.2~2.3倍和2.1~3.3倍。

3.2.2　针对评价密钥的优化

Joon[31]优化服务器生成密钥的策略,采用分层旋转密钥系统,降低CKKS方案和BFV方案中的旋转密钥生成开销。客户端只需生成和传输一小部分旋转密钥到服务器,服务器就可以从公钥和客户端发送的小部分旋转密钥中生成任何所需的旋转密钥,显著降低了客户端和服务器之间的通信成本,以及客户端的运算成本。以使用CKKS方案实现标准ResNet-18网络处理ImageNet数据集为例,原方法[32]需要145.1 s来生成所有旋转密钥,总大小为115.7 GB;使用2级分层旋转密钥系统,旋转密钥集的大小可以减少到2.91 GB(减少到1/39.8),生成时间减少到3.74 s(加快38.8倍);使用3级分层旋转密钥系统,旋转密钥集的大小可以进一步减少到1.54 GB(减少到1/75.1),生成时间减少到1.93 s(加快75.2倍)。常规的旋转密钥系统和分层旋转密钥系统如图8所示。

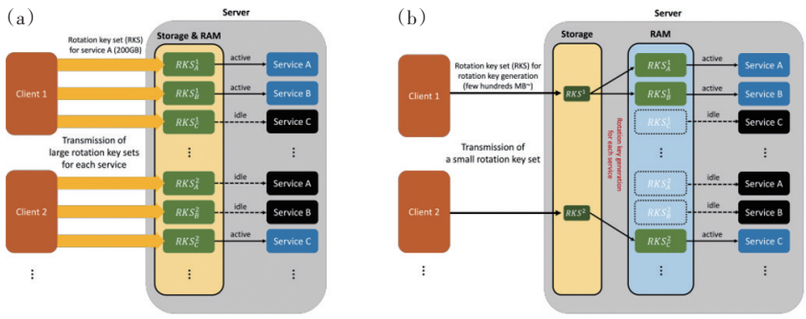

图8　常规的旋转密钥系统(a)和分层旋转密钥系统(b)

Binwu[33]采用盲旋转算法,针对LWE和RLWE(ring learning with errors)采用本方法后可以提高自举操作的运行速度,减少评价密钥的大小。在实现LWE密文的引导算法后,作者将其与FHEW/AP[34]和TFHE/GINX[35]2种引导算法进行了比较。在128位安全参数的三进制密钥分布下,该引导算法仅需存储18.65 MB的评价密钥,为FHEW/AP的1/89.8,为TFHE/GINX的1/2.9。此外,该引导算法的运行时间为112 ms,比FHEW/AP快3.2倍,比TFHE/GINX快2.1倍。

3.2.3 利用新的数学结构进行优化

Okada等[36]主要研究了BGV和BFV这2种广泛使用的全同态加密(FHE)方案。这2种方案具有相同的明文空间,并具有丰富的代数结构。利用Galois自同构性质,提出一种新的评估算法,可以将度数为d的多项式在密态下完成噪声规约计算,只需要3log(d)(在某些情况下仅为2log(d))次密文-密文乘法和自举操作,可以实现BFV中自举操作的加速。常规的自举操作和本方案优化后的自举操作如图9所示。

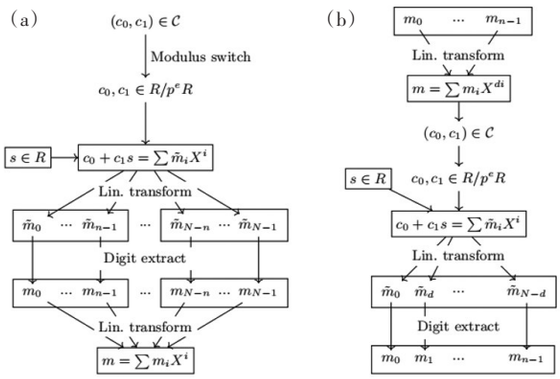

图9 常规的自举操作(a)和本方案优化后的自举操作(b)

实验结果显示,Hao[37]提供的实现相比,该工作获得了1.6倍的速度提升。

Geelen等[38]通过利用零多项式的性质和优化多项式表示,加速了BGV和BFV同态加密方案的引导过程,得到比原始多项式少50%系数的稀疏多项式。此外,文章提出了一种将数字提取分解为一组多项式求值的新方法,相比于HELib算法库有2.6倍的速度提升。

采用系数映射变换多项式构造的ShaftStop算法[39],引入单向映射函数,通过函数来模拟大量随机变量,可以构建足够的破解计算复杂度,满足安全性要求,并且在引入单向函数构造随机变量后,大大降低高次多项式的项数,从而实现快速计算。基于该算法实现的密文数值计算库HENumpy[39],基本算子库使用Go语言编写,采用Python语言封装调用接口,性能对比Numpy明文计算见表2,性能差距普遍在1~2个数量级,在某些强调安全并可以忍受一定性能损失的应用场景具有价值。

表2 HENumpy与明文算子性能对比

序号	函数名	函数说明	进行1万次运算的平均执行时间/ms	
			HENumpy	Numpy
1	add(+)	加法	97.870	5.691
2	mul(*)	乘法	99.590	5.545
3	sin	正弦函数	92.033	5.426
4	pow(**)	幂运算	86.237	5.946
5	sum	累加和	96.045	14.384
6	trapz	定轴积分	95.977	32.875
7	cross	向量叉积	126.352	90.003
8	var	方差	95.131	65.665

注:硬件配置为CPU 13th Gen Intel(R)Core(TM)i5-13400 2.50 GHz内存16 GB。

4 结论

作为一种解决数据安全和隐私泄露问题的理想技术手段,同态加密算法受到了学术界和科研界的广泛重视,但由于其内在原理导致的原因,同态加密算法性能较差,影响了其应用。近年来,在计算机体系结构和密码学研究的顶级学术会议上,围绕同态加密的研究开始加速,这种趋势在2023年显得尤为突出,达到了高峰,在硬件加速、算法优化方面涌现出多个比较重要的研究成果,一方面,以专用集成电路(application specific integrated circuit,ASIC)技术路线为代表的硬件加速效果明显;另一方面,从算法角度进行优化工作进展也很显著,同时中国在围绕同态加密的硬件加速和算法优化方面也涌现出Poseidon、ShaftStop等较好的成果。

可以预测,在几年内,由于中国对于数据流通的重视,对数据作为一种新型生产要素的大力推动,同态加密将随着各行各业对数据安全与隐私保护的强烈需求,被应用到跨行业、行业总分机构数据协作和利用中,借助与人工智能技术的结合,打破数据孤岛,并为基于各种数据利用产生商业价值奠定基础,铺平道路。

参考文献(References)

[1] Freitas L, Tonkikh A, Bendoukha A A, et al. Single secret leader election for PoS blockchains[EB/OL]. (2023-01-30)[2024-01-13]. https://eprint.iacr.org/2023/113.

[2] Cong K, Das D, Nicolas G, et al. Panacea: Non-interactive and stateless oblivious RAM[EB/OL]. (2023-06-12)[2024-01-13]. https://eprint.iacr.org/2023/274.

[3] Charles G, Joseph V, Dalton S, et al. Accelerated encrypted execution of general-purpose applications[EB/OL]. (2023-05-12) [2024-01-13]. https://eprint. iacr. org/2023/641.

[4] Lam K Y, Lu X, Zhang L, et al. Efficient FHE-based privacy-enhanced neural network for AI-as-a-service[EB/OL]. (2023-05-08) [2024-01-13]. https://eprint. iacr. org/2023/647.

[5] Cheon J H, Kang M, Kim T, et al. High-throughput deep convolutional neural networks on fully homomorphic encryption using channel-by-channel packing[EB/OL]. (2023-05-04) [2024-01-13]. https://eprint. iacr. org/2023/647.

[6] Song B. High-throughput deep convolutional neural networks on fully homomorphic encryption using channel-by-channel packing[C]//CCS '23: Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security. Copenhagen, Denmark: ACM, 2023: 2930-2944.

[7] Kim L. Microsoft/SEAL[EB/OL]. (2023-01-11)[2024-01-13]. https://github.com/microsoft/SEAL.

[8] Cheon J H, Cho W, Kim J, et al. Homomorphic multiple precision multiplication for CKKS and reduced modulus consumption[C]//Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security. Copenhagen, Denmark: ACM, 2023: 696-710.

[9] Mouchet C V. Lattigo: A multiparty homomorphic encryption library in go[C]//Proceedings of the 8th Workshop on Encrypted Computing and Applied Homomorphic Cryptography. Online: Homomorphic Encryption. org Consortium, 2020: 64-70.

[10] Samardzic N, Feldmann A, Krastev A, et al. F1: A fast and programmable accelerator for fully homomorphic encryption[C]//MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture. Virtual Event Greece: ACM, 2021: 238-252.

[11] Kim S, Kim J, Kim M J, et al. BTS: An accelerator for bootstrappable fully homomorphic encryption[C]//Proceedings of the 49th Annual International Symposium on Computer Architecture. New York: ACM, 2022: 711-725.

[12] Samardzic N, Feldmann A, Krastev A, et al. CraterLake: A hardware accelerator for efficient unbounded computation on encrypted data[C]//Proceedings of the 49th Annual International Symposium on Computer Architecture. New York: ACM, 2022: 173-187.

[13] Kim J, Lee G, Kim S, et al. ARK: Fully homomorphic encryption accelerator with runtime data generation and inter-operation key reuse[C]//2022 55th IEEE/ACM International Symposium on Microarchitecture (MICRO). Chicago: IEEE, 2022: 1237-1254.

[14] Jung W, Kim S, Ahn J H, et al. Over 100x faster bootstrapping in fully homomorphic encryption through memory-centric optimization with gpus[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021(4): 114-148.

[15] Kim J, Kim S, Choi J, et al. SHARP: A short-word hierarchical accelerator for robust and practical fully homomorphic encryption[C]//Proceedings of the 50th Annual International Symposium on Computer Architecture. Orlando: ACM, 2023: 1-15.

[16] Agrawal R, De Castro L, Juvekar C, et al. MAD: Memory-aware design techniques for accelerating fully homomorphic encryption[C]//56th Annual IEEE/ACM International Symposium on Microarchitecture. Toronto: ACM, 2023: 13.

[17] Fan S Y, Wang Z W, Xu W Z, et al. TensorFHE: Achieving practical computation on encrypted data using GPGPU[C]//2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA). Montreal: IEEE, 2023: 922-934.

[18] Shivdikar K, Bao Y H, Agrawal R, et al. GME: GPU-based microarchitectural extensions to accelerate homomorphic encryption[C]//56th Annual IEEE/ACM Interna-

tional Symposium on Microarchitecture. Toronto: ACM, 2023: 670-684.

[19] Yang Y H, Zhang H Z, Fan S Y, et al. Poseidon: Practical homomorphic encryption accelerator[C]//2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA). Montreal: IEEE, 2023: 870-881.

[20] Agrawal R, De Castro L, Yang G W, et al. FAB: An FPGA-based accelerator for bootstrappable fully homomorphic encryption[C]//2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA). Montreal: IEEE, 2023: 882-895.

[21] Guimarães A, Pereira H V L, Van Leeuwen B. Amortized bootstrapping revisited: Simpler, asymptotically-faster, implemented[C]//Advances in Cryptology-ASIACRYPT 2023: 29th International Conference on the Theory and Application of Cryptology and Information Security, Guangzhou, China, December 4-8, 2023, Proceedings, Part VI. Guangzhou: ACM, 2023: 3-35.

[22] Lee Y, Micciancio D, Kim A, et al. Efficient FHEW bootstrapping with small evaluation keys, and applications to threshold homomorphic encryption[EB/OL]. (2023-06-10)[2024-01-13]. https://eprint.iacr.org/2022/198.

[23] Liu Z Y, Wang Y H. Amortized functional bootstrapping inLess than 7 ms, with $\tilde{O}(1)$ polynomial multiplications [C]//Advances in Cryptology-ASIACRYPT 2023: 29th International Conference on the Theory and Application of Cryptology and Information Security, Guangzhou, China, December 4-8, 2023, Proceedings, Part VI. Guangzhou: ACM, 2023: 101-132.

[24] Liu Z, Micciancio D, Polyakov Y. Large-precision homomorphic sign evaluation using fhew/tfhe bootstrapping. In Advances in Cryptology[C]//ASIACRYPT 2022: 28th International Conference on the Theory and Application of Cryptology and Information Security. Taipei: Springer, 2023: 130-160.

[25] Kluczniak K, Schild L. FDFB: Full domain functional bootstrapping towards practical fully homomorphic encryption[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2023, 2023(1): 501-537.

[26] Guimares A, Borin E, Aranha D F. Revisiting the functional bootstrap in TFHE[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021(2): 229-253.

[27] Liu K, Xu C G, Dou B N, et al. Optimization of functional bootstrap with large lut and packing key switching[EB/OL]. (2023-07-26)[2024-01-13]. https://eprint.iacr.org/2023/631.

[28] Lu W, Huang Z, Hong C, et al. Pegasus: Bridging polynomial and non-polynomial evaluations in homomorphic encryption[C]//42nd IEEE Symposium on Security and Privacy (SP 2021). Online: IEEE, 2021: 1057-1073.

[29] Kim M, Lee D, Seo J, et al. Accelerating HE operations from Key decomposition technique[M]//Advances in Cryptology-CRYPTO 2023. Cham: Springer Nature Switzerland, 2023: 70-92.

[30] Han K, Ki D. Better bootstrapping for approximate homomorphic encryption[M]//Topics in Cryptology-CT-RSA 2020. Cham: Springer International Publishing, 2020: 364-390.

[31] Joon W L. Rotation key reduction for client-server systems of deep neural network on fully homomorphic encryption[C]//The 29th Annual International Conference on the Theory and Applications of Cryptology and Information Security (ASIACRYPT 2023). Guangzhou, China: IACR, 2023: 36-68.

[32] Lee E. Low-complexity deep convolutional neural networks on fully homomorphic encryption using multiplexed parallel convolutions[C]//International Conference on Machine Learning (ICML 2022). Baltimore, USA: Curran Associates, 2022: 12403-12422.

[33] Binwu X. Fast blind rotation for ebootstrapping FHEs [C]//2023 International Cryptology Conference (CRYPTO 2023). Santa Barbara, USA: Springer, 2023: 3-36.

[34] Ducas L. FHEW: Bootstrapping homomorphic encryption in less than a second[C]//Advances in Cryptology-EUROCRYPT 2015-34th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT 2015). Sofia, Bulgaria: Springer, 2015: 617-640.

[35] Chillotti I, Gama N, Georgieva M, et al. TFHE: Fast fully homomorphic encryption over the torus[J]. Journal of Cryptology, 2020, 33(1): 34-91.

[36] Okada H, Player R, Pohmann S. Homomorphic polynomial evaluation using Galois structure and applications to BFV bootstrapping[C]//The 29th Annual International Conference on the Theory and Applications of Cryptology and Information Security (ASIACRYPT 2023). Guangzhou, China: IACR, 2023: 69-100.

[37] Hao C. Homomorphic lower digits removal and improved fhe bootstrapping[C]//Advances in Cryptology-EURO-

CRYPT 2018-37th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT 2018). Tel Aviv, Israel: Springer, 2018: 315-337.

[38] Geelen R, Iliashenko I, Kang J Y, et al. On Polynomial Functions Modulo and faster bootstrapping for homomorphic encryption[C]//Advances in Cryptology-EUROCRYPT 2023-42th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT 2023). Lyon, France: Springer, 2023: 257-286.

[39] 无锡芯光同态信息科技有限公司. 十年磨剑,一朝试锋!首个基于高效密文算法的机器学习产品"御龙"全新发布[EB/OL]. (2023-11-07)[2024-01-14]. https://mp.weixin.qq.com/s/5T3KukovQVFhj69HUKQjyA.

Annual review of advances of full homomorphic encryption technology

FAN Ruiqi[1], CHEN Mingzhi[1], NIU Xinli[2], DONG Wenkuo[1], LI Xiaolin[1], LIU Shuo[1], LIU Jing[1], ZHAO Ming[3], CAI Jiayue[2], YAN Wei[1], ZHU Shuyong[1], ZHENG Kewei[2], XU Peng[4], HAO Qinfen[1*], SUN Ninghui[1]

1. Institute of Computing Technology, Chinese Academy of Science, Beijing 100086, China
2. Wuxi Xingguangtongtai LTD., Wuxi 214104, China
3. Wuxi Institute of Integrate Chip and Interconnect Technology, Wuxi 214104, China
4. School of Cyber Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract In the era of big data and artificial intelligence, homomorphic encryption methods are widely recognized as an ideal technology to solve data security and privacy leakage problems. However, there are currently issues such as poor computational efficiency and ciphertext inflation, which seriously affect application and promotion of this technology. On the basis of summarizing the current research status, this paper reviews and analyzes the relevant research progress in 2023 from two aspects: hardware acceleration for homomorphic encryption algorithms and optimization of homomorphic encryption algorithms. Significant acceleration effects are attributed to the dedicated integrated circuit technology route; substantial progress has been made in optimization from the algorithm perspective. It can be predicted that in the next few years homomorphic encryption will be combined with artificial intelligence to deliver more value in cross-industry and industry division data collaboration and utilization.

Keywords fully homomorphic encryption; data security; privacy leakage protection; computer architecture; cryptography

(责任编辑 王微)