

中国密码学发展报告 2023

中国密码学会 编

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

内 容 简 介

本书是中国密码学会组织编写的重要文献。从 2008 年起每年出版,集中反映上一年国内外密码学最新研究进展。本书分为两部分:第一部分介绍了 2022 年国际五大密码年会(CRYPTO、EUROCRYPT、ASIACRYPT、CHES、FSE)的主要成果,对五大密码年会发表的论文进行了分类整理,分为对称密码、基础理论、公钥密码、(后)量子密码、安全协议、应用密码、密码实现、设备安全/物理攻击/故障注入攻击/掩码攻击等八大类;第二部分介绍了全同态加密,包括全同态加密的 5 篇综述论文,具体为全同态加密应用发展报告、全同态加密的安全模型与安全性分析、全同态加密技术综述、全同态加密构造发展报告、全同态加密算法实现技术综述等。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

中国密码学发展报告. 2023 / 中国密码学会编.
北京:电子工业出版社, 2024. 7. -- ISBN 978-7-121-
-48378-3
I. TN918.1
中国国家版本馆 CIP 数据核字第 2024KY6160 号

责任编辑:张 冉

特约编辑:武瑞敏

印 刷:

装 订:

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编:100036

开 本:787×980 1/16 印张:21 字数:512 千字

版 次:2024 年 7 月第 1 版

印 次:2024 年 7 月第 1 次印刷

定 价:118.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888, 88258888。

质量投诉请发邮件至zltts@phei.com.cn, 盗版侵权举报请发邮件至dbqq@phei.com.cn。

本书咨询联系方式:(010) 88254439, zhangran@phei.com.cn, 微信号:yingxianglibook。

前 言

《中国密码学发展报告》是中国密码学会组织编写的重要学术文献，从 2008 年开始每年出版，集中反映上一年国内外密码学最新研究进展，是密码学研究和密码科技工作者重要的学术参考资料，对推动我国密码理论和技术创新具有重要意义。

《中国密码学发展报告 2023》分为两部分：第一部分介绍 2022 年国际五大密码年会（CRYPTO、EUROCRYPT、ASIACRYPT、CHES、FSE）的主要成果；第二部分介绍全同态加密。

本书的第一部分首次引入 CHES、FSE 两大国际密码年会的主要成果内容，由往年介绍的国际三大密码年会（CRYPTO、EUROCRYPT、ASIACRYPT）扩展至国际五大密码年会（CRYPTO、EUROCRYPT、ASIACRYPT、CHES、FSE）。国际五大密码年会反映了世界先进密码技术和研究的发展方向，归纳整理年会的主要学术成果，有助于我国密码工作者掌握最前沿的密码技术，开展具有中国特色的密码学术研究与技术创新。本书对 2022 年国际五大密码年会发表的论文进行了分类整理，分为对称密码、基础理论、公钥密码、（后）量子密码、安全协议、应用密码、密码实现、设备安全/物理攻击/故障注入攻击/掩码保护等八大类，每个大类又细分为多个研究方向，分别邀请长期从事该方向研究的专家进行归纳、总结和点评，希望有助于读者准确把握每篇论文的研究成果和各研究方向的前沿进展。

随着人工智能、云计算、区块链等技术的广泛应用，数据安全和用户隐私保护问题日益突出。全同态加密可以实现对明密文的任何同态运算，是解决隐私保护和安全计算问题的有效技术，有着广阔的应用前景。本书组织相关领域专家撰写了关于全同态加密的 5 篇综述论文，包括全同态加密的应用发展报告、全同态加密的安全模型与安全性分析、全同态加密技术综述、全同态加密构造发展报告、全同态加密算法实现技术综述，希望有助于读者全面和系统地了解全同态加密领域的研究进展。

高飞教授和郁昱教授分别负责本书第一部分和第二部分的组稿工作，感谢两位老师的精心组织，感谢执笔的各位老师；感谢中国密码学会王瑶副秘书长、毕雪妮同志在协调本书的组稿、编辑和出版过程中付出的诸多心力。

中国密码学会副理事长

中国密码学会学术工作委员会主任委员

戚文峰

2024 年 6 月 16 日

目 录

第一部分 2022 年国际五大密码年会主要成果

对称密码·····	4
基础理论·····	55
公钥密码·····	62
（后）量子密码·····	93
安全协议·····	114
应用密码·····	148
密码实现·····	163
设备安全/物理攻击/故障注入攻击/掩码保护·····	166

第二部分 全同态加密

全同态加密应用发展报告·····	174
全同态加密的安全模型与安全性分析·····	215
全同态加密技术综述·····	235
全同态加密构造发展报告·····	263
全同态加密算法实现技术综述·····	287

第一部分

2022 年国际五大密码年会主要成果

本书第一部分介绍 2022 年在国际五大密码年会(CRYPTO、EUROCRYPT、ASIACRYPT、CHES、FSE)上发表的论文。国际五大密码年会代表了世界先进密码技术和研究方向的发展,通过整理与归纳国际五大密码年会论文的研究方向及其主要贡献,有助于我国密码研究工作者掌握最前沿的密码技术,促进我国密码事业健康快速发展。

2022 年,CRYPTO 会议共收到 453 篇投稿论文,录用 99 篇,录用率约为 21.9%;EUROCRYPT 会议共收到 371 篇投稿论文,录用 85 篇,录用率约为 22.9%;ASIACRYPT 会议共收到 363 篇投稿论文,录用 98 篇,录用率约为 27.0%;CHES 会议共收到 284 篇投稿论文,录用 83 篇,录用率约为 29.2%;FSE 会议共收到 188 篇投稿论文,录用 57 篇,录用率约为 30.3%。

在 CRYPTO 录用的文章中,中国高校和科研院所为第一单位的共 5 篇,作者分别来自香港大学(杨如鹏、于作霞、区文浩)、中国科学院信息工程研究所(曹书蛟、薛锐、牛钟锋、史丹萍)、中国科学院大学密码学院(孙思维)、国防科技大学(刘韵雯、李超)、清华大学(包珍珍、董晓阳),研究方向主要集中在基础理论、公钥密码、量子密码、对称密码和密码攻击。此外,国内学者参与完成的论文有 11 篇,海外华人华侨参与完成的有 11 篇。

在 EUROCRYPT 录用的文章中,中国高校和科研院所为第一单位的共 7 篇,分别来自清华大学(董晓阳、秦岭月、王小云、陈一镭)、中国科学院大学密码学院(孙思维)、山东大学(孙玲、王薇、王美琴)、中国人民解放军战略支援部队信息工程大学(史臻、金晨辉、张际焱、崔霆、丁林、金羽)、台湾“中研院”(钟楷闵)、上海交通大学(姚立、郁昱)、电子科技大学(王煜宇),研究方向主要集中在对称密码、后量子密码和安全协议。此外,国内学者参与完成的论文有 5 篇,海外华人华侨参与完成的有 16 篇。

在 ASIACRYPT 录用的文章中,中国高校和科研院所为第一单位的共 21 篇,作者分别来自山东大学(崔佳敏、魏普文、王美琴、何家辉)、上海交通大学(沈耀斌、刘晗林、郁昱、吕林、刘胜利、谷大武、韩帅、贾琰雪、孙士锋)、清华大学(秦岭月、董晓阳、王安宇、华佳良、王小云、包珍珍)、中国科学院信息工程研究所(刘美成、马丽、张娜娜、杨倩倩、史丹萍、赵佳豪、胡磊、许军、黄震宇、邓燚、李帅帅)、中国科学院大学密码学院(孙思维)、中国科学院大学网络空间安全学院(张心轩)、暨南大学(宋凌、翁健、温金明、赖俊祚、颜俊)、华东师范大学(陈洁、李宇)、鹏城实验室(黄正安、曾功贤、王宇)、密码科学技术全国重点实验室(杨糠)、电子科技大学(王煜宇)、浙江大学(张聪、周哲磊、张秉晟、任奎)、北京航空航天大学(郑志明),研究方向主要集中在对称密码、基础理论、公钥密码和安全协议。此外,国内学者参与完成的论文有 4 篇,海外华人华侨参与完成的有 13 篇。

在 CHES 录用的文章中,中国高校和科研院所为第一单位的共 10 篇,作者分别来自上海交通大学(郁昱)、清华大学(赵灿坤、张能、王汉宁、朱文平、李政东、尹首一、魏少军、刘雷波、杨博翰、陈相任)、山东大学(王伟嘉、郭淳、季凡杰、苏杨)、武汉大学(唐明)、南京农业大学(李延斌、黄玉鑫、朱嘉杰)、蚂蚁集团(Wei Fu)、浙江大学(吴茜眉、张帆、冯天祥、黎治圻、任奎、赵新杰)、无锡沐创集成电路设计有限公司(朱敏)、南京航空航天大学(张吉鹏、刘哲)、广东省数据科学与技术交叉应用重点实验室(黄军浩、赵浩嵩、陈东龙)、台湾实验研究院(Ming-Hsien Tsai)、深圳大学(施晓牧、刘嘉祥)、香港大学(张泽松)、台湾大学(Po-Jen Chen)和台湾“中研院”(黄柏文、王柏尧、杨柏因、周彤、Jin-Han Liou),研究方向主要集中在后量子密码和应用密码。此外,国内学者参与完成的论文有 4 篇,海外华人华侨参与完成的有 9 篇。

在 FSE 录用的文章中,中国高校和科研院所为第一单位的共 15 篇,作者分别来自山东大学(郭淳、王伟嘉、孙玲、王薇、王美琴、倪博煜)、微位(深圳)网络科技有限公司(宗瑞)、北京工业大学(李铮)、中国电子信息产业集团有限公司第六研究所(陈怀凤)、惠州大学(罗宜元)、中国电信股份有限公司(王斯)、华东师范大学(刘富康)、清华大学(何乐、林晓恩、于红波、秦岭月、董晓阳、王小云、贾珂婷、尤启迪)、国防科技大学(刘韵雯)、密码科学技术全国重点实验室(巩欣欣)、中国科学院软件研究所(张斌)、中国科学院信息工程研究所(孙瑶)、湖北大学(陈思维、向泽军、曾祥勇、张莎莎)、上海交通大学(郁昱)、西南大学(杨铮)、重庆工程学院(包致婷)、南京航空航天大学(刘哲),研究方向主要集中在对称密码、理论基础、后量子密码和应用密码。此外,国内学者参与完成的论文有 4 篇,海外华人华侨参与完成的有 1 篇。

国际五大密码年会录用的论文基本涵盖了密码技术的主要研究领域。本书第一部分把录用论文的研究领域分为八大类:对称密码、理论基础、公钥密码、(后)量子密码、安全协议、应用密码、密码实现和设备安全/物理攻击/故障注入攻击/掩码保护等,每个大类包含多个具体的研究点。根据每个研究点,相应研究领域的专家对会议录用论文进行归纳、总结和点评,希望有助于我国密码工作者理解和掌握每篇会议论文的知识。

最后,感谢所有参与撰写和点评的专家以及相关工作人员。

高飞(北京邮电大学)、王婧(华中科技大学)

2023 年 7 月

对称密码

1 对称密码构建

对称密码是现代密码体制的重要组成部分，对称密码的设计与安全性证明研究一直以来都是对称密码领域的主流方向之一。在 2022 年，一些密码学者在该领域取得了一些新的研究进展，部分国内学者在其中也做出了重要的贡献。本文将对该领域 2022 年度的最新成果进行简要综述，以飨读者。

1.1 对称密码设计理论的最新进展

在对称密码设计理论方面，本年度的研究进展主要聚焦在反射密码、分组密码工作模式、分叉密码和认证分组密码算法等。

在 2012 年亚密会上，Borghoff 等人提出了一个分组密码算法——Prince 算法。该分组密码算法具有一个特征，命名为反射特征，即使用一个密钥进行加密与使用一个相关密钥进行解密是一样的，具有这种反射特征的分组密码被称为反射密码。在 2022 年美密会上，Beyne 等人首次从可证明安全的视角对反射密码进行了分析^[1]。一方面，Beyne 等人在理想置换模型

的条件下证明了两轮反射密码的安全性并给出了具体的匹配攻击，得出的安全界为 $O\left(\frac{qp^2}{2^{2n} + \frac{q^2}{2^n}}\right)$ ，其中， q 为构造评估的次数， p 为直接访问隐含置换的次数。由于两轮反射密

码已经达到较好的安全下界，该分析成果对基于单个公开置换构造反射密码具有启发意义。另一方面，针对反射密码提出了一个广义的密钥长度扩展方法，该方法提供了一个具有吸引力的 FX 构造方案的替代方案，可用于 Prince 和其他反射密码，新方案通过对已有方案进行最微小改动即可达到更好的安全强度。其中安全性证明是在理想密码模型的条件下的且需要依赖弱化假设，即两轮 Even-Mansour 密码仅有一个轮密钥，得益于提出的一个新的 Sum-Capture 定理，Beyne 等人将单密钥 Even-Mansour 密码的目前已知的最优界提升了两轮。

目前, 众多的对称密码算法在实际使用时采用了分组密码工作模式, 这要求设计的分组密码算法应当很难与随机置换区分开来, 而且如果用随机置换替代分组密码算法, 该工作模式应当具有可证明的安全性。由此可知, 从更加抽象的角度来看, 分组密码成为密码分析与可证明安全之间的边界。在 2022 年亚密会上, Băcuieți 等人提出了一个新的对称密码算法重构方法^[2]。该方法基于 deck 函数, 而非分组密码算法, 其中 deck 函数是一类输入输出规模任意且具有递增性质的带密钥函数。与基于分组密码算法的工作模式相比, 基于 deck 函数的新工作模式更容易分析, 且在实现上表现更优, 这主要得益于目前已提出的 deck 函数的优异的实现性能。作为副产品, Băcuieți 等人为认证加密 (Authenticated Encryption, AE) 定义了一个新的理想模型, 命名为 Jammin Cipher。与在线认证加密 2 (Online Authenticated Encryption 2, OAE2) 安全模型不同, Jammin Cipher 既是一个可操作的理想方案, 也是一个安全参照方案, 并且能为双向通信和多密钥安全等现实场景提供解决方案。

分叉密码 (Forkcipher) 是一个 n 比特输入、 $2n$ 比特输出的带密钥的可调函数, 其中 $2n$ 比特输出由两个置换的输出串联而成, 可被用于设计面向短消息的认证加密方案。一种典型的分叉密码设计方式是迭代-分叉-迭代 (Iterate-Fork-Iterate, IFI) 范式, 但关于这种设计方式的证明安全性研究还比较缺乏。在 FSE 2022 会议上, Kim 等人通过将 IFI 范式应用于可调 Even-Mansour 密码^[3], 基于公开置换提出了一种构造分叉密码的方法, 得到了一个分叉的可调 Even-Mansour (Forked Tweakable Even-Mansour, FTEM) 密码, 并证明了一个 (1,1)-轮 FTEM 密码在理想置换模型下的安全界是 $2^{2n/3}$ 。其中, (1,1)-轮 FTEM 密码由将一个单轮 TEM 应用于一个明文后, 再独立应用该单轮 TEM 两次后得到。

多重分叉密码 (Multi-Fork-Cipher, MFC) 是由 Andreeva 等人于 2019 年亚密会上提出的分叉密码的一个推广, 是一个对单个输入分组能够计算 s 个输出块的可调密码, 其中 s 是一个可任意选取的固定值。当 $s=1$ 时, MFC 即为可调分组密码; 当 $s=2$ 时, MFC 即为分叉密码。在 FSE 2022 会议上, Andreeva 等人定义了多重分叉密码在 ind-prtmfp 概念下的安全性^[4], 即与 s 个可调置换的不可区分性, 提出了广义的 CTR 加密模式——GCTR, GCTR 模式允许对 MFC 进行并行访问从而加密一个消息, 以及在 Peyrin 和 Seurin 于 2016 年美密会上给出的 nivE 安全概念下对 36 种简单而自然的 GCTR 变形进行了分析。结果表明, 大部分变形在随机值不误用的情况下能够提供完整的 n 比特安全性, 小部分变形在随机值误用的情况下仍能提供接近于 n 比特的安全性。最后, 将基于多重分叉密码 ForkSkinny 的 GCTR 模式与传统 CTR 模式以及近期提出的 CTRT 模式进行了比较分析, 结果发现, 所有 GCTR 变形在加密中等长度消息时都可以提升超过 20% 的效率。

1.2 新的对称密码算法

在 2022 年,一些新的对称算法密码被提出,包括 Forró 流密码算法、Atom 流密码算法、MOE 分组密码算法和 Rocca 认证加密算法等。

由于较高的软件实现性能,基于 ARX 的设计已成为现代密码算法设计的主流方式之一,这里 ARX 分别代表模加、循环移位和异或运算。Salsa 是由 Bernstein 于 2005 年设计的一个基于 ARX 的流密码算法,其减轮版本 Salsa 20/12 于 2008 年被遴选为欧洲 eSTREAM 工程最终胜选的 7 个流密码算法之一,目前已实际应用于 DNS 实现、Linux 内核、口令管理等。ChaCha 流密码算法是 Bernstein 于 2008 年设计的一个 Salsa 变体,目的是提升 Salsa 每轮中的混淆效率并提高密码算法抵抗各种攻击的能力,目前已实际应用于新 TLS 1.3、SSH、Noise、S/MIME 4.0、IKE 和 IPsec 等协议中。由于 Salsa 和 ChaCha 流密码算法的广泛应用,其安全强度评估成为一个重要的研究领域,得到了国际密码学者的关注。在 2022 年亚密会上,Coutinho 等人对 Salsa 和 ChaCha 流密码算法进行了安全性分析,得到了新的分析成果^[5]。具体而言,首先,通过将密码算法视作更简单的子轮的复合,提出了一种新的搜索线性逼近的方法,针对 ChaCha 给出了一个改进的差分-线性区分器;其次,提出了一种新的针对 Salsa 的攻击技术,命名为双向线性展开(Bidirectional Linear Expansion, BLE),应用该新技术改进了已有的针对 Salsa 20/7、Salsa 20/8 的差分-线性区分器和针对 Salsa 20/8 的密钥恢复攻击;再次,在考虑所有攻击结果的基础上,提出了一个新的流密码算法——Forró,安全强度评估表明与 Salsa 和 ChaCha 相比,Forró 能够达到更高的安全水平,并在资源受限的设备等众多平台上能达到更好的实现效率;最后,针对 Salsa、ChaCha 和 Forró 流密码算法等,提出了一个新的密码分析工具——CryptDances,该工具可以用于评估差分相关度、搜索线性逼近和计算攻击的复杂度等。

为了抵抗时间存储数据折中(Time-Memory-Data Trade-Off, TMDTO)攻击,内部状态规模应不低于密钥规模的两倍,这是一个广泛被接受的流密码算法的设计准则。在 FSE 2015 会议上,Armknicht 和 Mikhalev 基于 Grain 流密码算法提出了一个新的流密码算法——Sprout,其内部状态规模等于密钥规模,却能够抵抗 TMDTO 攻击。尽管后续的研究表明,Sprout 具有一些安全性缺陷,但开启了一个新的子领域的研究,即小状态流密码算法的设计与分析,一些新的流密码算法相继被提出。在 FSE 2022 会议上,Banik 等人提出了一个新的小状态流密码算法——Atom,该流密码算法的内部状态规模为 159 比特,声称提供 128 比特安全性^[6]。Atom 流密码算法同时使用了两个密钥过滤器,以保证该算法能够抵抗已知的各种攻击;同时,Atom 流密码算法是目前已知的能够提供 128 比特安全性的状态规模最小的流

密码算法。从结构上看, Atom 类似于 Grain 流密码算法, 但通过引入额外的密钥过滤器, 能够使得 Atom 流密码算法对目前已知的各种攻击免疫。

为了降低成本, 目前集成电路 (Integrated Circuit, IC) 的制造正大量委托给离岸代工厂, 这使其暴露于会被硬件特洛伊木马恶意修改的风险中。近期一些研究表明, 利用秘密分享和多方计算可以使非可信芯片具备抵抗特洛伊木马的能力, 但成本开销较大。在 FSE 2022 会议上, Bronchain 等人研究了通过对称密码高效实现上述能力的可能性^[7], 为了实现这一目标, 通过将模乘运算和二元矩阵乘法结合起来, 提出了一个简单的轮结构。基于此, 提出了一个新的抵抗特洛伊木马的分组密码算法——MOE, 其创新之处在于大部分运算为线性运算的情况下可以实现秘密分享的高效运算。

自 2020 年起, 5G 的商业应用已经在一些国家开始落地。事实上, 一些国家及国际组织对 6G 的研发在此之前已经开始了。就数据传输速率而言, 6G 的要求是不低于 100Gbps, 与 5G 相比提升了 10 倍以上, 现有的移动通信标准密码算法在加解密速率上无法满足 6G 的要求。在 FSE 2022 会议上, Sakamoto 等人针对 6G 的速率和安全需求, 提出了一个带关联数据的基于 AES 的认证加密算法——Rocca^[8]。为了达到超高的软件实现速度, Rocca 算法借鉴了认证加密算法 AEGIS 和 Tiaoxin-346 的设计思想, 充分利用了 AES-NI 和 SIMD 指令集的高速优势; 通过在轮函数中引入一个无代价的分组置换, 与 Jean 和 Nikolić 构造的机构相比, Sakamoto 等人得到了一些更加高效且状态规模更小的结构。基于这个新构造的轮函数, 在考虑针对 AEGIS 和 Tiaoxin-346 认证加密算法的已有攻击的基础上, Sakamoto 等人设计了一个提供 256 比特安全性的 AEAD 算法, 该算法的速率能够达到 138Gbps, 是基于 SNOW-V (国际 5G 候选标准加密算法) 的 AEAD 算法速率的 4 倍以上; 同时, Rocca 算法也比 AEGIS-256 和 AES-256-GCM 等 256 比特密钥的 AEAD 算法要快得多。因此, Sakamoto 等人声称, Rocca 算法是首个针对 6G 的专用密码算法, 具有 256 比特的密钥规模且速率能够达 100Gbps 以上。

1.3 对称密码安全强度评估与证明

在对称密码安全强度评估与证明方面, 本年度的研究进展主要聚焦在安全性证明工具、可搜索对称加密方案安全性证明和白盒密码安全性证明等。

随着 Keccak 算法被遴选为 SHA-3 竞赛的最终胜选算法, 基于公开置换的密码方案在国际密码学界获得了越来越多的关注。由于生日界的限制, 单个置换为了保持足够的安全强度通常需要具备较大的规模, 这限制了其在轻量级场景中的适用性, 因此基于两个公开置换的构造方式获得了大量的关注, 一些密码方案相继被提出, 如 2 轮 Even-Mansour 密码、可调分组密码 TEM、伪随机函数 SoEM、pEDM 和哈希函数 FPTP 等。然而, 伴随着每个新的构造方案的提出, 必须有一个专用的安全证明与之对应, 这表明现有的安全证明方法缺乏通用性。

在 2022 年亚密会上,针对基于两个公开置换的构造方案,Chen 提出了一个新的模块化的安全性分析工具,该工具基于经典的镜像理论 (Mirror Theory),在理想置换模型下可实现构造方案的安全性证明^[9]。Chen 在论文中提出了该分析工具的各种不同变体,可适用于各种不同的安全概念。同时,Chen 在论文中构建了一个一般性框架,用于指导如何使用文中提出的新分析技术,并给出了在应用公开置换镜像理论时应该排除的特殊情况。更进一步地,该论文将提出的框架应用于可调 Even-Mansour 密码、pEDM PRF 的两个置换版本和 nEHtM_p MAC 算法的两个置换版本,通过运用论文提出的新分析工具,以一种非常简单的方式证明了这些构造方案的多用户安全性。

通过利用可搜索对称加密 (Searchable Symmetric Encryption, SSE),用户可将一些文档外包存储于非可信的服务器上,并保持通过访问服务器搜索文档的能力;而利用动态可搜索对称加密时,用户还可以通过更新访问修改基础数据库中的内容,如添加或删除一些条目等。在 2022 年美密会上,Minaud 等人首次研究了动态存储高效的可搜索对称加密的相关问题^[10]。具体而言,涉及了 Local SSE 和 page 高效 SSE,创新地给出了这两种方案的一个关联。在该论文中,Minaud 等人引入了一个映射,命名为 Generic Local Transform,该变换可将一个包含某些特征的 page 高效 SSE 方案作为输入,输出一个具有强 locality 特征的 SSE 方案。Minaud 等人获得的主要结果如下:首先,对 page 规模为 p 的 page 高效 SSE 方案,建立了一个动态方案,命名为 LayeredSSE,其存储空间和 page 效率分别为 $O(1)$ 和 $\tilde{O}\left(\log_2 \log_2 \left(\frac{N}{p}\right)\right)$,该方案的技术创新性是提出了一个两选分配过程的新的加权扩展方案;其次,将 Generic Local Transform 和 LayeredSSE 进行结合,构建了一个动态 SSE 方案,在最长列表规模为 $O\left(N^{\frac{1}{\log_2 \log_2 4}}\right)$ 的条件下,其存储效率、locality 和读取效率可分别达到 $O(1)$ 、 $O(1)$ 和 $\tilde{O}(\log_2 \log_2 N)$;最后,通过将 Generic Local Transform 应用于 Tethys 方案,构建了一个无条件静态 SSE 方案,其存储效率、locality 和读取效率可分别达到 $O(1)$ 、 $O(1)$ 和 $O(\log_2^\varepsilon N)$,其中 ε 是一个任意小的常值并满足 $\varepsilon > 0$,这是目前为止最接近 Cash 和 Tessaro 在 2014 年欧密会给出的下界的构造方案。

白盒密码 (White-box Cryptography) 是由 Chow 等人于 2002 年提出的一项即使攻击者可以访问密码算法的实现设备时仍能保护数据不被非法窃取的密码技术。目前,白盒密码已经广泛应用于数字版权管理、移动支付和银行服务等场景。不可压缩性是白盒密码最基本的安全目标之一。近期,一些高效的不可压缩的分组密码算法先后被提出,如 SPACE、SPNbox 和 WhiteBlock, Hosoyamada 等人^[11]在 2022 年亚密会上证明了将不可压缩的 AEAD 工作模式归

约为不可压缩的分组密码算法的可行性,首次发现一些已有的 AEAD 工作模式[如 CCM、GCM(-SIV)和 OCB 等]即使采用不可压缩的分组密码算法,在面对白盒攻击者时仍是不安全的。基于这些发现,Hosoyamada 等人对 AEAD 方案和分组密码算法的基于不可压缩性的安全性定义进行了形式化,使得将设计的工作模式的安全性归约为底层密码算法的安全性成为可能。针对 AEAD,Hosoyamada 等人提出了一个新的安全概念,命名为 whPRI,这是黑盒条件下伪随机注入安全性的一个拓展。whPRI 保证了面对白盒攻击者时的强真实性,即无泄露的存在不可伪造性,这与已有的概念“要么不能保证真实性,要么只能保证普遍不可伪造性”不同。针对底层密码算法,Hosoyamada 等人通过拓展黑盒条件下的伪随机置换(Pseudo-Random Permutation, PRP)引入了一个新的概念——whPRP,发现一个实际 whPRI 安全的 AEAD 工作模式可以通过一个 whPRP 安全的分组密码算法构建而来。作为实例,基于 SPACE 是一个 whPRP 安全的分组密码算法这一猜想,提出了 SPACE 分组密码算法的一个 256 比特变体。

近期,利用非线性环节不覆盖整个状态空间的替代-置换网络(Substitution-Permutation Network, SPN)设计的密码置换和分组密码算法因各种场景中的优异实现性能获得了大量关注。在 FSE 2022 会议上,针对面向字的带固定线性层的部分 SPN(Partial SPN, P-SPN)方案,Grassi 等人^[12]对其线性层的具体细节如何影响结构的安全性这一问题进行了研究,就如何构建和避免基于概率 1 的无限长的截断差分的攻击分别给出了条件,与之前另一独立的工作相比,Grassi 等人给出了更加广泛的分析结果,就描述线性层的矩阵如何设计能抵抗已有攻击给出了严格的充分必要条件;同时,提供了一个工具,可用于检测给定的线性层是否能抵抗已有攻击。更进一步地,Grassi 等人针对线性层给出了一个充分条件,如果该条件满足,可以确保一定没有无限长截断差分存在。除了 P-SPN 方案,上述结论对 HADES 设计模式还具有借鉴意义。

近期,一些高级密码协议如多方安全计算和零知识证明等催生了一种新的对称密码设计方式,该设计面向大规模的有限域,可归为面向算术的密码(Arithmetization Oriented Ciphers)算法。这种设计方式需要尽力降低域上乘法运算的次数,这使得其容易遭受代数攻击的威胁,尤其是高阶差分攻击。因此,如何评估这类密码算法的代数次数就成为一个需要解决的问题。然而,由于通用且精准的评估方法的缺失,因此面向算术的密码算法的代数次数评估成为一个公开挑战。在 2022 年亚密会上,Cui 等人将二元域上的可分性特征(Division Property)拓展到了二元扩域上,提出了一个二元扩域上的代数次数上界评估框架,可用于检测面向算术的密码算法的代数次数,其主要思想是评估一个分组密码算法的多项式表示中是否包含一些特别的单项式^[13]。随后,Cui 等人将提出的新框架应用于 Feistel MiMC、GMiMC 和 MiMC 等面向算术的密码算法。针对 124 轮的 Feistel MiMC,首次提出了高阶差分区分器,大幅提升了已有的针对 80 轮的 Feistel MiMC 的区分器;同时,针对全轮的 Feistel MiMC,提出了一

个零和区分器,所需的数据量为 2^{251} 。针对 50 轮的 GMiMC,给出了高阶差分区分器。针对 MiMC,给出的分析结果与已证明的准确代数次数保持一致,并指出了 MiMC 设计报告中给出的算法轮数是不足以抵抗高阶差分攻击的,这为算法设计提供了一些指导。

AEGIS-128 和 Tiaoxin-346 是两个基于 AES 的认证加密算法,前者被遴选为 CAESAR 竞赛的最终胜选算法之一,而后者是进入 CAESAR 竞赛第三轮的算法之一,两者的共同特征就是将内部状态划分为若干个 128 比特字,并利用 AES 轮函数对这些字进行更新。虽然两个认证加密算法都采用了类似流密码的设计方式,但与一些知名的面向比特的流密码算法(如 Trivium 和 Grain 等)仍有很大区别。在 CAESAR 竞赛期间,针对 AEGIS-128 和 Tiaoxin-346 的初始化阶段,没有公开的分析成果。由于这两个认证加密算法具有相似性,刘富康等人在 FSE 2022 会议上对其初始化阶段的安全性进行了评估^[14]。刘富康等人通过将 128 比特字视为一个单元,写出了内部状态关于 nonce 和密钥的表达式,并深入研究了如何通过增加合适的条件简化这些表达式。作为结果,针对 5 轮 AEGIS-128 和 8 轮 Tiaoxin-346,发现了一些弱密钥簇,每簇包含 2^{96} 个弱密钥,并构造了时间复杂度和数据量皆为 2^{32} 的积分区分器。基于此区分器,恢复 5 轮 AEGIS-128 的弱密钥的时间复杂度为 2^{72} ;然而,针对 8 轮 Tiaoxin-346 的弱密钥恢复攻击需要依赖于算法使用一个弱常数,而算法使用这个弱常数的概率为 2^{-32} 。这些攻击都达到了 AEGIS-128 和 Tiaoxin-346 初始化轮数的一半,对理解 AEGIS-128 和 Tiaoxin-346 的设计具有积极价值。

1.4 量子安全的对称密码设计

近期的量子密码分析研究进展表明,一些在经典条件下被证明是安全的对称密码算法可以在多项式时间内被破译。同时,已有研究成果表明在经典条件下安全的可调分组密码算法能够通过安全的分组密码算法构造而来。然而,Kaplan 等人在 2016 年美密会上的成果表明,即使底层的分组密码算法是量子安全的,上层的可调分组密码算法在量子叠加态攻击时仍能在多项式时间内被破译。这就留下了公开问题,即是否存在一种构造量子安全的可调分组密码算法的模式。在 FSE 2022 会议上,Hosoyamada 等人在基于归约的可证明安全研究范式下解决了这一问题,基于量子安全的分组密码算法,首次给出了量子安全的可调分组密码算法的一种设计方式,并提供了一个可证明安全界^[15]。利用已有的压缩谕言机技术,证明了当底层是一个 n 比特量子安全的分组密码算法时,该设计方式针对能够进行多达 $O(2^{n/6})$ 次明文的量子叠加态访问的量子攻击仍是安全的。

1.5 面向对称密码的软件加密库设计

可编程逻辑控制器 (Programmable Logic Controller, PLC) 是广泛应用于工业自动化的控制设备,具体可用于电网、供水系统、核电站和制造系统等关键基础设施。在 FSE 2022 会议上,利用一个被称为结构化文本 (Structured Text, ST) 的标准 PLC 编程语言,杨铮等人构建了一个实现轻量级对称密码算法的软件加密库——PLCrypto^[16],这是目前已知的首个基于 ST 且在当前商用 PLC 可直接运行的加密库。PLCrypto 中嵌入了 10 个常用算法,包括单项函数、消息认证码、哈希函数、分组密码算法和伪随机生成器等。PLCrypto 可用于保护 PLC 上数据的机密性和完整性,不需要额外硬件和固件修改。此外,杨铮等人对 PLCrypto 中使用一些基本操作,如比特移位、比特循环、代替和置换等,给出了一般的优化实现方法和技术,这些从实践中提炼出来的优化技巧可用于指导其他计算密集型程序在 PLC 上的实现。最后,以一个被称为生存性证明的密码协议为例,演示了 PLCrypto 的实际使用方法。值得一提的是,软件加密库 PLCrypto 的源代码是公开的,系统操作员可以自由地使用该加密库。

本节作者:丁林(中国人民解放军战略支援部队信息工程大学)

参考文献

- [1] BEYNE T, CHEN Y. Provably Secure Reflection Ciphers[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 234-263.
- [2] BĂCUIETI N, DAEMEN J, HOFFERT S, et al. Jammin' on the Deck[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 555-584.
- [3] KIM H, LEE Y, LEE J. Forking Tweakable Even-Mansour Ciphers[J]. IACR Transactions on Symmetric Cryptology, 2020(4): 71-87.
- [4] ANDREEVA E, BHATI A, PRENEEL B, et al. Fork: Counter Mode Variants based on a Generalized Forkcipher[J]. IACR Transactions on Symmetric Cryptology, 2021(3): 1-35.
- [5] COUTINHO M, PASSOS I, VÁSQUEZ J, et al. Latin Dances Reloaded: Improved Cryptanalysis Against Salsa and ChaCha, and the Proposal of Forró[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 256-286.
- [6] BANIK S, CAFORIO A, ISOBE T, et al. Atom: A Stream Cipher with Double Key Filter[J].

IACR Transactions on Symmetric Cryptology, 2021(1): 5-36.

[7] BRONCHAIN O, FAUST S, LALLEMAND V, et al. MOE: Multiplication Operated Encryption with Trojan Resilience[J]. IACR Transactions on Symmetric Cryptology, 2021(1): 78-129.

[8] SAKAMOTO K, LIU F, NAKANO Y, et al. Rocca: An Efficient AES-based Encryption Scheme for Beyond 5G[J]. IACR Transactions on Symmetric Cryptology, 2021(2): 1-30.

[9] CHEN Y. A Modular Approach to the Security Analysis of Two-Permutation Constructions [C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 379-409.

[10] MINAUD B, REICHLE M. Dynamic Local Searchable Symmetric Encryption[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 91-120.

[11] HOSOYAMADA A, ISOBE T, TODO Y, et al. A Modular Approach to the Incompressibility of Block-Cipher-Based AEADs[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 585-619.

[12] GRASSI L, RECHBERGER C, SCHOFNEGGER M. Proving Resistance Against Infinitely Long Subspace Trails: How to Choose the Linear Layer[J]. IACR Transactions on Symmetric Cryptology, 2021(2): 314-352.

[13] CUI J, HU K, WANG M, et al. On the Field-Based Division Property: Applications to MiMC, Feistel MiMC and GMiMC[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 241-270.

[14] LIU F, ISOBE T, MEIER W, et al. Weak Keys in Reduced AEGIS and Tiaoxin[J]. IACR Transactions on Symmetric Cryptology, 2021(2): 104-139.

[15] HOSOYAMADA A, IWATA T. Provably Quantum-Secure Tweakable Block Ciphers[J]. IACR Transactions on Symmetric Cryptology, 2021(1): 337-377.

[16] YANG Z, BAO Z, JIN C, et al. PLCrypto: A Symmetric Cryptographic Library for Programmable Logic Controllers[J]. IACR Transactions on Symmetric Cryptology, 2021(3): 170-217.

2 哈希函数/消息认证码/伪随机函数

2.1 哈希函数

性质保留哈希 (Property-Preserving Hash, PPH) 是一类特殊的哈希函数, 提供一个评估函数 Eval, 对于两个消息 x_1 和 x_2 , 允许用户仅通过对应的摘要值判断原消息的某种属性 $P(x_1, x_2)$ 。基本的 PPH 应满足, 在敌手不知道使用的具体哈希函数时, 发现评估函数 Eval 输出和真实属性 $P(x_1, x_2)$ 不一致的概率可忽略。而安全 PPH 则要求在敌手已知具体哈希函数并可以适应性选择消息的前提下, 发现评估函数 Eval 输出和真实属性 $P(x_1, x_2)$ 之间矛盾的概率可忽略。

汉明距离谓词是一种二值谓词, 输入两个长度相同的比特串 (消息), 输出其汉明距离是否不小于 (或不大于) 某个阈值 t 。

在 2022 年欧密会上, Fleischhacker 等人提出了一种对于汉明重量谓词的安全 PPH, 与此前最好的安全 PPH 构造相比有许多优势^[1]。首先文中算法安全性的证明基于一个研究充分的格密码困难问题——小整数解 (Short Integer Solution, SIS) 问题, 而此前的构造基于较新提出的计算困难假设。从计算成本来看, 文中算法的 Hash 与求谓词过程只涉及少量模加等简单操作, 而之前的构造包括许多求幂以及大数域上高次多项式的插值及求值计算。此外, 文中算法需要的算法描述成本更少。文献[1]还给出了对汉明距离二值谓词的所有 PPH 算法的输出摘要长度下界, 并说明其构造的算法输出长度已接近最优。

该论文工作的核心是构造了一种安全的集合编码方案。首先任意长度为 n 的比特串 (消息) x 均等价于一个含 n 个正整数的集合 X , 每个元素取值为 $1 \sim 2n$ 。两个消息 x_0 与 x_1 汉明距离小于 t 等价于对应的集合 X_0 与 X_1 的对称差集元素数量小于 $2t$, 即 $|X_0 \Delta X_1| < 2t$ 。文中构造的编码方案是对可逆 BLOOM 查找表 (Invertible Bloom Lookup Table, IBLT) 的修改, 利用若干个 $2t$ -独立的哈希函数, 可以将集合 X 编码为规模为 $\tilde{O}(t)$ 的二维矩阵 H 。编码支持进行集合差运算, $X_0 \Delta X_1$ 对应编码为 $H_0 - H_1$ 。对于对称差的编码, 若 $|X_0 \Delta X_1| < 2t$, $H_0 - H_1$ 可以大概率地解码为正确的 $X_0 \Delta X_1$ 集合, 进而判断 $X_0 \Delta X_1$ 集合是否小于 $2t$ 。论文将该编码方案的安全性 (该方案任意适应性) 选择集合 X 的概率多项式时间敌手, 发现集合对称差解码结果不一致的概率可忽略, 归约到 SIS 问题, 从而证明了编码方案的安全性。

基于该安全的集合编码方案, 可以自然地构造出安全的 PPH 方案。该方案的输出长度为 $O(tl \cdot \log_2 n)$, 其中 $l=l(\lambda)$ 为安全参数的多项式。该方案可以处理的汉明距离谓词满足 $t = O(n / (l \cdot \log_2 n))$ 。

在 2022 年美密会上, Holmgren 等人构造了更好的安全 PPH, 并证明了其参数接近最优界^[2]。与之前的工作相比, 其构造依赖于更一般的假设。

文献[2]首先基于线性纠错码的故障列表解码构造了一种对汉明距离谓词的非安全的 PPH。其输出长度可以达到 $O(t \log_2 n + \lambda)$, 其中 t 为谓词中的汉明距离阈值, n 为输入消息长度, λ 为安全参数。该构造对于 $t = O(n)$ 的谓词是有效的。同时, 论文给出了一个任意 PPH 输出长度的下界, 并证明了其构造的非安全 PPH 输出长度基本最优。

接着, 文献[2]基于同态的抗碰撞的哈希函数对汉明距离谓词构造了一种安全的 PPH 方案, 输出长度最多是之前非安全 PPH 的 $\log_2 3$ 倍, 上界为 $O(t \log_2 n + l)$, 其中 l 为内部使用的同态抗碰撞哈希函数的输出长度。这个输出长度上界比之前的工作 (输出长度 $O(tl)$) 渐近更优。同时, 该方案对 $t = O(n)$ 有效, 而之前的方案只能处理 $t = O(n/l)$ 。

另外, 文献[2]首次基于最基础的标准抗碰撞哈希函数对汉明距离谓词构造出安全的 PPH 方案, 实现分为两步。首先论文基于标准抗碰撞哈希函数为一类更弱的谓词, 即阶越汉明距离谓词, 构造了安全的 PPH; 然后使用故障解码, 论文给出了一种将基于抗碰撞哈希函数对阶越汉明距离谓词的安全 PPH 转换为对汉明距离谓词的安全 PPH 的通用方案。最终得到的方案输出长度为 $O(tl \cdot \log_2(m/t))$, 与之前基于同态抗碰撞哈希函数的安全 PPH 相比稍差, 但达到了与此前工作相当的结果。

最后文献[2]还构造了对汉明距离谓词的随机化的安全 PPH, 即一个消息的 Hash 值是随机化的, 而不是确定的。该方案的输出长度为 $O(\sqrt{\lambda n} \log_2 n + t \log_2 n)$ 。

MD (Merkle-Damgård) 迭代结构是一种经典的哈希函数设计结构, 许多标准哈希函数如 MD5、SHA-1、SHA-256 等均是基于 MD 迭代结构进行设计的。为了对结构本身进行理论情况下的分析, 可以假设存在一个理想的哈希函数, 即其起始值 IV 是均匀随机产生的, 而且其中的压缩函数是一个随机谕言机 (Random Oracle) 的情况。对于这样的理想 MD 迭代结构哈希函数的碰撞攻击, 有一个 STB 假设: 若理想哈希函数的值域大小为 N , 攻击者最多有大小为 S 的预处理空间, 最多可以进行 T 次询问, 每个询问的输入长度不超过 B 个消息块, 则攻击者得到一组碰撞的概率不超过 $\tilde{O}(STB/N + T^2/N)$ 。此前的一些研究证明了 $B=1$ 和 $B=2$ 的情况。2022 年的美密会论文^[3]对这种情况下的理想 MD 迭代结构哈希函数进行了细致的分析, 通过假设存在敌手并进行归约, 对敌手的询问构建出函数转移图, 每个碰撞均对应图上的两条起点和终点均相同的路径, 通过这样的方式对起始值 IV 和哈希函数进行压缩, 而这两个都是完全随机的, 因此构造出了矛盾。通过这样的思路, 证明了在 B 为常数, 或者 $S^4 B^2 \in \tilde{O}(T)$ 的情况下, STB 假设是成立的, 这也说明了在这些情况下目前的攻击已经达到了理论最优的

上界。除此之外,2022 年美密会的另一篇论文^[4]采用将这个问题归约到连续的多个无预处理的 Hash 碰撞问题,给出了另一种情况下成功概率的界,为 $\tilde{O}(\text{STB}/N \cdot \max\{1, \text{ST}^2/N\} + T^2/N)$ 。也就是说,只要 $\text{ST}^2 \leq N$,对任意 $2 < B < T$,STB 假设也是成立的。

除了 MD 迭代结构,还有另一种可以用来设计哈希函数的 Sponge 结构,标准哈希函数 SHA-3 就是基于 Sponge 结构进行设计的。其结构中包含了长度为 c 的内部隐藏状态,避免了 MD 迭代结构存在的消息扩展攻击,一般来说,被认为是更优秀的结构。然而,2022 年的一篇美密会论文^[5]中对输入块数 B 具有限制的理想 Sponge 结构哈希函数进行了分析,给出了其在 $B=1$ 和 $B>1$ 两种情况下碰撞攻击。除此之外,其还证明了 $B=1$ 以及 $B=2$ 时的碰撞成功率上界。特别地,当 $B=1$,且参数满足 $\text{ST}^2 > 2^c$ 时,对 Sponge 结构的攻击成功率会高于 MD 迭代结构,这也是第一个在自然情况下,可证明的 Sponge 结构安全性弱于 MD 迭代结构的例子。

文献[6]中作者 Rothblum 和 Vasudevan 给出了关于哈希函数的(多)碰撞抗性的一些结论。一个 (t, l) -MCRH 指的是满足以下性质的哈希函数。

(1) 将输入的 n 比特压缩成了 $n-1$ 比特的摘要。

(2) 在多项式时间内无法以高于 $1/p(n)$ 的概率找到 t 个输入,它们的输出摘要都相同。

理想的哈希函数对应了 $t=2$ 的情况,而 t 更大对应了哈希函数的弱化情况。这篇文章证明了如果 t 较大的 MCRH 存在,那么也存在 t 更小的 MCRH,只不过一些相关的设置需要做相应的调整,即非均匀指运行 Gen 时改用概率多项式电路生成,io 指不再是对充分大的 n 都成立而是对无穷多个 n 成立。作者引入了部分域 MCRH 并给出了一个证明框架,最终证明了 3 个结论。

(1) 对于 $t=3$, $l = n/2 + \omega(\log_2 n)$ 的 MCRH 可以推出非均匀的 $(2,1)$ -ioMCRH。

(2) 对于 $t=4$, $l = 5n/6 + \omega(\log_2 n)$ 的 MCRH 也可以推出非均匀的 $(2,1)$ -ioMCRH。

(3) 对于任意的 k ,假如存在一个 $(t, l(n))$ -MCRH,则存在一个非均匀的 $(\max((2t \cdot \sqrt{k-1})^{2/3}, 24), \min(l(n) - n/k, l(kn) - n(k-1) - O(\log_2 n)))$ -ioMCRH。

其中,前两个结论能直接推到最理想的 $t=2$ 的情况,但是当已知条件的 $t \geq 5$ 时,由于得到的 $l(n)$ 是负的,因此结论没有意义;第三个结论对应了一般的情况,但只在已知条件 $t > 24$ 时才有意义。

文献[7]探讨了关于哈希函数结构的不可区分性的一系列问题。作者 Günsing 首先指出了一个之前文献中证明上的疏忽,该疏忽影响了之后的多篇文献, Günsing 进一步给出了该疏忽应如何避免和可能的解决方案。此外, Günsing 在修正了该疏忽后重新证明了基于分组密码的树 Hash 在原模式下的安全界,以及证明了在其他原来没有考虑的模式下的安全界,然后将这些结果应用在一些现有结构即基于截断 SHA-2、BLAKE3 和树 Sponge 的树 Hash 中,

并得到了更好的结果。其中,之前文献中证明上的疏忽在于,认为模拟器访问随机谕言机得到的结果是随机的。然而,因为区分者也能访问并可以提前访问随机谕言机,区分者询问模拟器的问题和模拟器访问随机谕言机得到的结果可能是相关的。从模拟器的角度来看,它访问随机谕言机得到的结果和之前某个问题的输入是一样的,这显然不是随机的。一种解决方案是,要求区分者必须一次性询问完模拟器,再询问随机谕言机,这样就不能根据随机谕言机的结果来构造问题询问模拟器了。此外,还可以对 Hash 引入截断等操作来避免这个问题。

2.2 消息认证码

消息认证码 (MAC) 是一种用于确保消息真实性的对称密码原语。MAC 通常由分组密码 (如 CBC-MAC、OMAC 和 PMAC) 或哈希函数构建。然而,当 MAC 使用分组长度为 64 位或更短的轻量级分组密码实现时, $2^{n/2}$ 的生日攻击的安全界就会为 2^{32} 或更小,因此在实际应用中是不安全的。为了克服生日攻击限制的障碍,人们提出了一系列超生日界安全的 MAC 算法,这些算法的结构都遵循一个类似的结构,称为 Double-block Hash-then Sum (DbHtS),包括 3kf9、Poly-MAC 等。

在 2022 年亚密会上,Shen 和 Sibleyras^[8]提出了两种基于 3kf9 的超生日界安全的消息认证码 n2kf9 和 n1kf9,它们是 3kf9 的缩简密钥版本。首先,作者通过回顾 3kf9 以及过去提出过的 3kf9 的一些不安全的缩减密钥版本,包括 2kf9 和 1kf9,分析了它们的基本结构: $T = E_{K_1}(\Sigma) \oplus E_{K_2}(A)$ 以及针对其中 fix 函数的攻击。之后作者在此基础上进行改进,将 2kf9 和 1kf9 中的 A 值乘以 2 (只需简单地将其左移 1 位),构造出了两种新的结构,即 n2kf9 和 n1kf9。将 A 乘以 2 可以打破 Σ 与 A 之间的关系,从而避免了单查询攻击和生日界的区分攻击。与 EMAC 及 CBC-MAC 相比,n2kf9 和 n1kf9 的额外开销非常小,它只需调用一个额外的分组密码并进行一个有限域中的加倍运算。之后,文献[8]对 n2kf9 和 n1kf9 进行了安全性分析,证明了 n2kf9 能抵抗 $2^{2n/3}$ 次询问,而且 n1kf9 在消息空间无前缀时,能抵抗 $2^{2n/3}$ 次询问。

在 2017 年,Cogliati 等人提出了两个结合可调分组密码 (Tweakable Block Ciphers, TBC) 和通用哈希函数的 MAC 算法: NaT、HaT。其中,NaT 的缺陷是:当 nonces 被误用时,它的安全性会降低到生日攻击的界 ($n/2$ 比特, n 代表 TBC 的分组长度); HaT 的缺陷是:即使 nonces 被正确使用,也不具有 n 比特安全性。

为了在对基础构造做出最小改变的前提下克服上述安全缺陷, Choi 等人^[9]提出了两个更加安全、高效的 MAC 算法: NaT2、eHaT。从结构上看: NaT2 是两个 NaT 的结合; eHaT 在 HaT 的结构上进行了两个扩展。

首先, 在消息前面添加一个 nonce 再计算 HaT 的输出; 其次, 把另一个 nonce 经过 TBC 加密后的结果和 HaT 的输出异或。当 nonce 可重复时, NaT2 和 eHaT 都可以提供超过 $n/2$ 比特的安全性。当 nonce 不重复时, 二者都可以提供接近 n 比特的安全性。此外, Choi 等人在分析 NaT2 的安全性时采用了扩展的镜像理论 (Mirror Theory)。

EWCDM (Encrypted Wegman-Carter with Davies-Meyer) 和 DWCDM (Decrypted Wegman-Carter with Davies-Meyer) 是两种使用 nonce 的消息认证码 (MAC)。EWCDM 使用 n 比特 nonce、Hash 密钥 K_h 以及两个加密密钥 K_1 和 K_2 , 进行一次带密钥的 Hash 以及两次加密操作。DWCDM 使用 $2n/3$ 比特 nonce、Hash 密钥 K_h 以及一个加密密钥 K_1 , 进行一次带密钥的 Hash 以及一次加密和一次解密操作。在此前的证明中, 这两种 MAC 模式都被证明在正确使用 nonce 的敌手中能够提供 $2n/3$ 的安全性, 在不正确使用 nonce 的敌手中能够提供 $n/2$ 的安全性。在文献[10]中, 作者使用扩展的镜像理论和二元仿射方程与不等式等技术, 并构造了更新的证明方式, 将两种模式的理论安全性上界都提高到 $O(2^{3n/4})$, 其中 DWCDM 的 nonce 长度也提高到 $3n/4$ 比特。

2.3 伪随机函数

伪随机函数是一类“看起来随机”的带有密钥的函数, 即给定任意的输入, 伪随机函数均可给出“看起来随机”的输出。作为现代密码学中最基本的概念和工具之一, 伪随机函数在过去几十年受到广泛的关注, 直到今天, 密码学界对它的研究仍然保持在一个比较活跃的状态。其中一个自然的研究方向是如何构造更加高效的伪随机函数。在文献[11]中, Banik 等人给出了一个新的伪随机函数的构造, 该构造在函数运算延迟方面取得了进展, 将函数计算延迟的纪录从 2.56ns 降低到 2.4ns。

如果伪随机函数本身是一个置换, 那么称其为伪随机置换。伪随机置换与一般的伪随机函数之间存在高效的转换。从伪随机函数到伪随机置换, 可以使用经典的 Luby-Rackoff 结构。但对于另一个方向, 特别是如果想要避免对伪随机置换的一些特定攻击, 就存在许多值得研究的问题。文献[12-14]针对这些问题进行了研究, 并提出了新的转换方法, 这些转换方法各具特色, 在效率、安全性等方面进行了折中。

可以定义伪随机置换的安全性, 即敌手在查询一些输入, 并观察函数在这些指定输入上的运算后, 无法将函数与真随机置换区分。此外, 也可以定义强伪随机置换的安全性, 即敌

手在看到函数在它指定的输入上的运算和在它指定输出上的求逆后无法区分函数与真随机置换。文献[15]提出了一个新的伪随机置换安全性不规则伪随机置换。它允许敌手同时查询函数运算与函数求逆,但是在敌手进行函数求逆查询时对可以查询的输出进行了一些限制。由于这些限制,不规则伪随机置换的安全性要比强伪随机置换的安全性弱,因此可以通过使用更简单和更高效的方法来构造不规则伪随机置换。另外,在某些构造,如对认证加密方案以及特定安全信道的构造中,不规则伪随机置换可以用来代替强伪随机置换。通过使用不规则伪随机置换这一新的工具作为中间桥梁,文献[15]构造了更加高效的认证加密方案与安全信道。

除了对标准伪随机函数或伪随机置换的研究,还研究了如何在伪随机函数中添加更多的功能,以适用于更广泛的应用场景。其中一个活跃的研究领域是可水印伪随机函数。相比于标准伪随机函数,可水印伪随机函数允许用户在伪随机函数的密钥中嵌入水印,并且保证敌手无法在不破坏伪随机函数运算的前提下去除水印。在可水印伪随机函数中,如果水印的嵌入与提取均不需要另外的私钥,就称其为公钥可水印伪随机函数。在文献[16]中,Yang 等人提出了公钥可水印伪随机函数的概念,并基于多种不同的假设,如不可区分混淆、标准格假设、单向函数的存在性等,给出了具有不同属性的公钥可水印伪随机函数的构造。除此之外,文献[16]还提出了构造可水印伪随机函数的新的框架,即使用带提示的可水印伪随机函数与健壮的不可混淆伪随机函数来构造可水印伪随机函数。新的框架大大简化了可水印伪随机函数的构造,并为新构造的提出带来了便利。

另一个对伪随机函数的扩展是有提示的伪随机函数/伪随机生成器。相比于普通的伪随机函数或伪随机生成器,有提示的伪随机函数或伪随机生成器在敌手可以获得关于密钥或种子提示的情况下仍然能保证函数的伪随机性。在文献[17]中,Alamati 与 Patranabis 提出了对有提示的伪随机函数或伪随机生成器的新构造,并将新构造应用于具有密钥相关安全性的加密方案的构造中。相比于之前的构造,Alamati 与 Patranabis 提出的新构造更加简单且安全性证明更简单。不仅如此,通过提出的新构造方法,文献[17]给出了第一个基于同源问题的有提示的伪随机函数与伪随机生成器的构造。

相关伪随机函数可以使用一对相关的密钥产生相关的输出,即通过使用具有特定关系的一对密钥,在任意的输入上均可以产生一对具有特定关系的输出。通过将这对具有特定关系的密钥分配给不同的参与方,可以使用相关伪随机函数产生大量具有特定关系的伪随机数。例如,如果相关伪随机函数产生的伪随机数满足茫然传输关系,那么持有相关密钥双方的其中一方可以产生 (s_0, s_1) ,而另一方则会产生 (b, s_b) ,这里 s_0 、 s_1 和 b 均为伪随机数。基于合适的相关伪随机函数,可以在协议构造时节省一些操作(如茫然传输操作),进而可以用来构造更高效的安全多方计算协议。文献[18]给出了对相关伪随机函数的新的构造。相比于之前的相关伪随机函数的构造,文献[18]提出的新构造具有更好的效率,并且能支持更多的相关关

系。除此之外，方案的安全性是基于带噪声的奇偶校验（LPN）学习问题，有希望获得抗量子安全性。

除了对伪随机函数和伪随机性的研究，许多密码研究还涉及如何在各种场景下产生真随机数。文献[19]研究了如何在孤立的环境下通过利用硬件的信号抖动产生真随机数。通过使用文献[19]提出的新方法，可以以 298Mbps 的速度产生随机数，并且在供电电压为 0.8V 的条件下，产生 1 比特随机数仅需要耗费 1.46J 的能量。不管是随机比特的产生速度，还是产生每比特所耗费的能量，文献[19]所提出的方法都优于已有的方法。

文献[20]考虑了多方协同产生随机数的场景，即如何在可能存在恶意参与方的情况下产生公共的、无偏的且可以公开验证的随机数。具体来说，文献[20]考虑了在使用区块链的情况下，当限定每个用户仅能公开发布一次消息时，如何产生公共随机数。之前的工作在此设定下仅允许敌手随机地对用户进行腐化，即敌手能否控制某个用户取决于某个随机值。但是文献[20]考虑了可以任意选择用户进行腐化的敌手，大大提高了模型与安全性定义的普适性。在新的模型下，文献[20]构造了安全的协同随机数产生协议，协议在敌手可以腐化约 $1/6$ 用户时是安全的。如果限定敌手从腐化用户处获得的信息，协议在敌手腐化 $1/5$ 用户时是安全的。不仅如此，文献[20]还证明了在他们的模型下，不可能构造敌手在腐化 $1/4$ 用户后仍然安全的协议。

本节作者：于红波、陈怡（清华大学），杨如鹏（澳大利亚伍伦贡大学）

参考文献

- [1] FLEISCHHACKER N, LARSEN K G, SIMKIN M. Property-preserving hash functions for hamming distance from standard assumptions[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 764-781.
- [2] HOLMGREN J, LIU M, TYNER L, et al. Nearly optimal property preserving hashing[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 473-502.
- [3] GHOSHAL A, KOMARGODSKI I. On Time-Space Tradeoffs for Bounded-Length Collisions in Merkle-Damgård Hashing[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 161-191.

[4] AKSHIMA, GUO S, LIU Q. Time-Space Lower Bounds for Finding Collisions in Merkle-Damgård Hash Functions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 192-221.

[5] FREITAG C, GHOSHAL A, KOMARGODSKI I. Time-Space Tradeoffs for Sponge Hashing: Attacks and Limitations for Short Collisions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 131-160.

[6] ROTHBLUM R D, VASUDEVAN P N. Collision-Resistance from Multi-Collision-Resistance[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 503-529.

[7] GUNSING A. Block-Cipher-Based Tree Hashing[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 205-233.

[8] SHEN Y, SIBLEYRAS F. Key-Reduced Variants of 3kf9 with Beyond-Birthday-Bound Security[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 525-554.

[9] CHOI W, INOUE A, LEE B, et al. Highly Secure Nonce-based MACs from the Sum of Tweakable Block Ciphers[J]. IACR Transactions on Symmetric Cryptology, 2020(4): 39-70.

[10] DATTA N, DUTTA A, DUTTA K. Improved security bound of (E/D) WCDM[J]. IACR Transactions on Symmetric Cryptology, 2021(4):138-176.

[11] BANIK S, ISOBE T, LIU F, et al. Orthros: A Low-Latency PRF. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021(1): 37-77.

[12] DUTTA A, NANDI M, TALNIKAR S. Permutation Based EDM: An Inverse Free BBB Secure PRF. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021(2): 31-70.

[13] CHOI W, KIM H, LEE J, et al. Multi-user Security of the Sum of Truncated Random Permutations[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Spring Nature Switzerland, 2023: 682-710.

[14] GRASSI L, MENNINK B. Security of Truncated Permutation Without Initial Value[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Spring Nature Switzerland, 2023: 620-650.

[15] DEGABRIELE J P, KARADZIC V. Overloading the nonce: rugged PRPs, nonce-set AEAD, and order-resilient channels[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 264-295.

[16] YANG R, YU Z, AU M H, et al. Public-Key Watermarking Schemes for Pseudorandom Functions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 637-667.

[17] ALAMATI N, PATRANABIS S. Cryptographic Primitives with Hinting Property. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 33-62.

[18] BOYLE E, COUTEAU G, GILBOA N, et al. Correlated pseudorandomness from expand-accumulate codes[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 603-633.

[19] PEETERMANS A, VERBAUWHEDE I. An energy and area efficient, all digital entropy source compatible with modern standards based on jitter pipelining. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 88-109.

[20] NIELSEN J B, RIBEIRO J, OBREMSKI M. Public Randomness Extraction with Ephemeral Roles and Worst-Case Corruptions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 127-147.

3 对称密码可证明安全

3.1 底层原语的进展

1. 伪随机置换和分组密码

Beyne 和 Chen 等人评估了自反映 (Reflection) 类型分组密码的可证明安全性, 成果发表于 2022 年美密会^[1]。文章有两个贡献点。第一, 作者首次在理想置换 (Ideal Permutation) 模型下研究密钥交替 (Key-Alternating) 自反映型分组密码。作者证明了 2 轮结构下的安全边界

为 $O\left(\frac{qp^2}{2^{2n}} + \frac{q^2}{2^n}\right)$, 其中 q 是结构询问的个数, p 是底层公开置换询问的个数。作者提供了明确

的攻击方法, 论证了该边界的紧致性。第二, 作者针对自反映类型分组密码, 提出了一种通用的密钥扩展方法, 可以有效替换目前 PRINCE 等分组密码使用的 FX 结构。这种方法在原

有构造的基础上,只需要做很小的改动,即可提升安全边界。作者在论证其安全过程中,提出了新的 Sum-Capture 定理,改进了 2 轮单密钥的 Even-Mansour 结构的安全边界,进而在此基础上提供了该方法的安全归约证明。

Degabriele 和 Karadzic 针对可调分组密码 (Tweakable Block Cipher) 提出新的安全概念,介于伪随机置换 (PRP) 和强伪随机置换 (SPRP),称为 Rugged Pseudo-Random Permutation (RPRP),成果发表于 2022 年美密会^[2]。RPRP 假设加密算法的输入并不是一整串的字符,而是被区分开的两部分。这类算法在若干场景中更加适合实际应用。作者主要关注可变长度的可调分组密码,可以更好地实现不同安全性质要求的关联数据的认证加密 (Authenticated Encryption with Associated Data)。作者提出了密文更加紧致的 RUP-安全认证加密码。在此基础上,作者提出了认证加密码的安全概念——Nonce-set AEAD,并提出了具体的构造。Nonce-set AEAD 更加适合安全通信通道 (Secure Channel),如 QUIC、DTLS 等。

郭淳等人分析了 Partial SPN (Substitution-Permutation Networks) 类型分组密码的可证明安全性,成果发表于 FSE 2022^[7]。作者针对 Partial SPN 开展了系统的安全分析,关注了可证明的强伪随机安全性、不可能差分攻击和零相关线性攻击,取得了一系列的成果:5 轮 P-SPN 满足强伪随机安全性;4 轮 P-SPN 可以抵抗不可能差分攻击和零相关线性攻击。

2. 伪随机方程

Choi 等人分析了基于截断随机置换和的伪随机方程多用户安全性,成果发表于 2022 年亚密会。伪随机置换 (PRP) 和伪随机方程 (PRF) 的转换问题是密码学的经典问题。基于 PRP 构造 PRF 主要有两种方法:PRP 取和,即 Summation 构造;PRP 截断,即 Truncation 构造。作者关注了两种方法的混合结构,截断 PRP 取和的部分输出比特,提出了 SaT1 和 SaT2。其中,SaT1 基于单个 PRP,使用了输入域分离的方法;SaT2 使用了两个独立的 PRP。作者证明了 SaT1 的多用户安全边界为 $O(\sqrt{uq}/2^{n-0.5m})$,SaT2 的多用户安全边界为 $O(\sqrt{uq}^{1.5}/2^{2n-0.5m})$,其中 n 是分组密码分组长度, m 是最终输出长度, u 是用户个数, q 是询问个数。此外,作者证明了基于独立 PRP 的 XORP3 结构,证明了其多用户安全边界为 $O(\sqrt{uq}^2/2^{2.5n})$ 。

Banik 等人提出了一个 128 比特伪随机方程 Orthros,成果发表于 FSE 2022^[4]。这个设计主要关注整个实现电路的延迟。因此,作者采用了并行结构,使用了两个加密置换。随机置换的轮方程类似于低延迟分组密码 Mirodi,更进一步减少了延迟,并且通过设定不同的轮数,大幅度提升了安全强度。作者开展了全面的安全分析,并提供了硬件实现。根据实验测试,Orthros 实现了目前最低的延迟。例如,在 STM90nm 工艺实现情况下,Orthros 的延迟大概为 2.4ns,低于 Prince 的 2.56ns、Midori-128 的 4.1ns、QARMA9-128- σ_0 的 4.38ns。

Dutta 等人提出了一个基于单个置换的超越生日界安全的伪随机方程 pEDM,成果发表于 FSE 2022^[5]。这个设计基于 EDM 构造,底层置换使用 $2n$ 比特密钥和 n 比特置换。不同于已有的基于单个置换的伪随机方程,pEDM 只调用底层置换,并没有使用置换的逆运算。作

者证明了 pEDM 的安全边界达到了 $2n/3$ 比特安全, 并提出了具体的攻击, 论证了该边界的紧致性。此外, 作者提出了 pEDM 的 n 比特版本的安全性, 特别指出生日界攻击是否存在是值得进一步研究的课题。

Datta 等人关注了两个消息认证码 EWCDM 和 DWCDM, 提出了新的安全证明, 提升了这两个消息认证码的安全边界, 相关成果发表于 FSE 2022^[6]。EWCDM 和 DWCDM 是基于分组密码的结构。EWCDM (Encrypted Wegman-Carter with Davies-Meyer) 是 Cogliari 和 Seurin 在美密会 2016 上提出的, 在随机数不重复的情况下达到 $2n/3$ 比特安全; 在随机重复的情况下达到 $n/2$ 比特安全。EWCDM 使用了两个独立分组密码。Datta 等人在美密会 2018 上提出了 DWCDM, 将 EWCDM 改进为单个分组密码, 并保留了同样的安全边界。在这篇论文中, Datta 等人证明了 EWCDM 和 DWCDM 在随机数不重复的前提下, 可以达到 $3n/4$ 比特安全。新的安全证明基于改进版的镜像 (Mirror) 证明框架。

3. 可调 (Tweakable) 分组密码

Hosoyamada 和 Iwata 关注了量子安全的可调分组密码, 采用归约可证明安全框架, 基于量子安全的分组密码, 提出了量子安全的可调分组密码构造, 成果发表于 FSE 2022^[8]。作者提出的构造 LRWQ 的结构简单, 是基于 LRW 的可调分组密码结构。如果底层分组密码是 k 比特密钥和 n 比特分组长度, 那么 LRWQ 是 $3k$ 比特密钥、 n 比特调柄和 n 比特分组长度。作者证明了 LRWQ 的量子安全边界为 $O(2^{n/6})$ 。这个结果也论证了量子安全可调伪随机置换可以归约到量子安全伪随机置换。

3.2 杂凑函数的进展

Gunsing 关注了基于分组密码的树形杂凑函数, 进一步讨论之前理论存在的固有缺陷, 并论证了修复方法, 成果发表于 2022 年美密会^[9]。作者关注了此前 Daemen 等人关于树形杂凑函数的安全要求的成果问题, 提出了这个缺陷来自 indifferntiability 证明框架, 指出同样的问题存在于此前多篇论文中。作者讨论了基于分组密码的树形杂凑函数的通用结束方程, 并重点讨论了 5 种方法: Normal Truncation、Truncation without Subtree-Freeness、Chopping、Enveloped、Feed-Forward, 提供了这些方法的安全边界。

Rothblum 和 Vasudevan 关注了杂凑函数的安全性质, 抗碰撞安全性 (Collision-Resistance) 和抗多碰撞安全性 (Multi-Collision Resistance) 之间的关联性^[10]。抗碰撞安全性是杂凑函数的经典性质, 支撑着很多应用的安全性。近年来, 一些杂凑函数的设计关注了抗多碰撞安全性。一般来讲, 大家认为抗多碰撞安全性相对于抗碰撞安全性, 是更弱的安全性质。作者在这篇论文中进行了更加严格的论证, 提出了存在着构造可以将 $3/4$ -抗碰撞安全性转换为抗碰

撞安全性。需要指出的是,作者成功论证了这种构造的存在性,但并没有给出具体的构造。此外,作者也论证了一般性的关联,即 t -抗碰撞安全性和 t' -抗碰撞性的关联,其中 $t > t'$ 。

Lefevre 和 Mennink 关注了 Sponge 结构杂凑函数的抗原像攻击安全性,给出了紧致的安全边界,成果发表于美密会 2022^[11]。Sponge 结构已经成为构造杂凑函数的通用结构,基于公开随机置换,证明了随机谕言机 (Random Oracle) 的生日界不可区分性 (Indifferentiability)。作者发现生日界安全对于抗原像攻击并不紧致,因此开展了更加深入的讨论,提供了改进的抗原像攻击安全边界。这个结果对于很多轻量级密码杂凑函数有实际的应用价值。

3.3 加密工作模式的进展

Cogliati 等人提出了一种基于可调分组密码的加密工作模式 CTET+, 成果发表于 FSE 2022^[12]。作者基于可调 SPN 模式和秘密 S 盒,构造了 2 轮的可调分组密码,并证明了其超越生日界安全。CTET+在使用 n 比特分组密码时,可以加密 wn 比特消息 ($w \geq 2$), 达到 $2n/3$ 比特安全。和以往的工作相比较,CTET+仅使用一个置换,通过弱化线性层,提升了整体的效率。CTET+是第一个基于单个置换超越生日界的加密工作模式。

3.4 消息认证码工作模式的进展

沈耀斌和 Sibleyras 关注了 3kf9 消息认证码的密钥减少变换版本,提出了新的构造,并提供了安全证明,成果发表于亚密会 2022^[13]。3kf9 使用了 3 个密钥, CBC 类型的消息认证码,超越生日界安全。密码学家尝试减少密钥个数,提出了 2kf9 和 1kf9。这两个构造相继被发现了生日界攻击,即不能达到超越生日界安全。沈耀斌和 Sibleyras 深入分析了之前构造的安全缺陷,提出了新的变换版本,证明了 $2n/3$ 比特超越生日界安全。此外,作者还证明了 n2f9 的多用户安全。

Chakraborty 等人关注了 PMAC 的安全边界和询问消息长度的关系,发现了 Naito's PMAC 变换的安全证明存在着问题,并明确了完善安全证明的关键问题,相关成果发表于 FSE 2022^[14]。作者证明了 sPMAC 在弱条件下的与消息长度无关的安全边界,并针对 PMAC1, 通过改变线性层,实现了与消息长度无关的安全边界证明。

3.5 认证加密码工作模式的进展

Hosoyamada 等人关注了基于分组密码的认证加密码的不可压缩性 (Incompressibility), 相关成果发表于亚密会 2022^[15]。不可压缩性是白盒密码 (White-box Cryptography) 的基础安

全性质。近年来,密码学家提出了若干白盒环境下的分组密码,如 SPACE、SPNbox、WhiteBlock。作者论证了认证加密码在白盒环境下的不可压缩安全性可以归约到底层分组密码的不可压缩性。目前主流的认证加密码包括 CCM、GCM (-SIV) 和 OCB 在白盒环境下并不安全,即使底层分组密码满足不可压缩性。因此,作者正式定义了认证加密码和分组密码的不可压缩性,提出了安全性质定义 whPRI 和 whPRP,并证明了 whPRI 安全的认证加密码可以基于 whPRP 安全的分组密码。

Chen 等人关注了认证加密码的密钥安全问题,相关成果发表于亚密会 2022^[16]。认证加密码集成加密和认证等功能,是对称密码的关键算法。在算法设计时,认证加密码使用的密钥是非常确定的。然而,在实际应用环境下,密钥管理中心需要同时管理大量的密钥,必须要确定使用哪个密钥去解密相应的消息。作者首次提出了带有密钥确定方法的基于随机数的认证加密码。在解密过程中,根据密文和密钥,算法可以辨别密钥的正确性,并在密钥正确的前提下可以解开密文。Chen 等人提供了相关的安全性质定义,并提出了若干算法设计和安全证明。

本节作者:王磊(上海交通大学)

参考文献

- [1] BEYNE T, CHEN Y L. Provable Secure Reflection Ciphers, CRYPTO 2022, LNCS 13510, pp. 234-263, 2022.
- [2] DEGABRIELE J P, KARADZIC V. Overloading the Nonce: Rugged PRPs, Nonce-Set AEAD, and Order-Resilient Channels, CRYPTO 2022, LNCS 13510, pp. 264-295, 2022.
- [3] CHOI W, KIM H, LEE J, et al. Multi-User Security of the Sum of Truncated Random Permutations, ASIACRYPT 2022, LNCS 13792, pp. 682-710, 2022.
- [4] BANIK S, ISOBE T, LIU F, et al. Orthros: A Low-Latency PRF, Fast Software Encryption 2022, IACR Transactions on Symmetric Cryptology, Volume 2021, No. 1, pp. 37-77, 2022.
- [5] DUTTA A, NANDI M, TALNIKAR S. Permutation Based EDM: An Inverse Free BBB Secure PRF, Fast Software Encryption 2022, IACR Transactions on Symmetric Cryptology, Volume 2021, No. 2, pp. 31-70, 2022.
- [6] DATTA N, DUTTA A, DUTTA K. Improved Security Bound of (E/D)WCDM, Fast Software Encryption 2022, IACR Transactions on Symmetric Cryptology, Volume 2021, No. 4, pp. 138-176, 2022.

[7] GUO C, STANDAERT F-X, WANG W J, et al. Provable Security of SP Networks with Partial Non-Linear Layers, Fast Software Encryption 2022, IACR Transactions on Symmetric Cryptology, Volume 2021, No. 2, pp. 353-388, 2022.

[8] HOSOYAMADA A, IWATA T. Provably Quantum-Secure Tweakable Block Ciphers, Fast Software Encryption 2022, IACR Transactions on Symmetric Cryptology, Volume 2021, No. 1, pp. 337-377, 2022.

[9] GUNSING A. Block-cipher-based Tree Hashing, CRYPTO 2022, LNCS 13510, pp. 205-233, 2022.

[10] ROTHBLUM R D, VASUDEVAN P N. Collision-Resistance from Multi-Collision Resistance, CRYPTO 2022, LNCS 13509, pp. 503-529, 2022.

[11] LEFEVRE C, MENNINK B. Tight Preimage Resistance of the Sponge Construction, CRYPTO 2022, LNCS 13510, pp. 185-204, 2022.

[12] COGLIATI B, ETHAN J, LALLEMAND V, et al. CTET+: A Beyond-Birthday-Bound Secure Tweakable Enciphering Scheme Using a Single Pseudorandom Permutation. Fast Software Encryption 2022, IACR Transactions on Symmetric Cryptology Volume 2021, No. 4, pp. 1-35, 2022.

[13] SHEN Y B, SIBLEYRAS F. Key-Reduced Variants of 3kf9 with Beyond-Birthday-Bound Security, ASIACRYPT 2022, LNCS 13792, pp. 525-554, 2022.

[14] CHAKRABORTY B, CHATTOPADHYAY S, JHA A, et al. On Length Independent Security Bounds for the PMAC Family, Fast Software Encryption 2022, IACR Transactions on Symmetric Cryptology, Volume 2021, No. 2, pp. 423-445, 2022.

[15] HOSOYAMADA A, ISOBE T, TODO Y, et al. A Modular Approach to the Incompressibility of Block-Cipher-Based AEADs, ASIACRYPT 2022, LNCS 13792, pp. 585-619, 2022.

[16] CHEN J L, GRUBBS P, RISTENPART T. Authenticated Encryption with Key Identification. ASIACRYPT 2022, LNCS 13793, pp. 181-209, 2022.

4 对称密码分析

4.1 经典分析技术

密码分析是评估密码系统安全强度的重要手段。自从 Biham 和 Shamir^[1]提出差分分析之后,陆续又有不少基于差分分析思想的分析方法。在这些分析方法中,有一类方法将密码算

法划分为多个子密码算法,并研究这些子密码算法的性质以及它们之间的相互影响。这类方法一般需要寻找子密码算法的高概率差分迹,并将它们高效复合起来以攻击完整算法。一些著名的分析方法包括飞去来器 (Boomerang) 攻击^[2-3]、矩阵 (Rectangle) 攻击^[4-5]、不可能差分攻击和反弹 (Rebound) 攻击等,这些分析方法广泛应用于分析分组密码。例如,矩阵攻击用于分析 Serpent、Kasumi; 不可能差分攻击用于分析 AES、CLEFIA、Camellia、LBlock、Simon、ARIA、Rijndael-160 和 Rijndael-224; 反弹攻击则用于分析 Whirlpool、Grøstl、Keccak; 飞去来器攻击则用于分析单密钥和相关密钥模型下的 AES。近年来,研究者又提出了回溯 (Retracing) 飞去来器攻击和扩展截断差分攻击并将其用于分析 AES。回溯飞去来器攻击由 Dunkelman 等在欧密会 2020 上提出,该方法试图额外再对子密码算法从空间上进行划分。扩展截断差分攻击从对角线子空间开始附加一轮,可以达到对 5 轮和 6 轮 AES 的区分与密钥恢复攻击。

在欧密会 2022 上,董晓阳等人^[6]对于采用线性密钥生成方案的分组密码提出了一类新型的矩形攻击密钥猜测方法。董晓阳等人发现当针对该类分组密码进行矩形攻击时,构造的四元组需要满足特定的非线性方程组才能在密钥猜测阶段识别候选密钥。但是,一些四元组天然地与该非线性方程组相矛盾,即这些四元组不会提供任何候选密钥,称这类四元组为无效四元组。受先前矩形框架的启发,董晓阳等人发现在构造四元组之前猜测某些关键密钥比特可能会减少无效四元组的数量。然而,一次猜测很多关键密钥比特可能会丧失提前过滤等技术的优势,这可能会导致更高的整体复杂度。为了获得更好的权衡,董晓阳等人在该类线性密钥方案的分组密码上构建了一个新的矩形攻击框架,目的是降低整体复杂度或攻击更多轮数。在新模型中,影响整体复杂度的参数很多,特别是在生成四元组之前关键密钥比特猜测的数量和位置的选择。为了确定最佳参数,董晓阳等人以 SKINNY 分组密码为例构建了一个统一的自动化搜索工具,优化参数包括密钥恢复阶段的最佳矩形区分器、生成四元组之前猜测密钥比特的数量和位置、构建影响的密钥计数器的大小等。基于该自动化工具,董晓阳等人给出了 SKINNY-128-384 的 32 轮相关密钥恢复攻击,将先前的最佳攻击扩展了 2 轮。对于 SKINNY- $n-2n$ 或 SKINNY- $n-3n$ 的其他版本,董晓阳等人也比以前的工作多攻击一轮。此外,使用之前的矩形区分器,董晓阳等人实现了对约减轮数的 ForkSkinny、Deoxys-BC-384 和 GIFT-64 的更好的攻击。最后,董晓阳等人讨论了该新型矩形攻击框架应用于单密钥情况下的密钥恢复攻击,并给出了对 10 轮 Serpent 的新单密钥矩形攻击。

在亚密会 2022 上,秦岭月等人^[7]给出了对可调分组密码 SKINNYe-64-256 以及 SKINNYe-64-256 v2 的相关-密钥调柄 (Tweakey) 矩形攻击、中间相遇攻击以及相关-密钥调柄不可能差分攻击。该工作首先对 SKINNYe-64-256 以及 SKINNYe-64-256 v2 的新的密钥调柄编排算法进行了深入研究,利用该编排算法是线性运算的特点,轮密钥的每个 4 比特元组可由一个二元矩阵 A 乘上主密钥的某个 4 比特元组得到;然后对矩阵 A 的性质进行了分析,尤其是它的

秩对差分类区分器中的抵消以及密钥恢复阶段中的轮密钥猜测策略的影响；最后将发现的密钥调柄编排算法的非平凡的性质与之前的自动化工具相结合，从区分器与密钥恢复阶段方面对攻击进行了改进。

在亚密会 2022 上，宋凌等人^[8]提出了从统一的观点来研究矩形密钥恢复攻击，并提出了一个统一且通用的矩形密钥恢复算法，支持任意的攻击参数，不仅覆盖了此前提出的 4 种算法，还另外包含了此前被忽略的 5 种新的算法。随着新的密钥恢复攻击算法，宋凌等人提出了一种自动化框架来搜索最优的攻击参数，将这些参数输入到密钥恢复算法中就能得到最优的攻击复杂度。该论文将提出的密钥恢复算法应用到分组密码算法 Serpent、SKINNY、CRAFT 和 Deoxys-BC 上并取得了一系列的改进结果，将 CRAFT 的攻击提高了一轮并且得到了当前最好的攻击结果。

近年来，物联网（IoT）以及一些受限应用环境（如 RFID 系统）的安全性日益引起密码研究者的关注。传统解决方案可能并不适用于这些应用场景。因此，近年来，研究者设计了一些新的解决方案和密码基元。这方面的一个重要事件是 NIST 于 2019 年发起的轻量级密码（LWC）征集，其目的是轻量级杂凑函数和带关联数据的认证加密的标准化。数十年以来，研究者已经在轻量级分组密码基础研究方面做出了很多努力，并提出了很多优秀的分组密码算法，如 SKINNY、CRAFT、PRESENT、MIBS、SIMON、GIFT 等。

SKINNY 是一族基于 SPN 网络的轻量级可调分组密码。在 LWC 征集中，有 3 套参选方案以它作为密码基元，如 ForkAE。飞去来器攻击是对减轮的 SKINNY 最为有效的分析方法之一。近年来，密码分析者利用此方法取得了关于飞去来器攻击区分器阶段和密钥恢复阶段的一些研究进展。因此，有必要重新评估 SKINNY 算法对于飞去来器攻击的安全性。

CRAFT 是在 FSE 2019 会议上提出的一个可调分组密码算法。除了设计者的分析，还有一些独立研究者进一步拓展了对该算法的分析。例如，在 FSE 2019 会议上，有研究者给出了该算法更有效的差分、零相关和积分攻击。后来又有研究者给出了截断差分攻击。在 FSE 2020 会议上，有研究者给出了单密钥条件下的弱密钥调柄（Tweakey）截断差分区分器。该区分器利用一条概率为 2^{-54} 的 15 轮相关调柄差分特征，并且可以导出一个 19 轮的密钥恢复攻击。但是，在已公开的研究结果中，还缺少对于该算法的飞去来器攻击。因此，也有必要评估 CRAFT 算法对于飞去来器攻击的安全性。

在 FSE 2022 会议上，Hosein Hadipour 等人^[9]使用考虑了梯转换（Ladder Switch）效应的启发式飞去来器区分器搜索方法，从差分概率和轮数两方面大幅改进 SKINNY- $n-2n$ 和 SKINNY- $n-3n$ 的飞去来器区分器。例如，对于 SKINNY-128-256，Hosein Hadipour 等人给出的 18 轮飞去来器区分器概率为 $2^{-40.77}$ ，远优于之前的最好结果 $2^{-77.83}$ 。同时，Hosein Hadipour 等人给出的区分器最多可达 21 轮，而已有的最优结果是 19 轮。对于 SKINNY-128-384，已有最好的飞去来器区分器可达 24 轮且概率为 $2^{-104.86}$ ，Hosein Hadipour 等人给出了新区分器

在相同轮数情况下的概率为 $2^{-87.39}$ 。相同的搜索技术被应用于寻找 CRAFT 算法的飞去来器区分器。结果表明,相比于其他单调柄模型的区分攻击(如差分攻击),飞去来器攻击是对 CRAFT 算法的一种更有前景的分析方法。例如, Hosein Hadipour 等人给出了一个实际的 9 轮单调柄飞去来器区分器,其概率为 $2^{-14.76}$,而之前最好的 9 轮区分器概率为 $2^{-40.20}$ 。基于这些飞去来器区分器, Hosein Hadipour 等人给出了 SKINNY- $n-2n$ 和 SKINNY- $n-3n$ 的相关调柄密钥矩形攻击,以及 CRAFT 算法的单调柄矩形攻击。例如,对于 SKINNY-64-128, Hosein Hadipour 等人给出了 23 轮的相关密钥调柄矩阵攻击,其时间存储复杂度分别为 $2^{120.7}$ 和 $2^{60.9}$,而之前最好结果的相应复杂度分别为 $2^{125.91}$ 和 2^{124} 。对于 CRAFT, Hosein Hadipour 等人给出了 18 轮的单调柄模型矩形攻击。这是对该算法的首个矩阵攻击,也是在攻击轮数意义下目前最好的分析结果。

在 FSE 2022 会议中, Gregor Leander 等人^[10]给出了 CRAFT 轮函数的等价刻画。在某些密钥空间上,该等价刻画与 Feistel 网络结构是非常相似的。在该弱密钥空间上, Gregor Leander 等人给出了 23 轮的单密钥差分攻击以及 26 轮的相关密钥差分攻击。

目前的研究表明,由于在子密码算法边界上的相互作用,飞去来器攻击中两条短差分是相互依赖的。Murphy 的工作表明,这种依赖性会导致两种结果:一是会导致不相容;二是可以通过 S 盒转换(S-box Switch)、梯转换(Ladder Switch)的方式改进攻击的轮数。这种方法最后被推广为三明治(Sandwich)攻击。这种新的密码分析方法引入了一些新密码学指标,如飞去来器连接表(BCT)、Feistel BCT 和飞去来器分布表(BDT)。一个有趣的分析技术是 Biham 等人提出的用以分析 Skipjack 算法的 Yoyo 游戏(Yoyo Game)。该分析技术结构上与飞去来器攻击相似,但是无须将原密码算法划分为子密码算法。在亚密会 2017 上, Yoyo 游戏被用于构建 2 轮一般 SPN 网络的确定性区分器,并基于此得到了关于 AES 的 5 轮密钥恢复攻击。Yoyo 游戏的技术随后被进一步推广,并用于分析已知密钥设定下的 AES 和未知密钥设定下的 ForkAES。

在 FSE 2022 会议上, Mostafizar Rahman 等人^[11]进一步研究了 Yoyo 技术并改进了分析的轮数。新的方法可以看作将 Yoyo 游戏嵌入飞去来器迹中。其中,上面部分的飞去来器迹与 Yoyo 一致,下面部分是经过特别调整的标准差分迹。在上下迹的边界上,则应用 S 盒转换和梯转换技术。主要的出发点是构造一条上面部分的迹并满足密文上的差分可以在上下迹的边界处产生一个 Yoyo 字交换(Word-Swap),反过来恰好满足 Yoyo 准则。因此,与典型的 Yoyo 技术一样,这会以概率 1 返回 Yoyo 且可以在上层验证。Mostafizar Rahman 等人证明了 S 盒转换和梯转换如何实现字交换。得益于边界字交换技术,飞去来器的复杂度为 pq^2 。换句话说,根据 Yoyo 性质,复杂度降低了一个因子 p 。这些新技术可用于分析 AES 算法。通过将 Yoyo 嵌入飞去来器中, Mostafizar Rahman 等人给出了 AES-128 的 6 轮密钥恢复攻击,时间

复杂度为 2^{78} 。同时,给出了基于 AES 的可调分组密码 Pholkos 的 10 轮密钥恢复攻击。此外, Mostafizar Rahman 等人也研究了与欧密会 2020 上提出的回溯飞去来器攻击的理论关系。

Cube 攻击最早由 Dinur 和 Shamir^[12]在欧密会 2009 上提出,其主要思路是通过选取大量的明文(也可以是 IV 或 Nonce),从密码算法的输出比特中构建关于密钥的方程式,从而提取出部分密钥信息。Cube 攻击主要有两个阶段:在第一阶段,也称为线下阶段,攻击者需要选择合适的 Cube,并恢复出这些 Cube 在输出比特中的超级多项式(Superpoly);在第二阶段,也称为线上阶段,此时真实的密钥是未知的,但攻击者可以通过多次访问加密机,获知在真实的密钥下超级多项式的值,这就建立了一个关于密钥变量的方程组。通过求解这个方程组,攻击者可以获取到一部分密钥比特的信息,而剩下的密钥比特可以通过穷搜的方法恢复出来。自 Cube 攻击提出起,已经出现了许多变体,如动态 Cube 攻击、条件 Cube 攻击、相关 Cube 攻击等。

Cube 攻击的一大难点在于如何恢复出超级多项式的表达式。随着密码轮数的增长,超级多项式的复杂度几乎呈指数级增长。作为积分攻击的一般化,Division Property 最早由 Todo^[13]提出,然后被引入 Cube 攻击中,用来探测超级多项式的结构。特别地,在亚密会 2020 上,单项式预测技术^[14]作为 Division Property 的代数解释被提出,其核心思想是对单项式路径进行计数,根据单项式路径数目的奇偶性来判断一个项是否存在于超级多项式中。单项式预测技术搭配上亚密会 2021 上提出的基于分而治之策略的嵌套框架,可以恢复出 845 轮 Trivium 的超级多项式。但是对于大于 845 轮的 Trivium,依然无能为力。

单项式预测技术的瓶颈在于不得不对所有的单项式路径进行枚举,为了缓解这个问题,在亚密会 2022 上,何家辉等人^[15]提出了核心单项式预测技术,相比于单项式预测,核心单项式预测牺牲了一部分精确性来换取效率上的提升,因此,为了保证最终的准确性,单项式预测技术被应用了两次。具体来说,他们将整个恢复超级多项式的方法分成了两部分,即系数求解器和项扩展器。在系数求解器中,他们首先将核心单项式预测与单项式预测结合,恢复出低轮的有价值的单项式,然后对这些有价值的低轮单项式,再应用单项式预测来恢复它们的超级多项式。而在项扩展器中,基于分而治之策略的嵌套框架被应用。得益于核心单项式预测带来的速度上的提升,他们可以恢复出直到 848 轮 Trivium 的超级多项式,该超级多项式包含项的数目超过 10 亿项。这就提出了另一个难点,即如何求解一个由复杂超级多项式构建出来的方程。尽管该超级多项式包含 10 亿多项,但相比于随机超级多项式,其依然可以被认为是稀疏的。为此,可以利用先猜测后确定的策略和莫比乌斯变换以快于暴力穷搜的复杂度恢复出密钥。这套方法也被应用于 Grain128AEAD、Kreyvium 等其他流密码算法,均得到了在这些密码算法上改进的 Cube 攻击的新结果。他们的这项工作不仅将之前超级多项式恢复的时间大幅缩短了,而且提高了在 Trivium 等算法上立方攻击的轮数。超级多项式的恢复

本质上是一个建模的问题,核心单项式预测的优势在于其提供了一个相比于单项式预测更加简单的 MILP 模型,从而更有可能被 Gurobi 在限定的时间内求解。

近似碰撞攻击作为差分分析的一个变种被广泛应用于对哈希函数、分组密码的安全性分析中。近年来,许多学者将近似碰撞攻击的思想应用于序列密码,如利用密钥流差分与内部状态近似碰撞直接的相关性来恢复内部状态比特。利用近似碰撞攻击方法,张斌等人^[16-17]先后给出了国际标准序列密码算法 Grain-v1 和 A5/1 的内部状态恢复攻击结果。在 FSE 2022 会议上,Patrick Derbez 等人^[18]指出上述两个近似碰撞攻击结果均是错误的,即 Grain-v1 攻击的复杂度高于穷搜复杂度;A5/1 攻击复杂度也远高于预期。张斌等人在 Grain-v1 和 A5/1 近似碰撞中频繁使用列表融合来过滤内部状态,同一列表多次参与融合,因此融合后的列表元素存在很强的相关性。但在进行过滤强度估计时,张斌等人将融合后强相关的列表当作是相互独立的,从而导致估计出来的过滤强度显著高于实际情况。该工作通过理论和实际实验两方面揭示了上述问题,从而彻底推翻了张斌等人关于 Grain-v1 和 A5/1 的近似碰撞攻击结果。该工作的意义在于首次确定了近似碰撞攻击的两个基本原则:①近似碰撞性质所起到的对内部状态的过滤强度完全由参与近似碰撞的比特数决定,参与近似碰撞的比特进行排列组合并重复比对并不能提升过滤强度;②在测算近似碰撞区分器强度时不能忽视比特间的相关性。该工作还进一步强调了验证实验的重要性,指出充分的验证实验能够有效避免类似错误的发生。

SNOW-V^[19]是 SNOW 系列序列密码算法的新成员,具有 256 比特密钥和极高的软硬件实现效率,用于应对 5G 及以上的安全性和效率需求。自 2019 年问世以来,SNOW-V 就受到了密码学术界的广泛关注,科研人员使用各类密码分析技术对 SNOW-V 的安全性进行了一系列研究。典型的分析技术包括猜测确定攻击、线性分析等^[20-21]。

在 FSE 2022 会议上,杨静等人^[22]给出了两项针对全轮 SNOW-V 的猜测确定攻击以及一项针对简化版 SNOW-V 的线性区分攻击结果。该工作的意义在于针对 SNOW-V 算法的特点提出了定制化的猜测确定策略,即首先通过猜测 3 个 128 比特的状态变量得到仅含一个未知状态变量的非线性方程,然后通过特定顺序的字节枚举对非线性方程进行求解,并在求解过程中综合运用预计算、多密钥流字过滤、猜测路径提前终止等技术,从而大幅降低了 SNOW-V 猜测确定攻击的复杂度。此外,杨静等人还进一步研究了将 SNOW-V 的 32 比特模加简化为异或运算时的线性相关性质,通过消除至多 9 个活跃 S 盒构造具有高偏差的线性掩码,从而得到数据复杂度为 2^{303} 的对简化版 SNOW-V 的线性区分攻击。

快速相关攻击是一种针对序列码工作阶段的密钥恢复攻击,它是利用工作阶段线性逼近的一类统计攻击。在欧密会 2022 中,针对 SNOW-V 和 SNOW-Vi,史臻等人^[23]将连续 3 拍工作阶段所构成的复合函数拆解成 6 个子函数,并结合分组密码算法中自动化分析的方法,史臻等人分别对 6 个子函数中的 8 比特 S 盒和 32 比特模加的线性逼近分别建模;并基于

SAT/SMT 求解器,提出了一种自动化搜索线性逼近的模型。针对 SNOW-V 和 SNOW-Vi 的连续 3 拍工作阶段,史臻等人利用上述自动化搜索模型,找到了高相关性的线性逼近,并基于该线性逼近,给出了时间复杂度低于 2^{256} 的快速相关攻击。史臻等人的研究表明,在忽略密钥流长度的限制下,SNOW-V 和 SNOW-Vi 无法保证 256 比特安全。

在 FSE 2022 会议上,巩欣欣和张斌^[24]研究与对比了 SNOW 2.0 和 SNOW 3G 的大单位字线性逼近及逐比特线性逼近的性质,利用新搜索得到的逐比特线性逼近给出了 SNOW 3G 更好的快速相关攻击结果。该工作首先考查了大单位字线性逼近和逐比特线性逼近的关系,以 SNOW 2.0 为例指出大单位字线性逼近相比于逐比特线性逼近更易得到大的线性偏差。然后,通过对比 SNOW 2.0 和 SNOW 3G 的逐字节线性逼近与逐比特线性逼近,发现很多逐字节线性逼近与逐比特线性逼近的线性偏差大致相同,由于逐比特线性逼近可以更有效地进行线性攻击,因此改进了 SNOW 2.0 和 SNOW 3G 的基于大单位字线性逼近攻击。

中间相遇分析方法是从两个独立方向计算同一个中间状态并计算匹配的方法,杂凑函数的原像攻击方法多应用了中间相遇的思想。Aumasson 等人^[25]结合中间相遇方法对约减轮的 MD5 提出了原像攻击。随后,Sasaki 和 Aoki^[26]进一步结合中间相遇思想与 local-collision 成功地给出了 HAVAL 的原像攻击,并提出了 splice-and-cut 和 initial structure 的中间相遇分析技术的概念^[27-28],并且成功攻破了全轮 MD5 的原像攻击的理论安全性。中间相遇原像攻击逐渐成为杂凑函数原像攻击非常有效的方法,攻破了多个杂凑函数申明的理论安全性强度,包括全轮的 MD4^[29]、MD5^[27]。针对 AES 杂凑函数模式,文献[30]一般化了传统的初始结构技术,提出了起始点的思想,同时提出了将列混淆和轮密钥异或运算整合为一个运算的思想,并且基于混合整数线性规划(MILP)模型,提出了中间相遇原像攻击的自动化分析方法。

在美密会 2022 上,包珍珍等人^[31]观察到一些状态可以描述为两个方向的状态的线性组合,从而提出了 superposition 的概念,并且自动化了猜测确定技术。该文为提高搜索效率,提出了一些优化策略,并将模型应用于改进 AES 类杂凑函数的原像攻击和分组密码的密钥恢复攻击。对 ISO 杂凑函数标准 Whirlpool,将原有最长 6 轮原像攻击改进到 7 轮。在单密钥模型下,得到了 ISO 可调分组密码标准 SKINNY 的最长攻击轮数的结果。在美密会 2022 上,André Schrottenloher 和 Marc Stevens^[32]提出了另一种中间相遇原像攻击的基于 MILP 的建模搜索方法。该文不同于文献[30]中定义部件的状态传播原则,而是从整体上看待状态属于哪个方向并优化计算复杂度。该方法可以找到置换的经典和量子攻击结果,并可以应用于 Present-类、Feistel 等结构中,但是该方法的限制是该搜索方法仅适用于置换,不能利用密钥编排中的密钥自由度。该搜索模型能够重现文献[6]中最好的结果,并对 AES 类杂凑函数 Haraka v2 和 Grøstl 提出了新的量子原像攻击结果。

差分-线性攻击是一种组合分析技术,它结合了差分攻击和线性攻击这两种对称密码分析最有效的手段。这一方法从 1994 年由 Langford 和 Hellman 提出,至今已有 30 年的历史,但

它始终没有跳出使用标准差分和线性特征的原始框架。在欧密会 2021 中,刘韵雯等人提出了利用旋转差分特征替换差分-线性攻击中标准差分特征的方法,给出了当输出掩码为单位向量时计算旋转差分-线性区分器相关度的方法,并提出了当输出掩码为任意向量时如何计算旋转差分-线性区分器相关度的公开问题。牛钟锋等人^[33]在美密会 2022 中解决了这一公开问题,并利用所提出的方法对 Alzette、SipHash、ChaCha 和 SPECK 进行了分析,大幅改进了已有的分析结果。同时,利用该方法,牛钟锋等人给出了多个 ARX 型算法实验型区分器的理论解释。

4.2 自动化密码分析

近年来,用来搜索具有高概率且覆盖尽可能多轮数的区分器的自动化建模(如 MILP、SAT/SMT、CP 模型)已经成为密码分析领域中的研究热点。在 FSE 2022 会议上,秦岭月等人^[34]对用来搜索可以直接改进密码分析结果的区分器的自动化模型进行了研究。该工作首先分析了限制一般的差分密钥恢复攻击的具体因素以及由 Zhao 等人提出的扩展矩形攻击模型。在数据复杂度以及时间复杂度均低于穷搜的前提下,最大化分组密码被攻击的轮数,其考虑的限制条件包括区分器的概率、被攻击的算法的输入输出差分不活跃比特的数量以及在拓展轮数中需要被猜测的密钥比特的数量。最后,提出了对 SKINNY 的新的相关密钥矩形攻击的自动化 MILP 模型。对 SKINNY 和 ForkSkinny 的一些减轮版本的密钥恢复攻击比之前最好的攻击多覆盖 1~2 轮。

一些自动化搜索 boomerang 区分器和 boomerang 攻击的 MILP 模型近年来被提出,但是此前的搜索模型只可以被应用于线性密钥编排的分组密码中。在亚密会 2022 上,Patrick Derbez 等人^[35]提出了一种新的搜索模型,该模型可以处理非线性密码编排的密码。虽然新的模型更复杂使得穷搜的时候搜索效率更低,但是只要引入一些提示,该模型能够找到现有最好的 AES-192 的相关密钥 boomerang 攻击的最好的攻击结果,将此前亚密会 2009 的攻击时间/数据/存储复杂度由 $2^{176}/2^{123}/2^{152}$ 变为 $2^{124}/2^{124}/2^{79.8}$ 。

三子集可分立方攻击 MILP 建模理论,首次建立了单项式系数与模型可行解个数奇偶性之间的联系^[45]。基于分治的超级多项式恢复方法,进一步提升了三子集可分立方攻击 MILP 模型的求解效率^[46]。结合成熟的 MILP 求解软件,人们已经能够有效地获得超级多项式的代数表达式。如果想要进一步提升攻击轮数,就需要筛选出优质的立方使得对应的超级多项式具有平衡的真值表、较少的项等优秀密码学性质。

在 FSE 2022 会议上,孙瑶^[47]提出在基于分治的超级多项式计算过程中,引入特定的筛选策略,从而得到具有平衡超级多项式的优质立方。筛选策略基于两项观察经验:在分治子系统中系数为 0 的单项式,有很高的概率最终系数为 0;当分治子系统中存在可行解的子系统

数量较少时,在子系统中系数为 1 的单项式最终系数为 1 的概率较高。该工作的意义在于基于上述筛选策略,给出了 Trivium 和 Kreyvium 序列密码算法当时最优的立方攻击结果,具体包括 808 轮 Trivium 的实际攻击、843 轮 Trivium 的理论攻击以及 893 轮 Kreyvium 的理论攻击结果。

在 FSE 2022 会议上,孙玲等人^[53]提出一种将 Matsui 界条件转化为布尔表达式的编码方法,该方法不需要引入新的辅助变量,只需要添加少量的约束条件便可以将多个界条件刻画到 SAT 模型中。该工作还研究了不同界条件集合下, SAT 模型自动化搜索的加速效果,根据大量的实验测试,提出具有优良加速求解效果的界条件集合的生成策略。该工作的重要意义在于给出了 SAT 模型的加速求解方案,可以更加高效地完成密码算法的差分和线性特征搜索,给出了 GIFT、Rectangle、LBlock、SPECK 等密码算法更好的安全性分析结果。

SHA-3 系列杂凑函数的碰撞攻击是业界关注度高但进展较为缓慢的一个问题。自从郭建等人在 JOC 2020 上提出对 SHA3-256 和其他变体的 5 轮碰撞攻击以来,没有其他显著的进展被公布。这种碰撞攻击扩展到更多轮次的挑战在于差分路径搜索的低效性。为了克服这一障碍,在亚密会 2022 上,郭建等人^[60]开发了一个基于 SAT 的自动化搜索工具。该工具被用于碰撞攻击的多个中间步骤,并在差分路径搜索和该过程中遇到的其他优化问题中表现出高效性。利用这一工具,郭建等人提出了对 SHAKE128 的第一个 6 轮经典碰撞攻击,时间复杂度为 $2^{123.5}$ 。同时对 6 轮的 SHAKE128、SHA3-224 和 SHA3-256 形成了量子碰撞攻击,量子时间分别为 $2^{67.25}/\sqrt{S}$ 、 $2^{97.75}/\sqrt{S}$ 和 $2^{104.25}/\sqrt{S}$,其中 S 代表量子计算机的硬件资源。经典碰撞攻击不适用于 6 轮 SHA3-224 和 SHA3-256 的事实表明,适用于量子碰撞攻击的差分覆盖范围更广,这与 Hosoyamada 和 Sasaki 在美密会 2021 上观察到的 SHA-2 的情况一致。碰撞攻击所采用的方法框架依然延续了 JOC 2020 上的方法框架,将攻击分为 3 个部分:高概率碰撞路径的搜索、高自由度连接器的构造和穷搜。SAT 方法应用于前两个步骤。为了缓解模型求解的压力,在碰撞路径的搜索上,只对活跃 S 盒的个数而非概率进行限制;在连接路径的搜索上,分两个阶段分别对不同轮的 S 盒的差分传播概率进行限制。差分路径搜索是杂凑函数和对称密码分析的关键环节,基于 MILP 和 SAT 的自动化搜索虽然给求解这一问题提供了一种直观的思路,但仍需要许多启发式和经验式的技巧控制模型规模使得问题可解。对 SHA-3 系列的攻击来说,大状态无疑给差分路径的搜索带来更多困难,这篇文章给出的基于 SAT 的搜索范式给后续工作提供了有效借鉴。

在 FSE 2022 会议上,倪博煜等人^[75]建立基于混合整数线性规划 (MILP) 的自动化搜索模型,用于计算 Simpira v2 的差分 and 线性最小活跃 S 盒个数。Simpira v2 是 Gueron 和 Mouha 在亚密会 2016 上提出的密码置换函数,其包含多个分组大小版本。基于该置换函数与典型哈希模式 Davies-Meyer 结构,Guéron 和 Mouha 构造了适用于短输入消息的哈希函数,该函数可用于基于哈希函数的签名系统等。作者发现 Simpira v2 置换设计者自分析给出的截断差分

路线存在局部不兼容性, 导致其给出的路线不存在。作者将该性质纳入 MILP 模型中, 计算出了比 Simpira v2 设计者自分析更紧的差分 and 线性活跃 S 盒下界。进一步, 作者利用该模型, 给出了两分支和四分支 Simpira v2 的多个新型截断差分路线并用于构造反弹攻击, 基于此作者给出了这两个版本 Simpira v2 的经典与量子碰撞攻击。对于两分支版本, 作者给出了 9 轮的经典和量子碰撞攻击。针对 4 分支版本, 作者给出了 11 轮经典和量子碰撞攻击。

4.3 量子攻击

对于对称密码的后量子安全性, 由于 Grover 量子搜索算法及其衍生算法一般认为只能达到二次提速, 因此通常认为将密钥大小翻倍是一种足够的预防措施。

在欧密会 2022 上, Xavier Bonnetain 等人^[54]的工作对这一结论造成了冲击。Xavier Bonnetain 等人第一次对分组密码进行了与最好的经典攻击相比有超过二次时间加速的量子密钥恢复攻击, 且只使用了经典查询。攻击对象是扩展的 FX 结构, 其中一个重要代表是 Gazi 和 Tessaro 在欧密会 2012 上提出的 2XOR-级联 (2XOR-cascade) 结构。这种结构是一种密钥长度扩展技术, 通过串联两个密钥长度为 $2n$ 比特的 n 比特分组密码并异或白化密钥得到一个 $2.5n$ 比特安全性的 n 比特分组密码, 且有理想模型下的安全性证明。本文作者对这种构造进行了 $\tilde{O}(2n)$ 时间的量子攻击, 与最佳经典攻击相比有 2.5 次的量子加速。攻击所用方法为 Bonnetain 等人在亚密会 2019 上提出的 offline-Simon 算法的扩展。这就意味着利用一些对称构造的结构可以突破 Grover 搜索二次提速的限制。同时意味着, 2XOR 级联结构只能提供与分组密码本身相同的安全性, 无法普遍用来增强分组密码的抗量子攻击能力。

在允许量子询问的场景下, 量子攻击可以实现指数加速, 但是在经典询问场景下, 很长时间的认知里对对称密码分析的最佳量子加速都停留在二次, Xavier Bonnetain 等人^[54]的工作将这一差别提升到了 2.5 次。Xavier Bonnetain 等人猜想 offline-Simon 算法可以对某些合适的问题实现 3 次加速, 但还没有发现相应的密码学场景, 这一极限到底为多少其实是公开问题。在增强对称密码算法的抗量子攻击能力上, 应仔细分析通用的密钥长度扩展技术, 探究是否存在特定的量子攻击削弱了对安全界的提升。

在量子计算领域, Simon 算法可以在多项式时间内找到函数的周期, Grover-meets-Simon 算法可以在一个函数族中搜索到一个周期函数及其周期。该算法用于对称密码分析时需要要对加密函数进行量子询问。offline-Simon 算法对此改进, 可以对加密函数进行经典询问。这些算法被用来攻击多种对称密码构造。

在 CHES 2022 上, Xavier Bonnetain 等人^[55]首次提出了 offline-Simon 算法的量子电路的完整描述, 并估计其攻击 MAC 方案 Chaskey、分组密码 PRINCE 和 NIST 轻量级 AEAD 入围方案 Elephant 的成本。Xavier Bonnetain 等人设计并实现了多种有独立意义的量子电路, 包

括：求解布尔线性方程的高效可逆电路；Chaskey 算法、PRINCE 算法的优化量子电路；Elephant、SPONGENT、KECCAK 算法中用到的两个置换的优化量子电路。在这些实现的度量下，对 PRINCE 和 Chaskey 的密钥恢复攻击需要 2^{65} 个量子操作，是目前已知最优的量子攻击。这说明了它们的安全性仅比 RSA-2048 高 30 多个比特（Shor 算法攻破 RSA-2048 需要 2^{31} 个类似的操作）。Elephant 算法由于状态更大，攻击的花销比较接近甚至超过穷搜。Xavier Bonnetain 等人建议为了抵抗 offline-Simon 攻击实现量子安全性，应该用更大的状态，而不只是更长的密钥；不应依赖对数据量的限制，因为其对量子攻击的影响很小；不要使用 Even-Mansour 和 FX 结构。

密钥的量子搜索并行性较差，再考虑到纠错产生的额外开销，一般认为并不构成对称密码的实际威胁。而作者给出了一种更具可行性的量子攻击途径，只比 RSA-2048 多 32 比特的安全性完全可以认为在后量子时代能被完全攻破。而且经典询问+量子计算可以形成在未来破译今天的通信系统的攻击场景，即一个有耐心的攻击者在今天收集足够多的数据，等未来有足够的量子计算资源时进行攻击。因此，为以长期安全性为目标的应用场景选择对称密码方案时要十分谨慎。

Simon 算法可以在量子计算模型下以多项式时间复杂度求出布尔函数的最小周期。因此，在量子计算模型下，人们可以基于密码体系构造具有周期性的布尔函数，使得该布尔函数的周期与密钥之间存在某种联系，然后利用 Simon 算法高效求出函数周期，从而得到密钥相关的信息^[56-58]。显然，构造基于密码体系的周期布尔函数是此类分析的关键，也是技巧性最强、最容易出现错误的部分。

在美密会 2022 上，Federico Canale 等人^[59]的一系列新技术与新理论，加深了人们对基于 Simon 量子算法的密码分析技术的理解。该工作的意义在于首次提出了自动化搜索周期布尔函数的方法，并利用这一方法，首次给出了 5 轮 Feistel-FK 和 5 轮 MISTY L-FK 的基于 Simon 算法的高效率密钥恢复攻击结果。该自动化搜索方法通过将密码体系描述成特殊的电路图结构，将函数的输入、输出以及基本运算统一看作是电路图上的“门”，然后遍历所有由 1,2,3,... 个门构成的电路图，判断其是否具有周期性，从而实现对周期函数的搜索。此外，该工作还讨论了将非标准哈达玛变换运用到 Simon 算法中进行密码分析的可行性，从理论上证明，非标准哈达玛变换无法直接套用现成的 Simon 算法进行对称密码分析，未来必须要设计与之相应的新型量子算法。

在欧密会 2020 上，日本学者 Hosoyamada 和 Sasak 首次提出针对 AES 类哈希函数的量子碰撞攻击。在亚密会 2020 上，我国学者董晓阳、孙思维等人进一步改进了该量子碰撞攻击算法。这些攻击均基于经典反弹攻击的量子化版本，其能够利用经典反弹攻击所不能利用的低概率差分路线。在 FSE 2022 会议上，Amit Kumar Chauhan 等人^[61]进一步研究基于 Hirose 双分组哈希模型下 AES-256（记为 HCF-AES-256，全轮 14 轮）的量子碰撞攻击。针对该算

法,经典攻击方法能够构造 9 轮 HCF-AES-256 的自由起始碰撞攻击。Amit Kumar Chauhan 等人给出了一个 10 轮的概率为 2^{-160} 的差分路线,并基于此构造了量子化的反弹攻击。该量子碰撞攻击所需的时间复杂度是 2^{85} ,所需的量子存储复杂度是 2^{16} 。另外, Amit Kumar Chauhan 等人还给出了一个不需要量子存储的量子碰撞攻击,时间复杂度是 2^{86} ,该攻击超越了 Chailloux、Naya-Plasencia 和 Schrottenloher 给出的通用量子碰撞攻击。

基于 AES 结构设计哈希函数是现在哈希函数主流设计方法之一。著名的算法包括 Whirlpool、Grøstl、ECHO、AES-MMO 等。反弹攻击 (Rebound Attack) 是由著名密码学家 Florian Mendel 等人于 2009 年提出的,是构造 AES 类哈希函数碰撞攻击的重要方法。该攻击一般包括一个 Inbound 阶段和两个 Outbound 阶段。如何拉长反弹攻击的轮数是一个重要的密码分析难题。相关研究包括 Gilbert 和 Peyrin 在 FSE 2010 会议上提出的超级 S 盒技术,以及 Sasaki 等人在亚密会 2010 上提出的非满活跃盒的超级 S 盒技术。然而,这些技术仅能将 Inbound 阶段扩展至 2 轮。

基于上述工作,董晓阳等人^[66]在美密会 2022 上提出了名为对角化反弹攻击 (Triangulating Rebound Attack) 的新型方法,该方法利用压缩函数中的消息自由度链接多轮 Inbound 阶段,极大地延长了反弹攻击的轮数。链接过程需要建立并求解一组关于消息的非线性方程组,董晓阳等人利用类似高斯消元 (对角化, Triangulation Algorithm) 方法,通过合理消耗消息自由度,高效地求解非线性方程组,从而成功构造多 Inbound 阶段的反弹攻击。该方法能够有效构造多轮 Inbound 阶段,被命名为超级 Inbound (Super-Inbound),而且该技术仅需要少量的内存,这使得该方法也能用于构造低量子存储的量子碰撞攻击。同时,董晓阳等人提出了基于可满足性问题 (Constrained Programming) 建立自动化搜索模型,高效地搜索多轮 Inbound 的反弹攻击路线。基于该方法,董晓阳等人首次将 AES-MMO 的碰撞攻击突破至 8 轮。另外,还给出了 Saturnin-Hash、Grøstl-512、SKINNY-128-384-MMO/MP 等哈希函数的多个减轮的经典和量子碰撞攻击。

将攻击对称密码使用的量子算法综合成量子电路,是精准评估量子攻击复杂度的重要手段。在亚密会 2022 中,黄震宇等人^[79]针对对称密码结构及组件的特点,给出了若干量子电路综合与优化的通用方法。第一,提出了利用出位 (Out-of-place) 子电路原位 (In-place) 实现迭代型对称密码可逆轮函数的通用方法。第二,给出了基于逻辑可满足问题 (SAT) 优化量子电路 CNOT 网络的方法。对于规模较小的线性量子电路,该方法可输出 CNOT 门数量达到理论最小值的量子电路。第三,提出了降低非线性向量布尔函数对应的量子电路的 T 门深度的一般方法。基于上述方法,黄震宇等人在国际上率先得到了 T 深度达到理论最低值的 AES S 盒的量子电路,并以此将微软量子计算团队在欧密会 2020 上给出的 AES 量子电路的 T 门深度从 60 降低到 30,将量子比特数从 512 比特降低到 371 比特,将 Toffoli 门数从 19788 降低到 17888,并将 Toffoli 深度从 2016 降低到 1558。

4.4 重要对称密码算法的分析结果

SHA 算法族是应用最广泛的杂凑函数族之一，并且已经被 NIST 形成 FIPS 标准。截至目前，共有 3 代 SHA 算法标准。SHA-1 算法由 NSA 在 1995 年设计，但在 2005 年被王小云等人证明不安全。SHA-2 系列算法由 NSA 在 2001 年设计，但由于它们与 SHA-1 算法有相似之处，因此其安全性仍然存在不确定性。鉴于此，NIST 在 2008 年发起了公开竞赛以征集新一代杂凑函数标准。经过 4 年的细致审查，Keccak 函数从 64 个参选算法中胜出并于 2015 年被确定为 SHA-3 标准。

Keccak 杂凑函数由 Bertoni 等人设计。该函数采用了一种新的海绵结构而不是 SHA-1 和 SHA-2 采用的 Merkle-Damgard 结构。Keccak 的设计者证明了，如果其内部置换 Keccak-f 可以被视作伪随机置换，那么海绵结构构造具有不可辨别性和抗碰撞性。尽管理论上不能证明 Keccak-f 是伪随机的，但是 24 轮的混合置换仍然可以很好地抵抗所有的密码分析。

关于 Keccak-f 的伪随机性，目前主要有两类区分器：差分特征区分器、零和区分器。在 FSE 2012 会议上，Duc 等人利用多条 5 轮的差分路径得到了一个 8 轮的差分区分器，其攻击的复杂度为 2^{491} 。2009 年，Aumasson 等人提出了关于 Keccak-f 的零和区分器，可以实现 9 轮的实际攻击和 15 轮的生日攻击。在 FSE 2011 会议上，Boura 等人提出了一个复杂度为 2^{1590} 的全轮零和区分器。关于碰撞攻击的大部分研究结果都依赖于差分迹。在 FSE 2012 会议上，Daemen 等人分析了 Keccak 的差分性质。随后，Kölbl 等人更深入地研究了 Keccak-f[800] 和 Keccak-f[1600] 的差分性质。2012 年，Dinur 等人提出了目标差别（Target Difference）算法。基于这项技术，他们得到了 4 轮 Keccak-224/256 实际碰撞。其后，乔珂欣、宋凌、郭建等人改进了这项技术并构造 5-轮 Keccak-224/256 实际碰撞。

2011 年，Naya-Plasencia 等人基于差分分析提出了首个 2 轮 Keccak 的原像和碰撞攻击。此后，研究者很快开始使用代数方面的技术对 Keccak 进行分析。例如，2012 年，Dinur 等人给出了一个碰撞搜索以及 4 轮实际攻击；2013 年，Morawiecki 等人提出通过 SAT 攻击的方法进行原像搜索。在 FSE 2013 会议上，Dinur 等人的论文讨论了碰撞搜索，并找到了 3 轮 Keccak-384 和 Keccak-512 的首个碰撞。以 Dinur 等人的工作为基础，乔珂欣等人在欧密会 2017 上给出了 Crunchy 竞赛算法实例的新的攻击；宋凌等人在美密会 2017 上给出了 5 轮 Keccak-224、6 轮 Keccak[1440,160] 的首个实际攻击，攻击使用了 160 比特摘要。最终，2016—2017 年，关于 Crunchy 竞赛中 Keccak 实例的一系列碰撞被找到，其后发表于密码学杂志上。自 2016 年起，郭建等人引入线性结构后，线性化技术就成为原像攻击的关键。在此工作的基础上，随后几年密码分析者又引入了交叉线性结构，并分析了 Crunchy 竞赛中的其他算法实例。对于更大规模的 Keccak 实例，Kumar 等人在亚密会 2018 上同样使用线性技术给出了

Keccak-512 的首个实际攻击。随后, Kumar 等人又改进了 2 轮 Keccak-284 的攻击。2019 年, Rajasree 改进了对于 3 轮和 4 轮 Keccak-384 和 Keccak-512 的攻击。同年, 李婷等人给出了 3 轮 Keccak-224 的首个实际攻击, 以及对于 3 轮和 4 轮 Keccak-224 和 Keccak-256 的非实际攻击。

关于 Crunchy 竞赛的现状, 目前仅有最小的 1 轮版本, 即 Keccak[40,160], 通过抵消轮常数可以被成功攻击。因此, 也有研究者建议将 Keccak 的最小版本应用于受限环境。目前, 最小版本实例的置换, 即 Keccak-p[200,nr]和 Keccak-p[400,nr]也用于构建认证加密算法, 如 KETJIE、ISAP 和 Elephant。因为目前研究者主要关注作为标准的 Keccak 实例, 所以很多的分析都针对 1600 比特状态而非 200 比特或 400 比特状态。因此, 目前还缺乏 Keccak 的小规模实例对碰撞攻击的安全性分析。

在 FSE 2022 会议上, Rachele Heim Boissier 等人^[63]研究了上述问题。该工作表明代数分析不仅可以用于原像攻击, 还可以用于碰撞攻击。Rachele Heim Boissier 等人使用了 FSE 2013 会议上提出的挤出攻击来生成内部状态碰撞。因此, 攻击可以适用于输出长度被扩展情况, 或者 Keccak 被当作 XOF 使用的情况。该工作也利用了一些具体性质, 如基于 θ 变换的特殊性质, 可以控制超过 1 轮的扩散; 通过分析 χ 变换的性质, 可以导出必要的线性方程, 并用它们计算满足内部状态属于输出集合的一个子集的 Keccak 状态。对于宽度较小的多个 Keccak 版本, 本文提出的攻击要优于已有的方法。事实上, 之所以已有的方法对 Keccak 其他版本实例的攻击多数有效, 是因为这些版本的比率允许相对较大的自由度。但是对于在搜索 Keccak[40,160]、Keccak[72,128]和 Keccak[144,256]的内部碰撞时却不适用。

在 FSE 2022 会议上, 何乐等人^[62]研究了对 Keccak 的原像攻击。围绕着 4 轮 Keccak-224/256 的最后两轮。对于第 3 轮, 何乐等人提出了一种高效的部分线性化策略, 可以重复使用某些公共的自由度。基于此策略, 可以使用 162 以内的自由度线性化特定位置 (Site) 的 56 比特, 并且可以穿过第 3 层 S 盒到达第 4 层 S 盒。对于第 4 轮, 何乐等人给出了 224/256 个输出比特和 56 个线性比特之间的关系。何乐等人全面地分析了线性位置可行与不可行的情况, 并通过随机摘要方式计算了 56 个位置的所有期望增益。综合部分线性化策略和对线性结构的期望分析, 何乐等人给出了 4 轮 Keccak-224/256 的完整原像攻击, 其复杂度为 $2^{192}/2^{218}$ 。

在 FSE 2022 会议上, 基于李婷和孙瑶^[64]之前对于 Keccak/SHA-3 的原像攻击, 林晓恩等人^[65]使用一种改进的策略, 给出了一种针对标准 3 轮 Keccak-224/256 进行实际复杂度的原像攻击。文献[64]中的策略是通过分两阶段构造第一和第二个消息块 (Block), 从而找到一个两块原像, 而不是之前常用的一块原像。在这个策略下, 李婷和孙瑶设计了一个具有 194 个自由度的新线性结构, 用于 2 轮 Keccak-224/256, 能够以 $2^{31}/2^{62}$ 的复杂度构造第二个消息块。然而, 该策略的瓶颈在于第一阶段的复杂度远大于第二阶段。因此, 林晓恩等人通过两种技术改进了第一阶段的复杂度。第一种技术是在第一阶段构建多块消息, 而不是一块消息, 以

达到更好的内部状态。林晓恩等人使用了迭代技术来生成第一个所需要的块。这个技术可以看成李婷和孙瑶的技术的进一步推广。第二种技术是设置更好的限制方程，这有助于对 3 轮 Keccak-256 的攻击。最终，3 轮 Keccak-224/256 原像攻击的复杂度从 $2^{38}/2^{81}$ 降低到 $2^{32}/2^{65}$ 。

Ascon 认证加密 (Ascon AEAD) 算法^[67]是美国国家标准技术局 (NIST) 发起的轻量级密码算法竞赛的第二轮候选算法 (Ascon 已于 2023 年 2 月被选为最终获胜算法)。Ascon AEAD 使用 MonkeyDuplex 结构^[68]，包含了初始化、吸收关联数据、加密和生成标签 4 个阶段。本文针对初始化阶段进行分析。Ascon AEAD 的初始化阶段接受 64 比特的 IV、128 比特的密钥和 128 比特的新鲜值 (Nonce)，之后运行 12 轮的 Ascon 置换得到输出。

之前对初始化阶段的分析达到 7 轮，主要有两种：一种是在 Nonce 重复使用的情形下攻击^[69]；另一种是条件 Cube 攻击^[70]，可以在 Nonce 不重复使用的情形下攻击。Nonce 不能重复使用是 Ascon AEAD 明确声明的；另外，Ascon 要求在同一个密钥下最多可以使用 2^{64} 个不同的 Nonce，但是文献[4]中使用的 Nonce 达到 2^{77} 。所以，之前的两个攻击都没有正确处理 Ascon 的使用限制。

在 FSE 2022 会议上，Raghvendra Rohit 等人^[71]给出了第一个满足了 Ascon 使用要求的 7 轮攻击。Raghvendra Rohit 等人使用 Cube 攻击，首先恢复出 64 次项的超级多项式，之后通过求解 64 个方程组恢复出 64 比特密钥。为了恢复出超级多项式，Raghvendra Rohit 等人使用了 Ascon 置换的一个重要性质：如果选择初始化的最后 64 比特输入作为 Cube 变量，那么 2^k 次项都只包含 2^k 个密钥信息。所以，对于 6 轮算法的输出的最高次项 (32 次项) 可以通过遍历的方式得到全部的超级多项式。而 Ascon 的 S 盒是 2 次，所以 7 轮的 64 次项的超级多项式可以通过组合所有可能的 6 轮 32 次项得到。通过一些并行技术，最终可以在 2^{123} 的复杂度下恢复 7 轮 64 次项的超级多项式。此外，Raghvendra Rohit 等人还改进了 Ascon AEAD 的区分攻击。Raghvendra Rohit 等人发现如果令 Ascon 的前 64 比特 Nonce 和后 64 比特 Nonce 相等，那么 Ascon 初始化的代数次数可以显著降低。最终，Raghvendra Rohit 等人给出了 7 轮 2^{60} 的区分器。该工作是目前在不误用 Ascon AEAD 设定的情况下最好的攻击。

在 FSE 2022 会议上，David Gérardt 等人^[72]针对 Ascon 认证加密家族及其内部置换，探讨了使用约束编程 (CP) 进行差分密码分析的方法。首先，David Gérardt 等人提出了一种使用 CP 自动找到 Ascon 优秀差分特征的搜索方法，并表明 CP 能够有效地产生优秀差分结果，与专用搜索启发式相比具有优势。基于他们的工具，David Gérardt 等人还参数化 CP 中的搜索策略以生成其他差分特征，目标是为 Ascon 内部置换的 4、5、6、7 轮形成有限生日区分器和矩形攻击。为了更好地区分这些区分器，David Gérardt 等人提出将它们分为黑盒和非黑盒两类。在这一基础上，他们使用 CP 找到的差分特征构建了 Ascon 置换的黑盒/非黑盒有限生日区分器和矩形区分器，涵盖了 4、5、6、7 轮的 Ascon 置换。此外，David Gérardt 等人再次利用他们的 CP 工具并进行特定参数化，找到了可以改进针对减轮 Ascon-128 和 Ascon-128a

伪造攻击的差分特征,相较于撰写时报告的最佳结果有所提高。最后,他们利用找到的最佳 2 轮差分特征,改进了最近对减轮 Ascon-Hash 的攻击。总体来说,David G rault 等人通过使用约束编程 (CP) 的方法,对 Ascon 认证加密家族进行差分密码分析,提出了一系列区分器和攻击方法,丰富了密码分析领域的研究工具和方法。虽然这项工作并未威胁到 Ascon 的安全性,但它有助于更好地了解这一候选算法在应对基于差分密码分析的攻击方面的抵抗能力。

在 FSE 2022 会议上,Erlacher 等人^[73]针对 Ascon 的差分和线性安全界进行了研究,提出了基于“束带模式”将整个搜索空间划分为大量子问题的搜索技术,并结合组合数学的项链理论给出利用 Ascon 算法中循环对称性质的方法。该工作的重要意义在于,提出了一种新的基于 SAT 问题的模型刻画策略,给出了约减轮 Ascon 算法更紧的差分和线性安全界,评估了纯 SAT 模型下的计数器实现和模型并行运算问题,为 SAT 模型的应用提供了有效指导。

在 FSE 2022 会议上,Raghvendra Rohit 等人^[74]给出了对 Ascon 的 7 轮区分和密钥恢复攻击,该攻击遵循 nonce 不重用的声明和每个密钥最多加密 2^{64} 个数据的限制。然而,其区分器需要 (不切实际的) 2^{60} 个数据,以及密钥恢复攻击的数据复杂度等于 2^{64} 。对于 7 轮的 Ascon,是否存在实际的区分器以及密钥恢复攻击 (数据少于 2^{64}) 仍然是一个开放问题。Raghvendra Rohit 等人对 Ascon 在弱密钥环境下的安全性进行了详细分析,并回答了上述问题。该工作首先给出了 7 轮 Ascon 的实际积分区分器,并改进了 5 轮、6 轮、7 轮的已有区分器的复杂度;然后给出了一个仅基于代数次数的 Ascon 弱密钥空间的理论构造框架;最后对 7 轮 Ascon 进行了两个密钥恢复攻击,其中第二个密钥恢复攻击的数据复杂度低于 2^{64} 。他们的研究表明,在对称密码分析领域,部分非线性层结构以及线性的密钥编排算法仍然没有得到充分理解,并且如何利用代数技术、自动化建模技术对算法进行分析仍然是一个重要问题。

4.5 新型对称密码分析

安全多方计算 (MPC)、零知识 (ZK) 证明和全同态加密 (FHE) 是近几年密码学界的研究热点,显示出了良好的应用前景。LowMC (欧密会 2015)、Rasta (美密会 2018) 等都是面向 MPC/FHE/ZK 的应用场景而专门设计的对称密码算法。其中,分组密码 LowMC 作为底层加密方案被应用于后量子签名方案 PICNIC (NIST 后量子密码竞赛第三轮备选算法) 中。在 PICNIC 数字签名方案中,LowMC 生成的明-密文对用作签名公钥,相应的加密密钥用作签名私钥。因此,在对 LowMC 的密码分析时,研究攻击者只能访问单个已知明-密文对的攻击情境对讨论 PICNIC 数字签名方案的安全性尤为重要。但这是一个具有挑战性的数学问题,因为在这种极端情况下,攻击者无法使用绝大多数对称密码标准分析技术,如线性分析和差分分析。

在 FSE 2022 会议上, Subhadeep Banik 等人^[76]给出了对 LowMC 的密钥恢复攻击。这些攻击都是在已知明文模型下对两个 LowMC 变体 (LowMC 密码分析挑战中的实例) 的复杂度为 1 的攻击。该工作首先给出了如何通过猜测一个关于 S 盒输入的二次表达式来有效地线性化 LowMC 的 S 盒。然后利用这个线性化技术, 对具有满 S 盒层的 2 轮 LowMC 实例与具有部分 S 盒层的 $0.8 \times \left\lfloor \frac{n}{s} \right\rfloor$ 轮的 LowMC 实例进行了攻击, 其中 n 代表分组长度, s 代表每轮中 S 盒的数量。最后, 该工作还将此线性化技术与中间相遇技术、3-xor 问题结合, 对上述的攻击进行了优化。

在亚密会 2022 上, 刘富康等人^[77]提出了一项“代数中间相遇攻击”的新技术, 通过利用部分非线性层的特征, 分析 LowMC 的安全性。之前在美密会 2021 上, 刘富康等人提出的代数技术可以从 LowMC 的差分路径中恢复全部密钥, 但其时间复杂度仍然是密钥长度的指数级。亚密会 2022 中的技术可以降低简单差异枚举攻击的存储复杂度, 另外, 可以在差分路径上存在足够多活跃 S 盒的情况下把时间复杂度降为常数时间。借助上述新技术, 作者改进了对 LowMC 和 LowMC-M (CRYPTO 2021) 的攻击, 并且首次攻破了 LowMC 的一些实例。

4.6 其他分析技术

差分故障分析 (DFA) 是对对称密码实现有力攻击的方式。鉴于 DFA 攻击具有非常实际的影响, 设计可抵御 DFA 的保护密码实现的策略具有高度的现实意义。在过去几年中, 人们已经探索了多种反制 DFA 的措施, 其中大多数实施于密码算法的实现层面。但是在亚密会 2021 上, Anubhab Baksı 等人^[37]提出了新颖的反制措施, 即使用具有线性结构的 S 盒来设计具有抵抗 DFA 固有属性的密码。这种策略期望以固有的方式限制攻击者可以学习到的信息量。与此同时, 他们给出了相应的分组密码实例 DEFAULT。设计者声称, 对于 DEFAULT, DFA 敌手最多只能学到 128 个密钥比特中的 64 个, 因此通过暴力破解方式恢复剩余密钥比特的复杂度是 2^{64} , 是不切实际的。

在欧密会 2022 上, Marcel Nageler 等人^[36]指出 DEFAULT 的设计存在一些漏洞。虽然每轮只能获得有限的信息量, 但可以跨轮组合信息。直觉上看来, 一个简单的密钥编排就可以防止对 DEFAULT 的信息组合攻击。因此, DEFAULT 密码的设计者提出使用具有全扩散性的强密钥编排, 以保证无法结合相邻两轮的信息, 这确实有助于防止朴素而直接的信息组合 DFA 攻击。但 Marcel Nageler 等人发现, 在有强密钥编排的情况下, 信息组合攻击仍然是可行的。Marcel Nageler 等人观察到, DEFAULT 这样的分组密码存在大量的等价密钥, 可以利用轮函数的性质, 使用线性方程以规范化的形式有效地表征这些密钥。这些规范化的密钥可以被用来将多轮信息结合起来, 从而破解有强密钥编排的 DEFAULT, 甚至是具有独立轮密

钥的理想化版本。攻击的一个关键点是，从密文开始，在每一轮，敌手要么可以获知关于轮密钥的一些额外信息，要么选择剩余候选值中的任意一个并继续反解下一轮。Marcel Nageler 等人通过优化 Fault 的植入位置，将 fault complexity 降低到少于 100 个 faulted computations，以恢复 DEFAULT-LAYER 的等价密钥，而无须额外的暴力破解。

这些工作表明，防止针对密码实现的 DFA 攻击具有极大的挑战性。虽然提供在密码设计层面上的保护措施将是一个非常好的解决方案，因为这样不会将责任推卸给密码算法的实现者，但这似乎需要采用与“线性结构”等方法有本质区别的想法。因此，如何调整 DEFAULT 设计方案以实现具有抵抗 DFA 攻击的固有安全性，将是下一步工作的重点。

对于现代对称密码，使用机器学习取得突破性分析成果的实例非常罕见。直到美密会 2019，来自德国联邦信息安全办公室的研究人员 Aron Gohr^[38]发表了基于深度学习分析现代密码的成功实例。Aron Gohr^[38]的工作表明，训练有素的神经网络可以执行密码分析中的区分任务，达到优于基于全差分分布的区分器的准确率。此外，应用非正统的密钥猜测策略，对现代分组密码 Speck32/64 缩减至 11 轮进行密钥恢复攻击，在保持相当的数据复杂度的同时，时间复杂度较已知最优攻击结果有明显的优势。这就引出了进一步的问题：基于机器学习 (ML) 的密码分析相对于传统方法的优势有多大？这种优势是否普遍存在于对现代密码的分析中？

为了回答这两个问题，在亚密会 2022 上，包珍珍等人^[39]基于 Gohr 的工作，进一步探索了差分-神经密码分析方法的优势。工作的一个关键点是增强基于 ML 攻击中的经典密码分析部分。利用同步中性比特集合 (SNBS)、条件中性比特 (CSNBS)、邻接差分路径中性比特 (SBfAD) 等多种广义中性比特，基于差分-神经网络区分器和贝叶斯优化密钥猜测过程^[1]，实现了对 Speck32/64 的首个 13 轮实际密钥恢复攻击，12 轮优化攻击。攻击有很多可调的参数，了解攻击过程中的一些变量的分布有助于合理选取关键参数。因此，工作^[1]也对攻击中一些关键变量的分布以及攻击参数对分布的影响进行了实验性的研究，给出了调节攻击参数的经验性结论。特别地，为了回答第二个问题，包珍珍等人^[39]给出了对 Simon32/64 的多种神经网络区分器及其与全差分分布区分器的对比和解释，并给出了对 Simon32/64 缩减到 16 轮的基于神经网络的实际密钥恢复攻击。因为注意到，对于 Simon，已知 r 轮输出的值，可以不用密钥就计算出 $r-1$ 轮输出的差分。所以，包珍珍等人^[39]提出，如果进行比较，应当用 r 轮接收密文对值的区分器与 $r-1$ 轮接收密文差分的区分器进行比较。在这样的比较下， r 轮接收密文值的神经区分器、 $r-1$ 轮接收密文值部分差分的神经区分器，并不比 $r-1$ 轮全差分分布区分器准确率更高。因此，给定密文值，神经网络有能力捕获反解“无密钥”的一轮操作之后中间数据的差分分布特征，但不会学习到更多的特征。

通过这些初步的探索，可以看到差分-神经密码分析大体上适用于不同类型的现代密码算法。通过增强传统密码分析和基于机器学习方法的结合方式，可以得到更好的密码分析结

果。最近机器学习取得了惊人的成功,但是,在这些成功案例中,机器学习只是作为更大更复杂问题系统的一个关键部分,需要与其他部分进行有效的结合,对于密码分析问题,更是如此。

目前基于机器学习的密码分析还有很多公开问题,如当考虑攻击尽可能多的轮数时,怎样进行理论投射,从对缩短轮的实际攻击的优势来理论评估对长轮攻击的优势呢?如何设计一个严谨的理论模型来刻画各种攻击参数与攻击复杂度和成功概率之间的关系,以便使用最好的攻击参数得到最优的攻击呢?差分-神经密码分析是选择明文攻击,那么如何基于机器学习方法构造已知明文区分器呢?还有哪些机器学习方法也能适用于这些密码分析问题呢?如何解释有效的机器学习结果,以便可以反过来提升传统密码分析技术呢?我们能用人工智能中哪些更高级的方法来更进一步地提升密钥搜索策略呢?这些都是下一步工作的内容。

后门(Backdoors)和故意弱点(Intentional Weaknesses)是两个重叠的且由来已久的有趣的密码学话题。之所以是由来已久的有趣的话题,有多方面的原因。首先是政治原因,如执法的原因等;其次是实际部署上的原因,如 DES、DualEC、GEA-1 这些应用实例;最后是从学术角度来看它也很有趣,如构造专门的设计和研发通用的设计框架。需要说明的是,后门和故意弱点并不是一件好事,实际上非常糟糕,但是对它们的公开学术研究确实是有价值的,因为可以更好地理解它们,以便能够更好地预防它们。继美密会 2020^[40]提出的构造具有故意弱点的可调分组密码框架 MALICIOUS 之后,以及欧密会 2022^[41]给出的对广泛部署的流密码算法 GEA-1 的实际攻击之后,在美密会 2022 上,Christof Beierle 等人^[42]对此课题做出了双重贡献。一方面是对解构现有故意弱点的研究,具体就是解释了 GEA-1 密码是如何被植入弱点的,以及它是如何被构造的,这是对欧密会 2022^[41]的工作所留下的公开问题的完整解答;另一方面,此工作展示了如何构造具有后门的可调分组密码,给出的构造方式更自然,因为它完全符合现代分组密码的设计原则。

具体地,一方面,流密码 GEA-1 被广泛应用于早前版本的 GSM,用于加密手机通信数据。继欧密会 2022^[41]的工作显示出 GEA-1 具有特殊的弱点,即经过线性注入会话密钥初始化之后,如果 3 个 LFSR 中的两个拿来一起看,只有 2^{40} 种可能的状态,而从整个密钥空间上看,它们本应该能有 2^{64} 种可能的状态。因此,联合熵的损失很大。而这种经过相消导致熵损失的可能性实际上是很小的。实验表明,在百万分之一的随机尝试中,最高的熵损失是 9 比特,而 GEA-1 的熵损失竟然达到 24 比特。因此,这种熵损失性质不是设计者不走运而发生的,而是设计者故意选择的结果。如何设计 GEA-1 这样的两个 LFSR,使得能产生这么大的熵损失实际上都是非常困难的问题。在美密会 2022 上,文献[42]的工作展示了一种构造这种弱点的技术。作者认为,这种方式应该就是 GEA-1 的设计者当初所采用的方式。技术框架就是研究向多个 LFSR 注入会话密钥时采用不同移位所带来的影响,并从满足熵损失出发,以更聪明的方式搜索 LFSR 的特征多项式。利用不同移位向不同 LFSR 注入会话密钥在很多流

密码的设计中都使用，如 GEA-1、GEA-2、A5/1。而这一工作显示，通过移位，再结合对反馈多项式和过滤器的特殊选择，就可以构造带故意弱点的流密码了。

另一方面，如何设计具有后门的可调分组密码。该工作基于美密会 2020^[40]上发表的 Malicious 构造框架。Malicious 构造框架的基本思想是使用特定的 Tweak，并将其散列，使用散列值来构造具有概率为 1 的差分传播的 LowMC 变体。而美密会 2022^[42]的工作不再使用一对 Tweak 来构造具有概率为 1 的差分传播，而是只使用一个 Tweak 来构造不变子空间（Invariant Subspace）或非线性不变子（Nonlinear Invariant）。基于构造不变子的这一基本思想，文章给出了两个具体工作：一是提出了 Malicious-AES，利用了人们很早就知道的 AES 的不变子空间，调整了密钥编排算法，使得使用一个 Tweak 会得到一个非常弱的密码，而对于其他 Tweak 会得到安全性非常强的密码；二是提出了 Boomelang，利用了连续两个轮函数的非线性不变子，这种非线性不变子很难被检测到，并且没有公开的工具可以自动检测这种弱点。但是本文提供的设计方法尚未达到美密会 2020^[40]提出的 4 个设计指标 [不可检测性（Undetectability）、不可发现性（Undiscoverability）、不可追踪性（Untraceability）、实际有效性（Practicability）] 中的不可追踪性。因此，为了达到不可追踪性，设计可以被多个 Tweak 开启的后门是下一步工作的内容。

算法替换攻击（Algorithm Substitution Attack, ASA）由 Bellare 等人在美密会 2014 上提出。在 ASA 中，攻击者有以下目标：将加密函数用一个替换版本替换掉，并利用替换版本下的密文恢复用户的密钥，而用户无法发现替换后的加密算法与原始加密算法之间的区别。不可检测性是 ASA 的重要目标，要求在用户进行的与应答器（Oracle）交互的可检测性实验中判断应答器是哪种算法的成功优势是足够小的。

一般而言，ASA 攻击依赖于攻击者和替换算法间共享密钥，而且原加密算法是随机的。现有 ASA 方案的替换算法除了有与原加密算法相同的输入，还有由特定密钥生成算法产生的嵌入密钥 ek 和提取密钥 xk ，以及一个状态变量作为输入；输出除了有与原算法相同的输出，还有更新后的状态变量作为输出。攻击者可以利用嵌入密钥和提取密钥以及用户输出的密文获取用户私钥。嵌入密钥 ek 和提取密钥 xk 不同的 ASA 方案称为非对称的 ASA，其中的嵌入密钥 ek 有被公开的可能；嵌入密钥与提取密钥相同的 ASA 方案称为对称的 ASA，这种情况下嵌入密钥认为不会被获取。

在实际的应用场景中，用户可以将 ASA 替换算法中用到的状态重置为任意一个之前出现过的状态。在这样的场景下，根据用户是否可以获取嵌入密钥 ek ，Philip W. Hodges 等人^[78]在 FSE 2022 的工作中定义了两种比前人更强的状态重置应答器用于可检测性实验，规范化地给出用户怎样利用状态重置来检测 ASA，并表明许多已公布的有状态的 ASA 方案是可以被状态重置的简单实际的方法检测到的。例如，对 Ateniese 等人在 CCS 2015 上提出的 ASA 方案，通过简单的状态重置，被攻击用户两次访问应答器得到的签名是相同的，而未受攻击

的用户大概率得到的是不同的签名。作者还给出了两种能抵抗状态重置检测的非对称 ASA 方案：第一种在对方知道嵌入密钥的情况下也无法被检测到；第二种只有在对方不知道嵌入密钥的情况下才无法被检测到，同时要求即使有另一方得到了嵌入密钥，也无法利用 ASA 方案获取有价值的信息。这种放松的要求使得 ASA 的设计有更多的灵活性，得到运行时间更短的 ASA 方案，从而在理论上更难被时间分析检测到。文章最后扩展了这一结果，允许任何满足某些属性的对称 ASA 方案都可以被转化为上述两种类型的非对称 ASA 方案。该工作揭示了多种现有的 ASA 方案是可以在实际的状态重置方法下被检测到的，对未来 ASA 方案的设计和评估工作有借鉴意义。无状态的 ASA 方案天然抵抗这种检测方法，但是否存在无状态的非对称 ASA 方案是一个公开问题。此外，现有的对 ASA 攻击的防护措施都假定了一些额外的受信任的组件，如何简单实现针对 ASA 攻击的防护措施也是未来的研究点。

在 FSE 2022 会议上，刘富康等人^[52]利用 Gimli 的对称性和弱扩散性，通过追踪单个内部状态的对称性和两个不同内部状态的对称性，仅利用两次查询就可以构造 18 轮 Gimli 置换的区分器，然后利用 SP 盒的新性质将该区分器扩展为 Gimli 的全轮区分器，改进了之前的区分攻击结果。另外，该工作进一步利用 Gimli 置换的弱扩散性，结合分治策略来加速约减轮 Gimli-Hash 和 Gimli-XOF-128 的原像攻击的穷举搜索，给出了目前最好的原像攻击结果。

4.7 可证明安全

通用攻击可以为密码结构提供安全下界，因而能够评估安全性证明所提供的安全上界是否是紧致的。在 FSE 2022 (ToSC 2020(4)) 上，Christoph Dobraunig 和 Bart Mennink^[43]提出通用攻击，结合之前以可证明方式得到的安全下界^[44]，评估了 Suffix Keyed 海绵 (Suffix Keyed Sponge, SuKS) 结构的安全界紧致度。顾名思义，SuKS 是一种基于海绵结构的构造，它在吸收数据之后在产生输出之前吸收密钥。将吸收密钥的函数记为 G ，那么函数 G 可以通过易于求逆的操作（如 XOR）或难以求逆的操作（如 PRF）来实现。使用具有难以求逆的方式吸收密钥的 SuKS 在抵抗侧信道攻击方面更有优势，并且这样的结构被 AE 方案 Isap 采用。

Christoph Dobraunig 和 Bart Mennink^[43]将 SuKS 的安全性视为 MAC，分别在函数 G 容易求逆和难以求逆的情况下构造密钥恢复攻击与伪造攻击。与之前针对海绵结构的安全性分析工作一样，多重碰撞在本文的工作中起着核心作用。与之前关于可证明安全性的研究不同的是，此工作推导并使用了在平均情况下找到至少一个多重碰撞的查询复杂度的界。针对以易于求逆方式吸收密钥的 SuKS 结构，作者提出了两种密钥恢复攻击：一种利用了 tag 上的多重碰撞；另一种利用了密钥吸收之前内部状态的多重碰撞。针对以难以求逆方式吸收密钥的 SuKS 结构，本文提出了伪造攻击。攻击中利用了密钥吸收之前内部状态的多重碰撞，较密钥恢复攻击更为复杂。伪造攻击的关键点可以总结如下。

(1) 多重碰撞有助于将问题转化为多目标问题,从而降低了恢复密钥吸收之后的部分状态的复杂度。

(2) 用于找到密钥吸收之前状态内部多重碰撞的明文和前期计算可以被重复利用,用于为多重碰撞中的每个元素找到状态外部的碰撞。

(3) 由于状态外部的碰撞可以被 G 函数保持,因此已经恢复的部分状态可以与在外部和多重碰撞中的元素发生碰撞的内部状态相结合,形成一个在密钥吸收之后的有效状态。在获得了密钥吸收后的有效状态之后就可以伪造有效的 tag 了。

对于这种在状态内部使用多重碰撞的攻击,离线复杂度只是略好于暴力搜索,但在线复杂度可以忽略不计。

这些通用攻击提供的安全下界与 Christoph Dobraunig 和 Bart Mennink 在 ToSC 2019(4)^[44] 中对 SuKS 提出的 PRF 安全界非常匹配,因此表明这些安全界中给出的各项是合理的。

本节作者: 闫海伦(中国科学院大学),胡凯(山东大学),史丹萍(中国科学院信息工程研究所),包珍珍、董晓阳(清华大学),乔珂欣(北京理工大学),郝泳霖(北京信息科学技术研究院),王森鹏(中国人民解放军战略支援部队信息工程大学)

参考文献

- [1] BIHAM E, SHAMIR A. Differential cryptanalysis of DES-like cryptosystems[J]. Journal of CRYPTOLOGY, 1991, 4(1): 3-72.
- [2] WAGNER D. The boomerang attack[C]. Fast Software Encryption: FSE 1999, Rome, Italy, March 1999. Cham: Springer Berlin Heidelberg, 1999: 156-170.
- [3] KELSEY J, KOHNO T, SCHNEIER B. Amplified boomerang attacks against reduced-round MARS and Serpent[C]. Fast Software Encryption: FSE 2000, New York, NY, USA, April 2000. Cham: Springer Berlin Heidelberg, 2000: 75-93.
- [4] BIHAM E, DUNKELMAN O, KELLER N. The rectangle attack-rectangling the Serpent[C]. Cham: Advances in Cryptology-EUROCRYPT 2001: EUROCRYPT 2001, Austria, May 6-10, 2001. Cham: Springer Berlin Heidelberg, 2001: 340-357.
- [5] BIHAM E, DUNKELMAN O, KELLER N. New results on boomerang and rectangle attacks[C]. Fast Software Encryption: FSE 2002, Leuven, Belgium, February 4-6, 2002. Cham: Springer Berlin Heidelberg, 2002: 1-16.
- [6] DONG X Y, QIN L Y, SUN S W, et al. Key guessing strategies for linear key-schedule algorithms in rectangle attacks[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT

2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 3-33.

[7] QIN L Y, DONG X Y, WANG A Y, et al. Mind the TWEAKEY Schedule: Cryptanalysis on SKINNYe-64-256[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 287-317.

[8] SONG L, ZHANG N N, YANG Q Q, et al. Optimizing Rectangle Attacks: A Unified and Generic Framework for Key Recovery[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 410-440.

[9] HADIPOUR H, BAGHERI N, SONG L. Improved rectangle attacks on SKINNY and CRAFT[J]. IACR Transactions on Symmetric Cryptology, 2021(2): 140-198.

[10] LEANDER G, RASOOLZADEH S. Weak Tweak-Keys for the CRAFT Block Cipher[J]. IACR Transactions on Symmetric Cryptology, 2022(1): 38-63.

[11] RAHMAN M, SAHA D, PAUL G. Boomeyong: Embedding Yoyo within Boomerang and its Applications to Key Recovery Attacks on AES and Pholkos[J]. IACR Transactions on Symmetric Cryptology, 2021(3): 137-169.

[12] DINUR I, SHAMIR A. Cube attacks on tweakable black box polynomials[C]. Advances in Cryptology-EUROCRYPT 2009: EUROCRYPT 2009, Cologne, Germany, April 26-30, 2009. Cham: Springer Berlin Heidelberg, 2009: 278-299.

[13] TODO Y. Structural evaluation by generalized integral property[C]. Advances in Cryptology-EUROCRYPT 2015: EUROCRYPT 2015, Sofia, Bulgaria, April 26-30, 2015. Cham: Springer Berlin Heidelberg, 2015: 287-314.

[14] HU K, SUN S W, WANG M Q, et al. An algebraic formulation of the division property: Revisiting degree evaluations, cube attacks, and key-independent sums[C]. Advances in Cryptology-ASIACRYPT 2020: ASIACRYPT 2020, Daejeon, the Republic Korea, December 7-11, 2020. Cham: Springer International Publishing, 2020: 446-476.

[15] HE J H, HU K, PRENEEL B, et al. Stretching Cube Attacks: Improved Methods to Recover Massive Superpolies[C]. Advances in Cryptology - ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 537-566.

[16] ZHANG B, XU C, MEIER W. Fast near collision attack on the Grain v1 stream cipher[C]. Advances in Cryptology-EUROCRYPT 2018: EUROCRYPT 2018, Tel Aviv, Israel, April 29-May 3, 2018. Cham: Springer International Publishing, 2018: 771-802.

- [17] ZHANG B. Cryptanalysis of GSM encryption in 2G/3G networks without rainbow tables[C]. Advances in Cryptology-ASIACRYPT 2019: ASIACRYPT 2019, Kobe, Japan, December 8-12, 2019. Cham: Springer International Publishing, 2019: 428-456.
- [18] DERBEZ P, FOUQUE P A, MOLLIMARD V. Fake near collisions attacks[J]. IACR Transactions on Symmetric Cryptology, 2020(4): 88-103.
- [19] EKDAHL P, JOHANSSON T, MAXIMOV A, et al. A new SNOW stream cipher called SNOW-V[J]. IACR Transactions on Symmetric Cryptology, 2019(3): 1-42.
- [20] GONG X X, ZHANG B. Resistance of SNOW-V against fast correlation attacks[J]. IACR Transactions on Symmetric Cryptology, 2021(1): 378-410.
- [21] JIAO L, LI Y Q, HAO Y L. A guess-and-determine attack on SNOW-V stream cipher[J]. The Computer Journal, 2020, 63(12): 1789-1812.
- [22] YANG J, JOHANSSON T, MAXIMOV A. Improved guess-and-determine and distinguishing attacks on SNOW-V[J]. IACR Transactions on Symmetric Cryptology, 2021(3): 54-83.
- [23] SHI Z, JIN C H, ZHANG J Y, et al. A correlation attack on full SNOW-V and SNOW-Vi[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 34-56.
- [24] GONG X X, ZHANG B. Comparing Large-unit and Bitwise Linear Approximations of SNOW 2.0 and SNOW 3G and Related Attacks[J]. IACR Transactions on Symmetric Cryptology, 2021(2): 71-103.
- [25] AUMASSON J P, MEIER W, MENDEL F. Preimage attacks on 3-pass HAVAL and step-reduced MD5[C]. Selected Areas in Cryptography: SAC 2008, Sackville, New Brunswick, Canada, August 14-15, 2008. Cham: Springer Berlin Heidelberg, 2009: 120-135.
- [26] SASAKI Y, AOKI K. Preimage attacks on 3, 4, and 5-pass HAVAL[C]. Advances in Cryptology-ASIACRYPT 2008: ASIACRYPT 2008, Melbourne, Australia, December 7-11, 2008. Cham: Springer Berlin Heidelberg, 2008: 253-271.
- [27] SASAKI Y, AOKI K. Finding preimages in full MD5 faster than exhaustive search[C]. Advances in Cryptology-EUROCRYPT 2009: EUROCRYPT 2009, Cologne, Germany, April 26-30, 2009. Cham: Springer Berlin Heidelberg, 2009: 134-152.
- [28] AOKI K, SASAKI Y. Preimage attacks on one-block MD4, 63-step MD5 and more[C]. Selected Areas in Cryptography: SAC 2008, Sackville, New Brunswick, Canada, August 14-15, 2008. Cham: Springer Berlin Heidelberg, 2009: 103-119.

[29] GUO J, LING S, RECHBERGER C, et al. Advanced meet-in-the-middle preimage attacks: First results on full Tiger, and improved results on MD4 and SHA-2[C]. Advances in Cryptology-ASIACRYPT 2010: ASIACRYPT 2010, Singapore, December 5-9, 2010. Cham: Springer Berlin Heidelberg, 2010: 56-75.

[30] BAO Z Z, DONG X Y, GUO J, et al. Automatic search of meet-in-the-middle preimage attacks on AES-like hashing[C]. Advances in Cryptology-EUROCRYPT 2021: EUROCRYPT 2021, Zagreb, Croatia, October 17-21, 2021. Cham: Springer International Publishing, 2021: 771-804.

[31] BAO Z Z, GUO J, SHI D P, et al. Superposition meet-in-the-middle attacks: updates on fundamental security of AES-like hashing[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 64-93.

[32] SCHROTTENLOHER A, STEVENS M. Simplified MITM modeling for permutations: New (quantum) attacks[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 717-747.

[33] NIU Z, SUN S W, LIU Y W, et al. Rotational Differential-Linear Distinguishers of ARX Ciphers with Arbitrary Output Linear Masks[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 3-32.

[34] QIN L Y, DONG X Y, WANG X Y, et al. Automated search oriented to key recovery on ciphers with linear key schedule: applications to boomerangs in SKINNY and ForkSkinny[J]. IACR Transactions on Symmetric Cryptology, 2021(2): 249-291.

[35] DERBEZ P, EULER M, FOUQUE P A, et al. Revisiting Related-Key Boomerang Attacks on AES Using Computer-Aided Tool[C]. Advances in Cryptology - ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Spring Nature Switzerland, 2022: 68-88.

[36] NAGELER M, DOBRAUNIG C, EICHLSEDER M. Information-combining differential fault attacks on DEFAULT[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 168-191.

[37] BAKSI A, BHASIN S, BREIER J, et al. DEFAULT: Cipher Level Resistance Against Differential Fault Attack[C]. Advances in Cryptology - ASIACRYPT 2021: ASIACRYPT 2021, Singapore, December 6-10, 2021. Cham: Springer International Publishing, 2021: 124-156.

[38] GOHR A. Improving attacks on round-reduced speck32/64 using deep learning[C]. Advances in Cryptology-CRYPTO 2019: CRYPTO 2019, Santa Barbara, CA, USA, August 18-22, 2019. Cham: Springer International Publishing, 2019: 150-179.

[39] BAO Z Z, GUO J, LIU M C, et al. Enhancing Differential-Neural Cryptanalysis[C]. Advances in Cryptology - ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland. 2022: 318-347.

[40] PEYRIN T, WANG H. The MALICIOUS framework: embedding backdoors into tweakable block ciphers[C]. Advances in Cryptology-CRYPTO 2020: CRYPTO 2020, Santa Barbara, CA, USA, August 17-21, 2020. Springer International Publishing, 2020: 249-278.

[41] BEIERLE C, DERBEZ P, LEANDER G, et al. Cryptanalysis of the GPRS Encryption Algorithms GEA-1 and GEA-2[C]. Advances in Cryptology-EUROCRYPT 2021: EUROCRYPT 2021, Zagreb, Croatia, October 17-21, 2021. Cham: Springer International Publishing, 2021: 155-183.

[42] BEIERLE C, BEYNE T, FELKE P, et al. Constructing and deconstructing intentional weaknesses in symmetric ciphers[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 748-778.

[43] DOBRAUNIG C, MENNINK B. Tightness of the Suffix Keyed Sponge Bound[J]. IACR Transactions on Symmetric Cryptology, 2020(4): 195-212.

[44] DOBRAUNIG C, MENNINK B. Security of the Suffix Keyed Sponge[J]. IACR Transactions on Symmetric Cryptology, 2019(4): 223-248.

[45] HAO Y L, LEANDER G, MEIER W, et al. Modeling for three-subset division property without unknown subset: Improved cube attacks against Trivium and Grain-128aead[C]. Advances in Cryptology-EUROCRYPT 2020: EUROCRYPT 2020, Zagreb, Croatia, May 10-14, 2020. Cham: Springer International Publishing, 2020: 466-495.

[46] HU K, SUN S W, WANG M Q, et al. An algebraic formulation of the division property: Revisiting degree evaluations, cube attacks, and key-independent sums[C]. Advances in Cryptology-ASIACRYPT 2020: ASIACRYPT 2020, Daejeon, the Republic Korea, December 7-11, 2020. Cham: Springer International Publishing, 2020: 446-476.

[47] SUN Y. Automatic search of cubes for attacking stream ciphers[J]. IACR Transactions on Symmetric Cryptology, 2021(4):100-123.

[48] DINUR I, SHAMIR A. Cube attacks on tweakable black box polynomials[C]. Advances in Cryptology-EUROCRYPT 2009: EUROCRYPT 2009, Cologne, Germany, April 26-30, 2009. Cham: Springer Berlin Heidelberg, 2009: 278-299.

- [49] TODO Y. Structural evaluation by generalized integral property[C]. Advances in Cryptology-EUROCRYPT 2015: EUROCRYPT 2015, Sofia, Bulgaria, April 26-30, 2015. Cham: Springer Berlin Heidelberg, 2015: 287-314.
- [50] HU K, SUN S W, WANG M Q, et al. An algebraic formulation of the division property: Revisiting degree evaluations, cube attacks, and key-independent sums[C]. Advances in Cryptology-ASIACRYPT 2020: ASIACRYPT 2020, Daejeon, the Republic Korea, December 7-11, 2020. Cham: Springer International Publishing, 2020: 446-476.
- [51] HE J H, HU K, PRENEEL B, et al. Stretching Cube Attacks: Improved Methods to Recover Massive Superpolies[C]. Advances in Cryptology - ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland. 2022: 537-566.
- [52] LIU F K, ISOBE T, MEIER W. Exploiting Weak Diffusion of Gimli: Improved Distinguishers and Preimage Attacks[J]. IACR Transactions on Symmetric Cryptology, 2021(1): 185-216.
- [53] SUN L, WANG W, WANG M Q. Accelerating the Search of Differential and Linear Characteristics with the SAT Method[J]. IACR Transactions on Symmetric Cryptology, 2021(1): 269-315.
- [54] BONNETAIN X, SCHROTTENLOHER A, SIBLEYRAS F. Beyond quadratic speedups in quantum attacks on symmetric schemes[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 315-344.
- [55] BONNETAIN X, JAUQUES S. Quantum Period Finding against Symmetric Primitives in Practice[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 1-27.
- [56] KAPLAN M, LEURENT G, LEVERRIER A, et al. Breaking symmetric cryptosystems using quantum period finding[C]. Advances in Cryptology-CRYPTO 2016: CRYPTO 2016, Santa Barbara, CA, USA, August 14-18, 2016. Cham: Springer Berlin Heidelberg, 2016: 207-237.
- [57] KUWAKADO H, MORII M. Quantum distinguisher between the 3-round Feistel cipher and the random permutation[C]. 2010 IEEE International Symposium on Information Theory, Austin, TX, USA, June 13-18, 2010. IEEE, 2010: 2682-2685.
- [58] KUWAKADO H, MORII M. Security on the quantum-type Even-Mansour cipher[C]. 2012 International Symposium on Information Theory and its Applications, HI, USA, October 28-31, 2012. IEEE, 2012: 312-316.
- [59] CANALE F, LEANDER G, STENNES L. Simon's algorithm and symmetric crypto: Generalizations and automatized applications[C]. Advances in Cryptology-CRYPTO 2022:

CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 779-808.

[60] GUO J, LIU G Z, SONG L, et al. Exploring SAT for Cryptanalysis: (Quantum) Collision Attacks Against 6-Round SHA-3[C]. Advances in Cryptology - ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Spring Nature Switzerland, 2022: 645-674.

[61] KUMAR C A, KUMAR A, KUMAR SANADHYA S. Quantum Free-Start Collision Attacks on Double Block Length Hashing with Round-Reduced AES-256[J]. IACR Transactions on Symmetric Cryptology, 2021(1): 316-336.

[62] HE L, LIN X E, YU H B. Improved Preimage Attacks on 4-Round Keccak-224/256[J]. IACR Transactions on Symmetric Cryptology, 2021(1): 217-238.

[63] HEIM B R, NOUS C, ROTELLA Y. Algebraic Collision Attacks on Keccak[J]. IACR Transactions on Symmetric Cryptology, 2021(1): 239-268.

[64] LI T, SUN Y. Preimage attacks on round-reduced Keccak-224/256 via an allocating approach[C]. Advances in Cryptology-EUROCRYPT 2019: EUROCRYPT 2019, Darmstadt, Germany, May 19-23, 2019. Cham: Springer International Publishing, 2019: 556-584.

[65] LIN X E, HE L, YU H B. Improved Preimage Attacks on 3-Round Keccak-224/256[J]. IACR Transactions on Symmetric Cryptology, 2021(3): 84-101.

[66] DONG X Y, GUO J, LI S, et al. Triangulating rebound attack on AES-like hashing[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 94-124.

[67] DOBRAUNIG C, EICHLSEDER M, MENDEL F, et al. Ascon v1. 2[J]. Submission to the CAESAR Competition, 2016, 5(6): 7.

[68] DAEMEN J. Permutation-based Encryption, Authentication and Authenticated Encryption. DIAC- Directions in Authenticated Ciphers, July 2012.

[69] LI Y B, ZHANG G Y, WANG W, et al. Cryptanalysis of round-reduced ASCON[J]. Science China Information Sciences, 2017, 60(3): 38102.

[70] LI Z, DONG X Y, WANG X Y. Conditional cube attack on round-reduced ASCON[J]. IACR Transactions on Symmetric Cryptology, 2017(1): 175-202.

[71] ROHIT R, HU K, SARKAR S, et al. Misuse-Free Key-Recovery and Distinguishing Attacks on 7-Round Ascon[J]. IACR Transactions on Symmetric Cryptology, 2021(1): 130-155.

[72] GÉRAULT D, PEYRIN T, TAN Q Q. Exploring Differential-Based Distinguishers and Forgeries for ASCON[J]. IACR Transactions on Symmetric Cryptology, 2021(3): 102-136.

[73] ERLACHER J, MENDEL F, EICHLSEDER M. Bounds for the Security of Ascon against Differential and Linear Cryptanalysis[J]. IACR Transactions on Symmetric Cryptology, 2022(1): 64-87.

[74] ROHIT R, SARKAR S. Diving deep into the weak keys of round reduced ASCON[J]. IACR Transactions on Symmetric Cryptology, 2021(4): 74-99.

[75] NI B Y, DONG X Y, JIA K T, et al. (Quantum) Collision Attacks on Reduced Simpira v2[J]. IACR Transactions on Symmetric Cryptology, 2021(2): 222-248.

[76] BANIK S, BAROOTI K, DURAK F B, et al. Cryptanalysis of LowMC instances using single plaintext/ciphertext pair[J]. IACR Transactions on Symmetric Cryptology, 2020(4): 130-146.

[77] LIU F K, SARKAR S, WANG G L, et al. Algebraic Meet-in-the-Middle Attack on LowMC[C]. Advances in Cryptology- ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 225-255.

[78] HODGES P, STEBILA D. Algorithm Substitution Attacks: State Reset Detection and Asymmetric Modifications[J]. IACR Transactions on Symmetric Cryptology, 2021(2): 389-422.

[79] HUANG Z Y, SUN S W. Synthesizing Quantum Circuits of AES with Lower T-depth and Less Qubits[C]. Advances in Cryptology - ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 614-644.

基础理论

1 单向函数

平均情况下的困难性构造单向函数是一个长期存在的公开问题。一种期望的结果是建立一个理想的双赢局面：要么（对称）密码学无条件存在，要么所有 NP 问题能在平均情况下得到有效解决。在这个看似十分困难的问题缺乏进展的背景下，文献[1]中 Brzuska 等人开始研究以下类型的较弱但有意义的候选结果：要么存在细粒度的单向函数（Fine-Grained One-Way Functions, FGOWF），要么所有 NP 问题在平均情况下效率可以得到次指数加速。FGOWF 只需要在函数的运行时间和求逆的运行时间之间存在一个固定的多项式差（而不是超多项式）。得到一个主要结果：Brzuska 等人表明，如果有一种 NP 语言具有非常强的平均困难度，将其称为块寻找困难度，那么 FGOWF 存在。通过证明这个非常强的平均困难度概念在随机语言中是成立的，来为这个概念提供启发式支持。

2 UC 安全模型

通用可组合性是密码协议模块化设计和分析中广泛使用的一种方法。证明协议在任意（多项式时间）上下文中都是安全的，以允许组合协议和重用安全结果。自从 Canetti 的原始 UC 模型以及 Pfitzmann 和 Waidner 的模型出现以来，已经提出多种不同的通用可组合模型，虽然所有模型遵循相同的基本思想，但在概念上存在巨大差异，这引出了基本和实际问题：一个模型中证明的概念和结果如何与另一个模型关联？不同的模型和其中制定的安全概念是否捕获相同类别的攻击？能否在一个模型中重用另一个模型中证明的结果，如何重用？在文献[2]中，Rausch 等人对此进行了一系列研究。其将 UC 模型嵌入 IITM 模型中，表明所有 UC 协议、安全性和组合结果都可转移到 IITM 模型中。因此，协议设计者可以从 IITM 模型的功能中受益，同时能够使用在 UC 模型中证明的所有结果。

颠覆攻击是通过将一个诚实方的合法实现替换为可以以无法检测的方式泄露信息的实现来破坏密码协议的安全性。所有当前已知的针对颠覆攻击的安全性的加密协议技术都存在一个重要限制，即无法保证协议会话在与其他协议同时运行的现实设置中的安全性。在文献[3]中，Chakraborty 等人在通用可组合框架中提供反向防火墙来解决上述问题。

3 安全归约

随机变量的关联对，或者简称相关性，是信息论密码学的一个核心概念。特别是两方求值函数可被安全归约为从这种相关性中采样，其满足信息论安全性。此外，这种归约为函数定义了重要的密码学复杂度度量，即在函数的安全两方计算协议中需要使用的相关性样本的数量。证明密码学复杂度的下限很困难，谨慎的做法是通过简化问题和模型温和地处理密码学复杂度问题。受此启发，在文献[4]中，Agarwal 等人提出了一种受限形式的安全归约，即安全非交互式归约 SNIR，并为其建立了一些基本结果和相关技术。

4 内存紧归约

内存紧归约是一种尽可能降低对敌手内存使用的策略，它只会产生比原始攻击者更小的内存成本，使我们能对那些受到内存限制的攻击者提出精确的安全声明。尽管这个概念非常重要，但我们目前只知道几种实现内存紧凑的技术，而且之前的研究成果显示，即使是最基本的归约方法，也不能实现内存紧归约。在文献[5]中，Ghoshal 等人介绍了一种新的“内存紧归约”方法，它通过在与攻击者的交互中使用随机字符串来隐藏状态信息，从而把内存成本转移到攻击者身上。

5 隐私电路

隐私电路是抗泄露计算的一个简单抽象。其简单性使得作为理论研究对象极具吸引力，能够对安全和效率之间可实现的权衡进行清晰而严格的分析。但同时，这一简单性使它成为现实的一个非常粗略的近似。令人出乎意料的是，隐私电路可用以保护嵌入式设备免受现实的侧信道攻击。由于在这些设备中生成新的随机性的成本很高，一些工作已开始研究如何使隐私电路的随机性复杂度最小化的问题。目前最好的已知上界为 $O(t^2 \log_2 ts)$ 随机比特，在文献[6]中 Goyal 等人将上界改进为 $O(t \log_2 ts)$ ，包括输入编码器使用的随机性，并将该界扩展到隐私电路的有状态变体上。

6 最小化密码计算开销

Ishai 等人提出了最小化密码学计算开销的问题，其主要结论是：在合理的假设下，大多数数密码原语可以通过恒定的计算开销来实现。然而，这忽略了可能与（具体）计算安全参数 λ 多项式相关的附加项，也就是说，常量开销协议的计算成本可能为 $O(n+\lambda^c)$ ，其中 n 是给定任务的大小，常数 $c>1$ 。因此仅当 $n \gg \lambda$ 时才会出现最佳效率；对于更小的实例，效率可能远非最优。在文献[7]中，Castro 等人研究了对于 n 和 λ 的所有选择，在多对数因子下获得最佳效率的问题，特别是当 $n=\lambda$ 时。并将实现此目标的多对数因子称为渐近似最优（AQO）密码学。

7 理想计算模型

在现代密码学中，随机谕言机模型（Random Oracle Model）和通用群模型（Generic Group Model）是两个重要的理想模型。通用群模型给多个群基的安全假设提供了明确的计算复杂度论证，以及多个安全假设之间的复杂度等价关系。在文献[8]中，Montgomery 和 Zhandry 证明了在量子环境下，离散对数问题和计算 Diffie-Hellman 问题是等价的。

关于通用群模型，现行的有多个变种，其中最知名的两个模型是 Shoup's GGM 和 Maurer's GGM。在文献[9]中，Zhandry 对这两个模型进行了细致的分析，指出了 Shoup's GGM 严格强于 Maurer's GGM，证明了在 Maurer's GGM 中，不存在高效的公钥加密算法、PRP 等。此外，Zhandry 首次给出 Shoup's GGM 与基于身份加密系统的黑盒分离证明。

关于随机模型，文献[10]中 Zhandry 提出了增广随机谕言机模型（Augmented Random Oracle Model），解决了随机模型无法实例化的问题。关于通用群模型的无法实例化问题，Fuchsabuar 等人提出代数群模型（Algebraic Group Model）。不同于通用群模型，在代数群模型中，无法计算困难问题的无条件安全论证，因此需要论证代数群模型下归约算法的有效性。然而，在文献[11]中，Zhang 等人证伪了代数群模型，并给出以下反例：构造不安全的假设 Beg，并证明在代数群模型中 Beg 和离散对数问题是等价的。

8 RSA 密码学分析

部分密钥泄露（Partial Key Exposure, PKE）攻击仅使用 RSA 私钥的部分连续最高有效位（MSB）或最低有效位（LSB）即可恢复整个私钥。由于 PKE 攻击通常依赖于指数的侧信道泄露，在实际应用中 CRT-RSA 常采用各种加盲措施，其中加性盲指数作为“经典指数盲”

法被广泛部署。在这一背景下，如何将 PKE 攻击扩展到具有加性盲指数的 CRT-RSA 上并利用尽可能少的指数连续 MSB/LSB 泄露是一个重要问题。在文献[12]中，Zhou 等人将 PKE 攻击扩展到了具有加性盲指数的 CRT-RSA 上。该扩展允许盲因子达到某种指数大小，且在特定情况下可仅使用盲指数的 $1/3$ MSB 或 LSB 进而恢复整个私钥。一般而言，该扩展 PKE 所适用的公开指数 e 几乎涵盖了实践中最常见的选择。

9 哈希到椭圆曲线点的映射

将任意字符串映射到椭圆曲线上一点（群元素）的哈希函数 H 是许多密码学原语与协议中必不可少的构造。同时，该哈希函数应当“像一个随机谕言机（Random Oracle, RO）一样”，即与 RO 不可区分（Indifferentiable）。因此如何在现实中实例化构造该函数并保证其在恒定时间内执行十分重要。到目前为止，大多数解决方案都存在同样的问题：它们至少要花费两个基域上的求幂开销。在文献[13]中，Jorge 等人针对域 F_q ($q \equiv 1 \pmod{3}$) 且阶为奇数或可被 4 整除) 上的所有曲线提出了新的不可微哈希构造：SwiftEC，其开销仅为基域上的一次求幂运算。

10 NP 论证

考虑以下场景：证明者有一批数量为 m 的 NP 语句 $x_1, \dots, x_m$ ，并试图说服验证者所有语句都是真的。一个朴素的解决方案是证明者向验证者提供 m 个 witness，即 $w_1, \dots, w_m$ ，并让验证者检查每对 (x_i, w_i) 上的 NP 关系。一个自然的问题是：能否更有效地做到这一点，即证明的大小随着实例的数量呈亚线性增长？考虑在公共参考字符串（CRS）模型中为 NP 语言构建非交互式批量论证（BARG），在文献[14]中，Waters 等人采取了一种直接的方法构建出第一个在双线性映射群的标准假设下的 NP BARG，并使用其构造的 BARG 在非对称双线性群中的合理假设下得到了 RAM 程序的非交互式授权方案，这是第一个可公开验证的 RAM 程序的非交互式授权方案。

11 编码

非延展编码（Non-Malleable Codes）可以保护加密工具免受相关密钥攻击（RKA）。在这种攻击中，敌手尝试修改或篡改加密的密钥。非延展编码的独特之处在于，即使攻击者尝试修改编码（或密钥），他们也无法创造出一个有效且与原始编码有关的新编码。然而传统的非延展编码存在一些限制，为此近年来研究人员提出了一种新的编码技术，称为连续非延展编

码 (Continuously Non-Malleable Codes)。目前只有少数几种连续非延展编码的构造方法,而标准的非延展编码已经有很多种,包括对各种篡改方式的防御。在文献[15]中, Brian 等人提供了连续非延展编码的新的构造方法,且可抵抗决策树篡改、有界多项式深度篡改的问题。

12 非交互式模拟

安全多方计算 (MPC) 允许互不信任的各方对其私有数据进行安全计算。MPC 协议通常将大多数加密和计算密集型组件转移到离线程序中。该离线程序的目标是从高度结构化的关联随机性中输出安全的样本。离线程序依靠公钥密码学来实现这一目标,因此其是计算和通信密集型的。提高计算效率的一个自然解决方案是:以非交互的方式将各种生成成本低廉的关联随机性,如来自噪声源的联合样本,转化为对安全计算有用的相关关系,同时产生较低的计算开销。受这一应用的启发,在文献[16]中, Hamidreza 等人引入了安全非交互式模拟 (SNIS) 的概念,并研究了基本的关联随机性之间的 SNIS。

本节作者: 张聪 (浙江大学)

参考文献

- [1] BRZUSKA C, COUTEAU G. On Building Fine-Grained One-Way Functions from Strong Average-Case Hardness[C]. Advances in Cryptology-EUROCRYPT 2022: 41st Annual International Conference on the Theory and Applications of Cryptographic Techniques, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 584-613.
- [2] RAUSCH D, KÜSTERS R, CHEVALIER C. Embedding the UC Model into the IITM Model[C]. Advances in Cryptology-EUROCRYPT 2022: 41st Annual International Conference on the Theory and Applications of Cryptographic Techniques, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 242-272.
- [3] CHAKRABORTY S, MAGRI B, NIELSEN J B, et al. Universally composable subversion-resilient cryptography[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 272-302.
- [4] AGARWAL P, NARAYANAN V, PATHAK S, et al. Secure non-interactive reduction and spectral analysis of correlations[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 797-827.

[5] GHOSHAL A, GHOSAL R, JAEGER J, et al. Hiding in plain sight: memory-tight proofs via randomness programming[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 706-735.

[6] GOYAL V, ISHAI Y, SONG Y. Private circuits with quasilinear randomness[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 192-221.

[7] DE CASTRO L, HAZAY C, ISHAI Y, et al. Asymptotically Quasi-Optimal Cryptography[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 303-334.

[8] MONTGOMERY H, ZHANDRY M. Full quantum equivalence of group action DLog and CDH, and more[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 3-32.

[9] ZHANDRY M. To label, or not to label (in generic groups)[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 66-96.

[10] ZHANDRY M. Augmented random oracles[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 35-65.

[11] ZHANG C, ZHOU H S, KATZ J. An Analysis of the Algebraic Group Model[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 310-322.

[12] ZHOU Y, VAN DE POL J, YU Y, et al. A Third is All You Need: Extended Partial Key Exposure Attack on CRT-RSA with Additive Exponent Blinding[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 508-536.

[13] CHÁVEZ-SAAB J, RODRÍGUEZ-HENRÍQUEZ F, TIBOUCHI M. Swift: Shallue-van de woestijne indifferentiable function to elliptic curves[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 63-92.

[14] WATERS B, WU D J. Batch arguments for NP and more from standard bilinear group assumptions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 433-463.

[15] BRIAN G, FAUST S, MICHELI E, et al. Continuously Non-malleable Codes Against Bounded-Depth Tampering[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 384-413.

[16] AMINI KHORASGANI H, MAJI H K, NGUYEN H H. Secure non-interactive simulation: feasibility and rate[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 767-796.

公钥密码

1 公钥加密

2022 年国际五大密码年会关于公钥加密的相关工作可分为两大类：一类工作提出许多新的加密概念，包括变形加密、不可压缩密码、多指定接收者公钥加密、知识加密；另一类工作则是聚焦于公钥方案设计以提升现有方案的安全模型、效率、安全归约等方面。现概述各工作如下。

在设计用于安全消息传输的公钥加密方案时，通常默认以下两个基本假设：一是接收者隐私假设，即敌手无法获得接收者的私钥；二是发送者自由假设，即发送者可以自由选择想要加密的消息。在现实世界中，若这两个假设不再满足（存在一个能够获得接收者私钥或强迫发送者发送指定内容的独裁者），是否还能利用公钥加密方案实现安全消息传输呢？针对该问题，Persiano 等人^[1]提出了变形加密（Anamorphic Encryption）的概念。变形加密方案使得通信双方能够在上述基本假设不成立的情况下完成隐蔽的消息传输。具体来说，Persiano 等人^[1]针对两个假设不成立的情况分别定义了接收者变形加密和发送者变形加密。

在接收者变形加密方案中，接收者隐私假设不成立。因此，发送者和接收者需要共享一个额外的对独裁者而言不可见的变形密钥。发送者利用变形密钥同时对正常消息和变形消息进行加密得到变形密文，而接收者可以利用变形密钥从变形密文中还原出变形消息。同时，变形密文和正常密文对独裁者而言应当满足不可区分性。Persiano 等人^[1]指出现有的 Noar-Yung 变换以及拒绝采样技术能够用于构造接收者变形加密方案。在发送者变形加密方案中，发送者自由假设不成立。此时，独裁者强制发送者向特定接收者发送特定消息，而发送者希望向额外接收者发送自己想要发送的消息。因此，该方案中存在一个特殊的随机数采样函数。当使用该函数生成的随机数对特定消息进行加密时，额外接收者能够从生成的密文中解密出发送者想要发送的消息。目前，仅有 LWE 和 Dual LWE 加密方案能够用于构造发送者变形加密方案。

安全漏洞无处不在，公钥密码方案中的私钥常常处于被泄露的风险中。为了应对私钥泄露所带来的安全隐患，Guan 等人^[2]提出了不可压缩密码（Incompressible Cryptography）的概念，具体包括不可压缩加密和不可压缩签名。其核心思想是通过增加攻击者存储密文或获取完整签名的成本以减少私钥泄露所造成的明文还原和基于已有签名的新签名伪造。

不可压缩加密中的密文可以具有足够大的长度,使得攻击者不得不花费大量的存储空间来保存密文。同时,若攻击者未能完整保存密文,即便攻击者拿到私钥也不能从中获得任何底层明文相关的信息。对于使用该方案的发送者和接收者而言,他们可以采用类似流密码的方式进行加密和解密,从而避免在极大的空间中对密文进行运算。具体地,Guan 等人^[2]首先基于功能加密和在线提取器构造了一个简单的不可压缩公钥加密方案,然后采用不可区分混淆构造了明密文长度比例接近于 1 的不可压缩公钥加密方案。

不可压缩签名方案中的签名同样可以具有足够大的长度,使得攻击者在未能获得一个完整签名的前提下,无法伪造任何消息的签名。类似地,签名者和验证者可以通过流式传输的方式完成签名的生成与验证。具体地,Guan 等人^[2]给出了一个简单的不可压缩签名方案的构造。该方案采用对流式传输友好的哈希算法对消息求哈希值,然后对哈希值进行签名。因此,整个方案所涉及的运算可以在小空间中进行。在此基础上,Guan 等人^[2]给出了签名和消息长度比例接近于 1 的不可压缩签名方案。

不同的公钥密码应用场景对底层的公钥加密方案提出了不同的功能以及性质需求。例如,在广播传输网络中,发送者希望仅有指定接收者能够使用各自私钥解密得到正确且相同的明文。受广播加密(Broadcast Encryption)方案和多指定验证者签名(Multi-Designated Verifier Signature)方案的启发,Maurer 等人^[3]提出了广播公钥加密(Public Key Encryption for Broadcast)和多指定接收者公钥加密(Multi-Designated Receiver Signed Public Key Encryption)的概念,并给出了具体构造。该工作的核心思路是从广播加密方案出发,设计一个具有更多性质和功能的广播公钥加密方案,然后结合多指定验证者签名方案,构造满足应用需求的多指定接收者公钥加密方案。

在广播公钥加密方案中,接收者使用私钥解密密文除了能获得底层明文,还可获得指定接收者集合。Maurer 等人^[3]提出的具体方案是以 Naor-Yung 方案为基础进行改进的,加强了原方案的安全性,并继承了底层公钥加密方案的匿名性。同时,该方案具有健壮性(非指定接收者的成员无法正确解密)及一致性(指定接收者集合中的接收者解密得到的信息是相同的)。对于多指定接收者公钥加密方案,核心设计思路是使用多指定验证者签名方案对消息进行签名,再通过广播公钥加密方案对消息及其签名加密得到密文。该具体方案继承广播公钥加密方案中的正确性、一致性和匿名性,并具备一些新性质:不可伪造性,即攻击者无法伪造一个来自诚实发送者的密文,以及非公开性,即一个不诚实的接收者无法让第三方相信收到的消息。最后,Maurer 等人^[3]指出上述方案可以用于构造首个基于标准假设的具有标识隐私性的多指定验证者签名方案。

Deng 和 Zhang^[4]首次提出了知识加密(Knowledge Encryption)的概念。具体来说,一个知识加密方案包含了一个 NP 语言集合 L 。一对公私钥的生成需要用到一个语言实例 x 和它的证据 w ,其中公钥需要包含实例 x ,私钥需要包含证据 w 。同时,知识加密方案需要满足 3

个性质：密文不可区分性、证据可提取性和公钥可模拟性。任何基于游戏安全的两轮不经意传输（OT）都能单独用于构造知识加密方案，并且此类不经意传输方案的构造能够基于很多标准困难性假设，包括 DDH、QR 和 LWE 假设。

凭借知识加密，Deng 和 Zhang^[4]给出了首个双边可模拟的三轮不经意传输协议。现有的三轮不经意传输协议要么只能实现针对发送者的单边可模拟性，要么只能实现双边的基于游戏的安全性。除了不经意传输协议，知识加密还可用于实现一系列具有可模拟安全性的密码协议，包括：具有延迟输入的三轮零知识声明；针对输入独立的功能的三轮两方安全计算；BPK 模型下的三轮并发零知识协议；选择公开攻击下安全的两轮承诺协议。与上述协议相关的现有工作要么要求额外假设，要么只能基于大整数分解假设。同时，该工作实现的可模拟性要强于现有的区分者依赖可模拟性。该工作的另一优势是在多项式的困难假设下实现了类多项式时间的可模拟性，而现有工作往往需要类多项式的或指数的困难性假设。

在 2019 年美国国际密码年会上，Ateniese 等人^[5]提出了匹配加密的概念。匹配加密方案要求通信双方在加密通信过程中完成属性与访问策略的匹配判断。只有双方在加解密算法的输入中实现了属性与访问策略的正确匹配，接收者才能正确解密密文。当属性由双方的身份标识组成且访问策略为允许解密的接收者标识时，匹配加密方案又被称为标识匹配加密方案。

考虑到现有的基于标识匹配加密方案，要么在随机谰言机模型下被证明是安全的，要么依赖于 q 型困难性假设，Chen 等人^[6]提出了首个标准模型下基于标准假设的标识匹配加密方案。其核心思想是通过双层标识加密方案在属性上与标识匹配加密的联系，以构造满足需求的匹配加密方案。研究人员在设计匹配加密方案时通常考虑两个基本性质：隐私性和真实性。隐私性要求敌手无法区分在不同的发送者标识和允许解密的接收者标识下加密不同明文产生的密文。上述方案的隐私性依赖于双层标识加密方案中第一层与第二层之间的匿名性。真实性要求敌手无法伪造一个明文的正确密文。上述方案的真实性关联于双层标识加密方案中第二层的不可伪造性。

在公钥加密的传统安全性质定义中，敌手只能获得单个挑战密文。选择开放（Selective-Opening）安全性质则考虑了敌手获得多个挑战密文情况下的安全性，因而相比于传统安全性质要更强。准确地说，在该安全模型中，一个敌手在获得多个挑战密文的同时，能询问其中部分密文的随机数。选择开放安全性确保敌手即使具有这种额外的能力，仍无法获得任何与未询问密文底层明文相关的信息。在形式化描述中，选择开放安全性有两种定义方式：基于不可区分游戏的选择密文攻击安全性（IND-SO-CCA）和基于模拟的选择密文攻击安全性（SIM-SO-CCA），其中满足后者的具体方案更为常见。

Pan 和 Zeng^[7]提出了 4 种具有紧凑公钥和密文的公钥及加密方案。这些方案均满足 SIM-SO-CCA 安全性，具体包括 3 种基于 Diffie-Hellman（DH）假设的直接构造和一种基于 FO 变换的通用构造。前三个具体构造基于不同的 DH 假设的变体（强 DH 假设、Twin DH 假设、

决策性 DH 假设), 并具有小常数因子的安全损失和紧凑的密文、公钥。第四种构造提供了一种通用构造范式, 即对有损加密应用 Fujisaki-Okamoto 转化实现 SIM-SO-CCA 安全的公钥加密方案。在效率方面, 与现有 SIM-SO-CPA/SIM-SO-CCA 安全公钥加密方案相比, 新方案在 NIST 安全标准下的安全强度或公钥和密文尺寸方面都具有一定的优势。然而, 如何在标准模型中构造紧凑且紧密的 SIM-SO-CCA 安全公钥加密方案仍是一个有趣的公开问题。

在公钥加密中, 匿名性要求在多用户场景下, 密文不会泄露底层公钥信息。腐败攻击是多用户场景下常见的安全威胁, 但是现有工作鲜有对腐败攻击下的匿名公钥加密进行研究。为此, Huang 等人^[8]提出了 ANON-RSOk&C 安全性以刻画腐败攻击下的匿名性, 以及 SIM-RSOk&C 安全性以刻画相同类型腐败攻击下的机密性。相应地, Huang 等人提出了一种新的密码原语——密钥和消息非承诺加密 (Key and Message Non-Committing Encryption, KM-NCE), 并使用该原语给出了在同类型腐败攻击下具有匿名性和机密性的公钥加密方案的通用构造框架。另外, 该密码原语的通用构造可以通过密钥可打开的哈希证明系统得到。Huang 等人^[8]使用矩阵 DDH 假设对该哈希证明系统进行了实例化, 由此得到了具体的在标准模型下同时满足上述两种安全性的公钥加密方案。

安全归约旨在利用密码方案的攻击者构造一个用于解决其他计算性困难假设的敌手, 从而证明密码方案的安全性。若攻击者和敌手的运行时间与优势大小相近, 则称该归约为紧 TA 归约。然而, 在 2017 年美密会上, Auerbach 等人^[9]提出归约过程中攻击者以及敌手的内存使用情况也应当被纳入考虑范围。若攻击者和敌手使用的内存大小也相近, 则该归约被称为紧 TAM 归约。Jaeger 和 Kumar^[10]针对多挑战多用户 CCA 安全性给出了首个具有紧 TAM 归约的公钥加密方案。该方案的设计思路是将 KEM-DEM 范式应用于 Hashed-ElGamal 和 Fujisaki-Okamoto 变换的变体。另外, Jaeger 和 Kumar^[10]在证明 Hashed-ElGamal 的 ECIES 拓展版本时采用了新的计算性 Diffie-Hellman 困难性假设。

Fujisaki-Okamoto (FO) 变换作为一种将公钥加密方案的安全性从 IND-CPA 安全提升到 IND-CCA 安全的范式, 在各种密码标准中被广泛应用。具体的转化过程需要对 IND-CPA 安全的加密算法进行去随机化和重加密的操作。因此, 对于加密开销较大的方案而言, FO 变换需要较大的计算开销。

然而, 一些应用场景可能并不需要公钥加密方案具有 IND-CCA 这么强的安全性。例如, TLS 1.3 协议只需要 IND-qCCA 安全性。IND-qCCA 安全性的强度介于 IND-CPA 与 IND-CCA 安全性之间。与 IND-CCA 安全性相比, IND-qCCA 安全游戏中的敌手只能进行 q 次解密询问。针对该安全性, Huguenin-Dumittan 和 Vaudenay^[11]提出了一种将 IND-CPA 安全性提升到 IND-qCCA 安全性的转化范式。与 FO 变换相比, 该转化范式不需要对加密算法进行去随机化和重加密, 因而更加高效。并且, Huguenin-Dumittan 和 Vaudenay^[11]在量子随机谕言机模型

下证明了其安全性。这意味着原本 IND-CPA 安全方案的抗量子攻击性质在经过该转化范式后依然维持。

本节作者：陈荣茂、王毅（国防科技大学）

参考文献

- [1] PERSIANO G, PHAN D H, YUNG M. Anamorphic Encryption: Private Communication Against a Dictator[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 34-63.
- [2] GUAN J, WICHS D, ZHANDRY M. Incompressible Cryptography[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 700-730.
- [3] MAURER U, PORTMANN C, RITO G. Multi-Designated Receiver Signed Public Key Encryption[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 644-673.
- [4] DENG Y, ZHANG X. Knowledge Encryption and Its Applications to Simulatable Protocols with Low Round-Complexity[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 334-362.
- [5] ATENIESE G, FRANCATI D, NUNEZ D, et al. Match Me if You Can: Matchmaking Encryption and Its Applications[C]. Advances in Cryptology-CRYPTO 2019: CRYPTO 2019, Santa Barbara, CA, USA, August 18-22, 2019. Cham: Springer Nature Switzerland, 2019: 701-731.
- [6] CHEN J, LI Y, WEN J, et al. Identity-Based Matchmaking Encryption from Standard Assumptions[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 394-422.
- [7] PAN J, ZENG R. Compact and Tightly Selective-Opening Secure Public-key Encryption Schemes[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 363-393.
- [8] HUANG Z, LAI J, HAN S, et al. Anonymous Public Key Encryption Under Corruptions[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 423-453.
- [9] AUERBACH B, CASH D, FERSCH M, et al. Memory-tight reductions[C]. Advances in Cryptology-CRYPTO 2017: CRYPTO 2017, Santa Barbara, CA, USA, August 20-24, 2017. Cham: Springer Nature Switzerland, 2017: 101-132.

[10] JAEGER J, KUMAR A. Memory-Tight Multi-challenge Security of Public-Key Encryption[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 454-484.

[11] HUGUENIN-DUMITTAN L, VAUDENAY S. On IND-qCCA Security in the ROM and Its Applications[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 613-642.

2 数字签名

Lyubashevsky 签名是基于 Fiat-Shamir 构造范式的一个签名方案,其核心思想是使用拒绝抽样 (Rejection Sampling) 技术,将原本与密钥相关的签名样本转换为来自 (或接近于) 一个独立的 (与密钥无关的) 目标分布的样本。文献[1]通过拒绝抽样技术研究 Lyubashevsky 签名方案,证明了在给定签名算法运行时间的要求下签名压缩的下界,以及完美拒绝抽样策略的预期运行时间。基于雷尼散度 (Rényi Divergence),文献[1]对 Lyubashevsky 签名的安全性重新进行了分析,发现其允许与目标分布有更大的偏差。同时,文献[1]提出超球均匀 (Hyperball Uniforms) 分布是一个很好的分布选择,能够渐近地达到签名的压缩下界,并为实际部署提供了有趣的特性。最后,文献[1]提出了一种新的拒绝抽样策略,其绕过了预期运行时间的下界,并提供了一个最坏情况下 (Worst-Case) 的运行时间保证。

在过去 20 年里,结合部分已知随机值 (Partially Known Nonces) 对 (EC) DSA 和其他相关的签名 (如 Schnorr 签名) 的格归约攻击 (Lattice Reduction Attack),一直是许多具体的密码分析工作的核心。文献[2]发现知道签名密钥的一部分比特信息可以使得格归约攻击更加高效,因此提出了一种改进攻击的新思想:对密钥的某些位进行穷举搜索。这将问题从某个格中的一个有界距离解码 (Bounded Distance Decoding, BDD) 实例转变为一个体积更大但边界相同的固定格 (Fixed Lattice) 中的多个 BDD 实例,由于后者的格是固定的,可以使用 BDD 求解器的批处理/预处理变体 (Batch/Preprocessing Variants),这远比在非预处理格 (Non-Preprocessed Lattices) 上重复格归约高效得多。因此,文献[2]的技术在使用格归约攻击可达到的参数范围内 (160 位组的 2 位泄露或 256 位组的 3 位泄露) 具有竞争力。此外,文献[2]表明这种技术也可以应用于猜测签名随机值 (Signature Nonces) 的某些位或过滤签名数据 (Filtering Signature Data),可得到类似的改进效果。最后,文献[2]在 TPM-FAIL 数据集上进行了实验,并利用改进的攻击方法进行密钥恢复 (Key Recovery)。虽然最初的攻击需要大约 40000 个签名,但通过猜测位 (Guessing Bits) 的方法,只需大约 800 个签名便可恢复密钥,这个改进效果是非常可观的。

白盒密码 (White-Box Cryptography) 是关于寻找密码算法的软件实现, 这些算法嵌入一个密钥, 从而不可能从软件中提取这个密钥。白盒密码是当今一个活跃的研究领域, 实际应用非常广泛, 且不需要特殊的安全硬件 (Secure Hardware)。文献[3]研究了椭圆曲线数字签名算法 (Elliptic Curve Digital Signature Algorithm, ECDSA) 的白盒实现。首先, 文献[3]考虑了打破这种实现的一些攻击途径, 特别提供了各种故障攻击 (Fault Attacks) 的系统概述, ECDSA 白盒实现特别容易受到这些攻击的影响。然后, 文献[3]提出了不同的数学对策 (Mathematical Countermeasures), 主要是对敏感变量 (Sensitive Variables) 进行掩藏/盲化 (Masking/Blinding), 以防止此类攻击 (或至少使此类攻击更加困难)。此外, 文献[3]简要介绍了 ECDSA 白盒场景中一些典型的实现对策及其面临的挑战。文献[3]的工作是由 CHES 2021 邀请赛之一的白盒密码算法设计竞赛 (WhibOx Contest) 发起的, 包括设计和打破白盒 ECDSA 实现。文献[3]通过提交挑战来说明研究的结果和发现, 并提供一个全面的概述, 可以以哪种方式解决哪些挑战。最后, 文献[3]对选定的挑战进行了更加详细的分析。

群签名 (Group Signature) 允许群中的授权成员单独代表群体进行签名, 而签名者的具体身份保持匿名。为了防止匿名性的滥用, 需要一个称为群管理员 (Group Manager) 的特殊实体可以追踪到签名者, 从而使小组成员对各自的签名负责。文献[4]提出了一个基于同源和格假设 (Isogeny and Lattice Assumptions) 的动态群签名 (Dynamic Group Signature), 或者称为可问责环签名 (Accountable Ring Signature), 可通过 CSIDH 群操作 (Group Action) 或基于 MLWE 的群操作对其进行实例化, 签名的大小为 $O(\log_2 N)$, 其中 N 是群中用户的数量。该方案基于一个有效的 OR-证明 (OR-Proof), 增加了一个额外的有效密文验证, 并证明了其非交互式零知识证明系统 (Non-Interactive Zero-Knowledge Proof System) 是可在线提取的 (Online Extractable), 实现了更紧致的安全性归约 (Tighter Security Reduction)。文献[4]提出的动态群签名方案满足 CCA 完全匿名性 (Fully CCA Anonymity)、可追踪性 (Traceability) 以及不可伪造性 (Unforgeability)。最后, 基于同源 (Isogeny-Based) 构造的方案中签名大小比所有已知的后量子群签名小一个数量级。相比之下, 基于格构造的群签名方案仍有一个较大的签名, 但是对于非常大的群来说仍然是比较小的。

ECDSA 签名的两种常见变体是附加密钥派生和预签名 (Additive Key Derivation and Pre-Signatures)。附加密钥派生是一种简单的机制, 用于从单个主密钥派生多个子密钥, 已广泛应用于密码货币当中。由于其线性性质, 附加密钥派生也可以在阈值设置中有效实现。使用预签名, ECDSA 签名算法中的秘密和公共随机值 (Secret and Public Nonces) 可被预先计算。在阈值设置中, 使用预签名和其他预计算的数据可为协议实现一个非常高效的“在线阶段”。最近, 关于这两种变体的研究备受关注, 有时会结合起来考虑。然而, 先前并没有对附加密钥派生进行安全性证明, 更不用说与预签名相结合的附加密钥派生了。

因此,文献[5]为这些变体(包括各自单独的变体和结合的变体)的安全性提供了一个详细的分析,该分析基于通用群模型(Generic Group Model, GGM),且没有以任何方式修改签名方案或削弱安全性的标准概念。文献[5]还提出了一个特定的椭圆曲线通用群模型(Elliptic Curve Generic Group Model, EC-GGM),该模型更好地模拟了 ECDSA 的一些特性,如转换函数和延展性(Conversion Function and Malleability)。此外,文献[5]发现了这些变体方案的一些安全弱点。例如,文献[5]证明了当这两种变体组合时,存在一个对 ECDSA 的立方根攻击,这比对普通 ECDSA 的最著名的平方根攻击(Square-Root Attack)要快得多。最后,文献[5]提出了两个缓解方法:重新随机的预签名(Re-Randomized Presignatures)和齐次密钥派生。这两个缓解措施都是轻量级的,当组合使用时,其安全性本质上与普通 ECDSA 相同。

部分盲签名是普通盲签名的扩展,在电子现金和电子投票中有着广泛的应用。迄今为止最高效的方案之一是 Abe-Okamoto 部分盲签名方案(CRYPTO 2000),其底层核心 OR 证明技术已成为多项研究工作的基础。文献[6]指出了该方案的安全性证明中几个细微的缺陷,重新提供了一个详细且严格的证明,并实现了与原始工作相似的界限。与此同时,该证明策略也可用于其他基于 OR 证明技术的密码方案的安全性分析。

环签名允许用户代表一组临时用户(一个环)签署消息,同时隐藏其真实身份。环签名最初的动机是匿名举报(ASIACRYPT 2001):一名高级政府雇员可以匿名泄露敏感信息,同时证明该信息来自可靠来源,即通过签名泄露。然而,基本上所有已知的环签名方案都要求所有环成员发布与该方案兼容的结构化验证密钥。这产生了一个悖论:如果用户不想被陷害举报,他们将远离支持环签名的方案。因此,文献[7]提出了通用环签名(Universal Ring Signature, URS)的形式化概念:URS 允许签名者针对环成员发布环签名,而与环成员所使用的签名方案无关。特别是,环中的验证密钥都不需要来自同一方案。因此,URS 提供了一种更为有效的举报机制。该工作主要研究了 URS 的可行性,特别是在标准模型中(没有随机预言机或公共参考字符串),且提出了几种可行的 URS 结构,并在所需假设、安全级别和签名大小之间提供了不同的权衡:第一个构造基于标准基元的超多项式困难假设且实现了紧凑签名,即签名大小仅与环的大小和涉及的签名方案的数量对数有关。随后,该工作仅从标准多项式困难假设构建了 URS 的可行性,并基于见证加密和其他标准假设构建了一个非紧凑型 URS 方案。最后,该工作展示了如何通过不可区分性混淆将非紧凑结构调整紧凑结构。

盲签名方案是研究最充分的隐私保护认证工具之一。然而,可证明的安全盲签名的已知构造要么依赖于非标准困难假设,要么参数长度随并行发布的签名数量呈线性增长,或者涉及效率极低的通用技术(如通用安全两方计算)。Katz、Loss 和 Rosenberg(ASIACRYPT 2021)提出了一种新技术:对于安全参数 n ,将 $O(\log_2 n)$ 并发执行安全的盲签名协议转换为对任何 $\text{poly}(n)$ 并发执行安全的盲签名协议。文献[8]消除了该转换的两个缺点:①盲签名协议

的通信复杂性随着签名交互的数量线性增长；②最终方案从底层方案继承了一个非常松散的安全边界，因此需要实际不可行的参数大小。在此之上，在保留原始变换的优点的同时，该工作的通信复杂性只随交互次数呈对数增长。基于 CDH 和 RSA 困难假设，该工作对具体效率进行了改进，并给出了①基于 BLS 的盲签名方案，基于标准大小的组，其中签名大小约为 3KB，每个签名的通信量约为 120KB；②基于 Okamoto-Guillou-Quisquater 方案的盲签名方案，签名的大小和通信量分别约为 9KB 和 8KB。

文献[9]提出了第一个实用的无双线性配对的三步盲签名方案，且满足并发安全、产生短签名（3 个或 4 个组元素/标量），以及基于普通或 one-more 离散对数假设（除了另外假设随机预言机），在通用群模型或代数群模型下该方案可证明安全。该工作还展示了所提出的构造方案的部分盲化版本。该工作所提出的方案不依赖于 ROS（Random inhomogeneities in a Overdetermined Solvable system of linear equations）问题或 mROS（Modified ROS）问题的难度。之前唯一使用这些属性的工作是 Abe 的签名方案（EUROCRYPT 2002），其在密码年会上被 Kastner 等人证明它是安全的（PKC 2022），但它产生的签名大小也是所提签名方案的两倍。该工作的安全性证明的核心是一个新困难问题：加权分数 ROS 问题（Weighted Fractional ROS, WFROS），该工作证明了其（无条件）指数下界。

文献[10]提出了一种新型的门限签名，称为“门限可追责的私用签名”，英文简称 TAPS。该类签名方案支持隐藏签名者的身份信息，同时允许持有验证私钥的用户对签名者追责。TAPS 是介于私用门限签名（PTS）和可追责门限签名（ATS）之间，且可以被描述为需要多名群成员签署同一消息时的群签名。值得一提的是，TAPS 有别于门限群签名，并非对管理员权限的分发，而是如门限签名，对签名者数量设置下界，仅当有足够数量的签名者同时签名才可被验证通过。TAPS 的运行流程可分为三部分，首先生成一个公钥、一个追踪密钥和对应 n 个签名者的 n 个私钥，当满足门限数量下限 t 个以上的签名者对消息进行签名时，生成一个合法签名，最后可凭公钥验证签名的有效性。TAPS 签名具有两大特性，公众无法得知签名者的数量及所需的数量下界 t ，但持有追踪密钥的用户可以追踪到签名者和签名者的数量。从安全性考虑，TAPS 方案应使得一组签名者无法通过欺骗追踪过程来构造出另一组签名者。文献[10]给出了 TAPS 的正式定义与通用构造方案。通过结合通用的公钥加密方案、ATS 签名和非交互零知识证明，能够构造出 TAPS 签名方案。文献[10]中指出 TAPS 通用构造方案的效率较低，并基于 Sigma 协议和 Bulletproofs 给出了两个效率较高的 TAPS 实例方案。

文献[11]首次提出了多模态私有签名的匿名签名原语，允许指定的开放机构了解关于签名者身份的部分信息 op 。在群签名中， op 可以被理解为完整的身份信息；在环签名中， op 则代表没有任何身份信息；而在一般情况下， op 可以被定义为某个泄露部分身份信息的公开函数 $G_j(id)$ 。多模态私有签名是一种高度抽象的签名原语，其涵盖了传统匿名导向的签名原语中跟踪的概念，将支持各种新型的隐私保护相关应用。文献[11]给出了多模态私有签名的形

式化定义和安全性要求,并利用密码学基础组件,提出了一种模块化的通用构造方案,展示了多模态私有签名方案设计的可行性。文中还给出了一种基于群的在标准模型下证明安全的实例化方案和一种基于格的在随机谕言机模型下证明安全的实例化方案。

文献[12]聚焦于以 BLS 和 FROST 签名为代表的非交互式的门限签名,对现有门限签名进行抽象,提出了通用的语法和分层次的安全性模型。文献[12]中利用提出的分层次安全性模型证明了 BLS 签名与 FROST 签名具有比目前被认为的安全性更高的安全性,同时指出 FROST 签名不能满足文中提出的最高等级安全性要求。作者对 FROST 签名进行了细微的调整,提出了新版本的签名方案 FROST2。新提出的 FROST2 方案仅在原有 FROST 方案上做了极少的修改,但具有更高的安全性,满足作者提出的最高等级安全性。

聚合签名能够将 N 个不同消息的 N 个签名压缩成一个短的聚集签名,从而将存储签名的空间复杂度从线性依赖于 N 降低到一个确定的常数。目前存在的问题是验证聚集签名需要访问所有 N 个消息,这将导致验证聚集签名的复杂度至少是 $\Omega(N)$ 。文献[14]提出了支持有效验证、本地可验证的聚集签名,即对 N 个消息构成集合 M 的一个短的聚集签名,验证者能够检查一个特定的消息是否在集合 M 中,其中计算的时间代价并不依赖 N 的大小,并且验证者也并不需要获知整个集合 M 的知识。文献[13]表明了本地可验证的聚集签名可广泛应用于证书透明日志、区块链、编辑签名,甚至单个用户产生的所有签名。文献[13]提出了在随机谕言机模型下两种单签名者、本地可验证聚合签名的构造方法。一种基于 RSA 假设,另一种基于 Bilinear Diffie-Hellman Inversion 假设。此外,文献[13]提出的应用该技术于身份基加密(IBE)中对密钥进行压缩,构造了将 N 个用户身份的私钥压缩到一个聚集密钥的 IBE 方案,其中该聚集密钥能够用于对任意 N 个用户的密文进行解密。

唯一签名是指每条消息只有一个唯一且有效的数字签名。大多数唯一签名的安全归约在基于非交互困难假设下选择消息攻击存在不可伪造(EUF-CMA)的安全模型中存在一个自然的归约损失。Guo 等人于 CRYPTO 2017 提出了一种特定基于链的唯一签名方案,其中唯一签名由 n 个类似区块链顺序计算的 BLS 签名组成。在 Computational Diffie-Hellman 假设下,对于 q_H 个哈希查询,该方案的归约损失是 $n \cdot q_H^{1/n}$ 。当 $n = \log_2 q_H$ 时,该方案达到了对数级别的紧归约。在基于链的唯一签名研究中,存在的问题是还不知道是否存在比对数级别的紧归约更好的归约结果。针对 Guo 等人提出基于链的唯一签名方案,文献[14]表明当唯一签名由 n 个 BLS 签名组成时,对于 q 个签名查询必须有 $q^{1/n}$ 的归约损失,并使用元归约基于任何非交互困难假设证明了在 EUF-CMA 安全模型中的这个下界。文献[14]还提出了基于 CDH 假设、EUF-CMA 安全、仅有 $4 \cdot q^{1/n}$ 安全归约损失的基于链的唯一签名方案,从而极大地改进了以前 $n \cdot q_H^{1/n}$ 最紧对数级别的归约损失。其归约的核心思想是为基于链的唯一签名设计一种独特的非均匀模拟。

Schnorr 签名方案是一种签名长度较短、高效的数字签名方案,其 $4k$ 比特的签名可以保证 k 比特的安全性。在长度为 $p \approx 2^{2k}$ 群上的 Schnorr 签名 σ 由一个元组 (s, e) 组成,其中 $e \in \{0, 1\}^{2k}$, 是哈希输出, $s \in Z_p$, 必须使用密钥计算。虽然哈希输出 e 需要 $2k$ 比特来编码,但 Schnorr 提出在不影响安全性的情况下也许可以截断哈希值。文献[15]在(Shoup's)通用群模型和可编程随机谕言机模型中证明了长度为 $3k$ 比特的短 Schnorr 签名提供了 k 比特的多用户安全性,并进一步分析了能够抵抗预处理攻击、支持公钥作为哈希前缀、短 Schnorr 签名的多用户安全性,获得了长度为 $3k + \log_2 S + \log_2 N$ 比特的签名,其中 N 代表用户数, S 代表预处理攻击者产生提示信息的大小。例如,当 $S = 2^{k/2}$, 可以为 $N \leq 2^{k/4}$ 用户组成的群获得 $3.75k$ 比特的签名。文献[15]提出的技术容易推广至其他基于 Fiat-Shamir 签名的方案,也可应用于 Chaum-Pedersen 签名和 Katz-Wang 签名获得类似结果。文献[15]还将该技术作为底层的构建块分析了在通用群模型中针对有预处理和没有预处理两种情况下的 1-out-of- N 离散对数问题。

本节作者:黄琼、马莎、李宏博、陈新坚(华南农业大学),黄建业(澳大利亚伍伦贡大学)

参考文献

[1] DEVEVEY J, FAWZI O, PASSELÈGUE A, et al. On Rejection Sampling in Lyubashevsky's Signature Scheme[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 34-64.

[2] SUN C, ESPITAU T, TIBOUCHI M, et al. Guessing Bits: Improved lattice attacks on (EC)DSA with nonce leakage[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 391-413.

[3] BAUER S, DREXLER H, GEBHARDT M, et al. Attacks Against White-Box ECDSA and Discussion of Countermeasures: A report on the WhibOx Contest 2021[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 25-55.

[4] BEULLENS W, DOBSON S, KATSUMATA S, et al. Group signatures and more from isogenies and lattices: generic, simple, and efficient[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 95-126.

[5] GROTH J, SHOUP V. On the security of ECDSA with additive key derivation and presignatures[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 365-396.

[6] KASTNER J, LOSS J, XU J. The Abe-Okamoto partially blind signature scheme revisited[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 279-309.

[7] BRANCO P, DÖTTLING N, WOHNIG S. Universal Ring Signatures in the Standard Model[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 249-278.

[8] CHAIRATTANA-APIROM R, HANZLIK L, LOSS J, et al. PI-Cut-Choo and Friends: Compact Blind Signatures via Parallel Instance Cut-and-Choose and More[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 3-31.

[9] TESSARO S, ZHU C. Short Pairing-Free Blind Signatures with Exponential Security[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 782-811.

[10] BONEH D, KOMLO C. Threshold Signatures with Private Accountability[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022:551-581.

[11] NGUYEN K, GUO F, SUSILO W, et al. Multimodal Private Signatures[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022:792-822.

[12] BELLARE M, CRITES E, KOMLO C, et al. Better than Advertised Security for Non-interactive Threshold Signatures[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022:517-550.

[13] GOYAL R, VAIKUNTANATHAN V. Locally Verifiable Signature and Key Aggregation[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 761-791.

[14] GUO F, SUSILO W. Optimal Tightness for Chain-Based Unique Signatures[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 553-583.

[15] BLOCKI J, LEE S. On the multi-user security of short Schnorr signatures with preprocessing[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 614-643.

3 认证/密钥协商/密钥封装/密钥交换

3.1 认证与密钥协商

Transport Layer Security (TLS) 在互联网领域被广泛用来为客户端和服务端建立安全信道。TLS 中的握手协议 (Handshake Protocol) 用于实现客户端和服务器的身份认证、密码组件与参数的协商和建立共享密钥。其中密钥调度算法 (Key Schedule) 是握手协议的重要组成部分, 用于为握手协议的每个阶段贡献相关密钥材料。文献[1]具体分析了 TLS 1.3 握手协议的安全性, 特别是对其中的密钥调度算法进行了深入研究。文献[1]为 TLS 1.3 所用密钥调度算法建立了形式化安全模型。该安全模型能够捕获具备控制客户端和服务端能力的敌手恶意行为 (例如, 敌手可适应性地选择客户端和服务器的密钥材料)。文献[1]使用基于游戏的安全性证明范式证明了密钥调度算法的安全性, 并将密钥调度算法的安全性归约到了伪随机函数和密钥派生函数的安全性。

通信复杂性旨在研究协议的参与者为了执行某些任务需要交换的通信比特数量。现有的通信复杂性计算模型一般要求参与者的输入都是正确的, 且敌手计算能力不受限。文献[2]研究了在计算受限的情况下, 能否通过计算困难假设来打破已知的通信复杂性下界问题。文献[2]讨论了两个通信复杂性计算模型, 即预设公共硬币 (Preset Public Coins) 模型与有状态的自由谈话 (Free Talk) 模型。在第一个模型中, 文献[2]假设存在一个公共随机性来源, 该模型的输入是由一个计算受限的敌对参与者在看到公共随机性后选择的。文献[2]证明了在第一个模型中, 若存在分布式抗碰撞哈希函数 (Distributional Collision-Resistant Hash Function), 则现有通信复杂性计算模型的已知通信下界可以被打破。另外, 文献[2]证明在预设公共硬币模型中不存在常数复杂度的通信协议。在第二个模型中, 文献[2]假设参与者在选择输入之前可以自由交流并保持秘密状态, 通信复杂度仅在交流结束之后才测量。文献[2]证明了在该模型中, 满足通信复杂度下界的相等性谓词协议意味着以建设性的方式达成秘密密钥协商协议。

持续群组密钥协商 (Continuous Group Key Agreement, CGKA) 支持一组群成员在异步网络环境下持续地协商新鲜会话密钥。为同时满足前向安全、盗用后安全等一系列安全性, 现有 CGKA 协议要求群成员频繁地执行密钥更新操作, 导致产生大量协议消息, 引发通信开销过大的问题。例如, CGKA 协议最主流的 TreeKEM 协议在一个群成员执行一次密钥更新时需广播 $\log_2 n$ 条消息, 其中 n 为群成员数。

文献[3]给出了一个称为 CoCoA (Concurrent Continuous Group Key Agreement) 的新的 CGKA 协议构造。该协议允许通信群组高效并发执行会话密钥更新操作。在该协议中, 当一

个群组需要进行 $T(T \leq n)$ 次并发密钥更新时, 在不牺牲安全性的前提下, 最优只需 2 轮通信 (最坏情况下需要 $\log_2 n$ 轮)。在通信开销方面, 当 n 个群成员同时需要执行密钥更新时, 任意群成员仅需处理大约 $(\log_2 n)^2$ 条协议消息。文献[3]将 CoCoA 协议的安全性归约到数字签名的不可伪造性上。与现有的 CGKA 构造相比, 该协议具有最小的通信复杂度。

3.2 密钥封装与密钥包装

在抗量子密码学的背景下, 密码群作用可用于构造基于同源的密码学方案, 如群作用哈希 ElGamal 密钥封装 (Group Action Hashed ElGamal Key Encapsulation Mechanism, GA-HEG KEM) 以及群作用哈希非交互密钥交换 (Group Action Hashed Diffie-Hellman Non-interactive Key Exchange, GA-HDH NIKE)。现有的 GA-HEG KEM 及 GA-HDH NIKE 仅在随机谕言机模型下被证明是被动安全, 其安全性基于群作用计算 Diffie-Hellman (Group Action Computational Diffie-Hellman, GA-CDH) 问题。然而, 目前还未有研究者分析 GA-HEG KEM 和 GA-HEG KEM 的主动安全性, 且如何在量子随机谕言机模型 (Quantum Random Oracle Model, QROM) 下形式化证明 GA-HEG KEM 和 GA-HEG KEM 是主动安全的这一开放问题。

针对上述问题, 文献[4]首先提出群作用强计算 Diffie-Hellman (Group Action Strong CDH, GA-StCDH) 问题的两个变体: 群作用部分量子访问强计算 Diffie-Hellman (Group Action Partial Quantum Access Strong CDH, GA-PQ-StCDH) 问题以及群作用完全量子访问强计算 Diffie-Hellman (Group Action Full Quantum Access Strong CDH, GA-FQ-StCDH) 问题; 其次首次基于 GA-PQ-StCDH 及 GA-FQ-StCDH 问题在 QROM 下形式化证明了 GA-HEG KEM 和 GA-HDH NIKE 这两个方案是主动安全的。

密钥包装 (Key Wrapping, KW) 可看作对称密码体制下的一种密钥封装形式, 即使用单个主包装 (Master Wrapping) 密钥去包装 (加密) 多个数据加密密钥。这些被包装的数据加密密钥可以用于在细粒度级别上保护数据 (如保护单个文件、单条通信消息、单条金融交易条目信息等)。目前, 该机制已经被 NIST 等机构制定了相关标准。

文献[5]将可穿刺技术和密钥包装相结合, 设计了一种被称为可穿刺密钥包装 (Puncturable Key Wrapping, PKW) 的新型密码学原语。该原语包含 4 个算法 $PKW = \{\text{Key Gen}, \text{Wrap}, \text{Unwrap}, \text{Punc}\}$ 。其中, Key Gen 是密钥生成算法, 用于生成 PKW 的主包装密钥; Wrap 是密钥包装算法, 用于使用主包装密钥包装 (加密) 多个数据加密密钥; Unwrap 是密钥解包装算法, 用于使用主包装密钥解包装 (解密) 数据加密密钥; Punc 是穿刺函数, 用于

更新主包装密钥。文献[5]给出了 PKW 的定义和通用构造,并基于可刺穿伪随机函数和认证对称加密给出了一个可证明安全的 PKW 实例协议。

3.3 (隐私保护认证)传统密钥交换

隐私保护认证密钥交换(Privacy-Preserving Authenticated Key Exchange, PPAKE)在保证会话密钥安全的同时实现通信实体身份隐私保护。现有 PPAKE 协议均采用广播信道,需 4 轮交互方可建立会话密钥,所需的通信复杂度与计算复杂度均与广播信道中用户总数线性相关。在文献[6]中,作者将健壮性概念(Robustness)引入 PPAKE,并提出了首个具有健壮性的隐私保护认证密钥交换协议。健壮性使得每个用户在协议运行的第一轮即可确认其是否为目标接收者,从而减少协议的后续交互轮中冗余的通信和计算开销。该协议在保护通信双方身份信息的前提下仅需 3 轮即可建立会话密钥,并且每轮用户广播的消息数量和每个用户的计算复杂度均为常数。协议的构造基于密钥封装机制(Key Encapsulation Mechanism)、数字签名(Digital Signature)、消息认证码(Message Authentication Code)、伪随机生成器(Pseudo-random Generator)和对称加密(Symmetric Encryption)等密码学组件。该 PPAKE 方案在标准模型下实现了会话密钥的前向安全性、显式身份验证、用户身份的前向隐私性(类似于会话密钥前向安全性,即某用户长期私钥泄露不会暴露之前会话中与该用户通信的实体身份信息)这些安全属性。

匿名性是一个抽象的安全目标,在广泛部署的会话协议(Session Protocol, SP)中尤为重要。传统的匿名性意味着会话的参与者不能被识别,即参与者的身份隐私被保护,这种匿名性概念是较弱的,在 SP 的实际部署中是远远不够的。在现实世界中,敌手可以通过故意在会话用户的设备上植入病毒或在会话参与者被扣留时进行物理访问等途径发起攻击,使用户持续的、短暂的暴露他们的秘密值。这类攻击手段对 SP 的匿名性提出了更高的要求,即很难将公开的用户秘密与公开可见的数据关联起来。

棘轮密钥交换(Ratcheted Key Exchange, RKE)是一类 SP 原语,可以在用户秘密值暴露的情况下提供保密性和真实性保证。这是因为 RKE 定期更新用户秘密值,以保障会话密钥盗用后的安全性和前向安全性。文献[7]首次借助 RKE 的优势,在用户秘密值短暂暴露的情况下实现了 SP 的强匿名性保证。文献[7]首先扩展了 RKE 的现有定义,以捕捉将网络上看到的密文与用户设备上暴露的秘密相互关联的攻击;其次,开发了两个新原语:可更新和可随机化的公钥加密(Updatable and Randomizable Public Key Encryption)以及可更新和可随机化的签名(Updatable and Randomizable Signatures);最后,基于这两个新原语构建了强匿名性保护、真实性保护和保密性保护的高效 RKE 方案。

椭圆曲线隐藏数问题（Elliptic Curve Hidden Number Problem, EC-HNP）被 Boneh 等人在 2001 年亚密会上首次提出。该问题被用来严格评估椭圆曲线迪菲赫尔曼密钥交换（Elliptic Curves Diffie-Hellman Key Exchange, ECDH）的比特安全性。该安全性捕获的是敌手恢复会话密钥的有效比特位的能力。目前，EC-HNP 问题也被用于在侧信道攻击下分析 ECDH 会话密钥的安全性。文献[8]回顾了被用于求解 EC-HNP 问题的“Coppersmith”算法，并给出了一个求解 ECDH 密钥比特位数的新界限。具体来说，对于任意的正整数 d 、大素数 p 和一个素数域 $\mathbb{F}_p$ 上的椭圆曲线，如果存在一个预言机输出 ECDH 密钥 X 坐标长度的前 $1/(d+1)$ 个有效位，就可以获得一个启发式的算法能够在 $\log_2 p$ 多项式时间内计算密钥 X 坐标的所有比特。当 d 大于 1 时，界限 $1/(d+1)$ 显著优于已知的严格界限 $5/6$ 和启发式界限 $1/2$ 。由于这种算法涉及的是启发式结果，文献[8]还无法在固定曲线上获得 ECDH 的比特安全性。但是，文献[8]通过实验在低维度格的 NIST 曲线上验证了这个启发式结果的有效性。

3.4 基于口令的认证密钥交换

口令基认证密钥交换（Password Authenticated Key Exchange, PAKE）可允许持有相同（可能是低熵）口令的两方实体建立一个共享会话密钥。非对称口令基认证密钥交换（Asymmetric PAKE, APAKE）作为 PAKE 的扩展，主要针对客户端-服务器场景（客户端持有口令而服务器持有口令的哈希值）建立安全会话密钥。Gu 等人在 2021 年美密会上提出的 APAKE 协议（又称为 KHAPE）被认为是目前计算成本最低的 APAKE。该协议计算成本与不带认证的密钥交换（Un-authenticated Diffie-Hellman Key Exchange, UDH-KE）近乎等价。然而，KHAPE 协议与之前 APAKE 相比并不是轮最优的。文献[9]指出最低计算成本的 APAKE 的实现并非要以牺牲轮代价为前提，并提出两个同时实现最低计算代价和最优轮复杂度的 APAKE 协议：OKAPE 和 AEKE。OKAPE 使用加盐（Salt）的哈希值（ $H(\text{pw}, \text{salt})$ ）在服务器发起会话的情况下仅需 2 轮即建立会话密钥。AEKE 则使用未加盐的哈希值（ $H(\text{pw})$ ）以单向同步流（客户端、服务器同时发送消息）的交互方式完成会话密钥建立过程，具有与 KHAPE 类似的最低计算代价及与 UDH-KE 等价的最优轮复杂度。

现有的强非对称口令基认证密钥交换（Strong APAKE, SAPAKE）加强了 APAKE 对被攻破服务器的安全性保证，但不保护客户端的存储。因此，SAPAKE 只能应用于非对称设置中，无法在对称设置中应用。文献[10]提出了解决这一问题的（强）身份绑定的基于口令的认证密钥交换〔（Strong）Identity-binding PAKE, （S）IPAKE〕概念。它可以防止任意一方被攻破。因此，可以在对称设置中应用。文献[10]定义了 UC 模型下 IPAKE 的理想功能函数，并构造了 CHIP 编译器。该编译器可以将任意对称 PAKE 编译为 IPAKE，将 APAKE 协议的安全性推广到所有参与方。CHIP 在强 Diffie-Hellman 假设下的可编程随机预言机模型中被证明是安

全的。文献[10]也定义了 UC 模型下 SIPAKE 的理想功能函数，并构造了 CRISP 编译器。该编译器可将任意对称 PAKE 编译为 SIPAKE，保护所有参与方口令的安全性。CRISP 在通用群模型和随机谕言机模型中被证明是安全的。

文献[11]中作者利用可交换群作用（Commutative Group Action）构造出两个可证明安全的 PAKE 协议：X-GA-PAKEI 和 Com-GA-PAKEI。在 X-GA-PAKEI 中，参与密钥协商的每个实体需在每轮发送口令后附加集合元素（每个口令比特后附加两个集合元素）。虽然 X-GA-PAKEI 中每轮发送的消息长度较传统 PAKE 协议有所扩展，但该协议可有效抵抗离线字典攻击。此外，X-GA-PAKEI 中参与密钥协商的两个实体仅需一轮交互即可协商出会话密钥。在 Com-GA-PAKEI 中，参与密钥协商的两个实体仅需对每个口令比特附加一个集合元素。为抵抗离线字典攻击，该协议规定其中一方（如服务器）需事先发送对后续发送口令的承诺。Com-GA-PAKEI 需要 3 轮为两个实体协商出会话密钥。为证明协议的安全性，作者在文献[12]中定义群作用的基础上定义出一新的群作用：(R)EGAT [(Restricted) Effective Group Action with Twist] 群作用。X-GA-PAKEI 和 Com-GA-PAKEI 均在该群作用下被证明有效抵抗离线字典攻击，其中 X-GA-PAKEI 协议的安全性归约到 GA-StCDH 问题和 SqInv-GA-StCDH（Square-inverse GA-StCDH）问题，而 Com-GA-PAKEI 协议的安全性归约到 GA-StCDH 问题及 GA-GapCDH（Group Action Gap Computational Diffie-Hellman）问题。

本节作者：张磊、张蕊、张明瑞、徐琳（华东师范大学）

参考文献

- [1] BRZUSKA C, DELIGNAT-LAVAUD A, EGGER C, et al. Key-Schedule Security for the TLS 1.3 Standard[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 621-650.
- [2] COHEN S P, NAOR M. Low Communication Complexity Protocols, Collision Resistant Hash Functions and Secret Key-Agreement Protocols[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 252-281.
- [3] ALWEN J, et al. CoCoA: Concurrent Continuous Group Key Agreement[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 815-844.
- [4] DUMAN J, HARTMANN D, et al. Group Action Key Encapsulation and Non-Interactive Key Exchange in the QROM[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 36-66.

[5] BACKENDAL M, GUNTHER F, PATERSON K G. Puncturable Key Wrapping and Its Applications[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 651-681.

[6] LYU Y, LIU S, HAN S, et al. Privacy-Preserving Authenticated Key Exchange in the Standard Model[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 210-240.

[7] DOWLING B, HAUCK E, RIEPEL D, et al. Strongly Anonymous Ratcheted Key Exchange[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 119-150.

[8] XU J, SARKAR S, WANG H, et al. Improving Bounds on Elliptic Curve Hidden Number Problem for ECDH Key Exchange[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 771-799.

[9] DOS SANTOS B F, GU Y, JARECKI S, et al. Asymmetric PAKE with Low Computation and Communication[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 127-156.

[10] CREMERS C, NAOR M, PAZ S, et al. CHIP and CRISP: Protecting All Parties against Compromise through Identity-Binding PAKEs[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 668-698.

[11] ABDALLA M, EISENHOFER T, KILTZ E, et al. RIEPEL D. Password-Authenticated Key Exchange from Group Actions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 699-728.

[12] ALAMATI N, DE FEO L, MONTGOMERY H, et al. Cryptographic Group Actions and Applications[C]. Advances in Cryptology-ASIACRYPT 2020: ASIACRYPT 2020, Daejeon, the Republic Korea, December 7-11, 2020. Cham: Springer Nature Switzerland, 2020: 411-439.

4 函数加密/属性加密/广播加密

4.1 函数加密

相对于传统公钥体制，函数加密（Functional Encryption，FE）实现了一种细粒度访问控制机制，即当利用解密密钥（与一个函数 f 关联）对密文（与一个明文 x 关联）进行解密时，

最终的解密输出为 $f(x)$ ，且此过程不会泄露除 $f(x)$ 之外的其他信息。因此，函数加密十分适用于隐私数据的分析与计算。当下，针对 FE 的研究主要集中在函数计算模型、安全提升以及衍生扩展上，这些方向在 2022 年的国际三大密码年会上均有对应研究成果发表。在函数计算模型方面，Ananth 等人提出了第一个支持随机访问机（Random Access Machine, RAM）的抗合谋函数加密方案^[1]；Datta 等人提出了一个支持属性权重和（Attribute-Weighted Sum, AWS）运算的非受限函数加密方案^[2]，其权重函数支持 Logspace 图灵机。在安全提升方面，Garg 等人首次提出了动态有界抗合谋函数加密的定义和安全模型^[3]；Kitagawa 和 Nishimaki 提出了函数加密的安全密钥租赁（Secure Key Leasing）概念以解决复制行为带来的隐私泄露问题^[4]。此外，在多用户扩展方面，Nguyen 等人提出了一个安全且更加高效的多用户函数加密方案^[5]，该方案同时实现了基于属性的访问控制和内积函数运算。

在文献[1]中，Ananth 等人提出了第一个公钥环境下支持 RAM 的抗合谋 FE 方案。该方案的安全性依赖支持通用电路的抗合谋 FE 和公钥双重高效私有信息检索（Public-key Doubly-efficient Private Information Retrieval, PK-DEPIR）^[6]。相比于以往的支持通用电路的 FE，支持 RAM 的 FE 方案在密钥生成算法和解密算法上提供了更好的效率保证，即方案参数大小并不会随最坏情况函数计算时间上界增长，且解密时间仅与安全参数和模型计算时间相关。Ananth 等人还设计了一个通用转化框架，可以将任一支持通用电路的（抗合谋）FE 转化为支持 RAM 的（抗合谋）FE。具体来说，他们的技术路线为：①为了实现简洁性，即参数大小和解密时间与最坏情况函数计算时间上界无关，他们构造一个更加简单高效的混淆 RAM 方案^[7]，其通过分别对 RAM 的计算程序和明文数据进行编码，使得泄露的有效信息仅包括函数计算结果，该方案等价于满足一次安全的支持 RAM 的私钥 FE；②结合定时加密方案^[8]和可重用的不经意 RAM^[9]这两种技术实现了可重用的混淆 RAM，可重用性表现在同一明文数据可用于多个不同的混淆程序计算；③类似于文献[10]中的通用转化框架，将支持通用电路的抗合谋 FE 方案转化为支持 RAM 的抗合谋 FE 方案，其中可重用混淆 RAM 被用于明文数据编码和 RAM 程序计算，通用电路 FE 则用于隐藏可重用混淆 RAM 的私钥信息和计算程序编码。至于安全性，根据文献[11-12]中的混合函数加密技术，首先对函数密钥中的混淆计算程序进行硬连线，然后调用混淆 RAM 方案的可重用安全性来证明 FE 方案的不可区分安全性。此外，上述技术同样可以扩展至支持 RAM 的有界抗合谋 FE 构造上，但缺点是方案效率无法媲美支持 RAM 的抗合谋 FE 方案。

文献[2]的研究重点放在了支持 AWS 的函数加密的计算受限问题上。支持 AWS 的函数加密是由 Abdalla 等人在 CRYPTO 2020 首次提出的^[13]，其运算特性在现实中具有广泛的应用。在该方案中，密文与一对属性 (x, z) 绑定（其中 x 为公开属性， z 为私有属性），解密密钥则与一个权重函数 f 绑定，最终的解密结果为 $f(x)z$ ，该解密过程并不会泄露关于 z 的任何信息。Abdalla 等人的方案在标准模型下达到半适应性模拟安全且函数 f 支持算数分支程序

(Arithmetic Branching Programs, ABP) 运算。随后, Datta 等人也提出了一个支持 AWS 运算的函数加密方案^[14], 与前者相比, 该方案的不同之处在于其实现了适应性模拟安全。然而, ABP 是一个非统一计算模型, 且两个方案均要求系统在初始化阶段确定公开属性和私有属性的向量长度, 这些问题也因此限制了该类函数加密方案的实际应用。

文献[2]在考虑并总结了前人工作的基础上, 构造了一个支持 AWS 的 FE 方案, 其同时具备以下优势: ①权重函数 f 支持计算能力更强的 Logspace 图灵机 (统一计算模型); ②方案支持非受限的属性权重和 (Unbounded AWS, UAWS) 运算, 即公开/私有属性支持任意长度, 无须在系统初始化阶段确定; ③密文尺寸是紧凑的, 不会随属性长度的增加而增加; ④依赖标准困难假设 (SXDH), 达到了标准模型下的适应性模拟安全。从技术角度上看, 文献[2]将既有框架^[14-15]拓展到了满足模拟安全和属性隐藏的函数加密 (支持 Logspace 图灵机) 环境下, 依赖一个支持 Logspace 图灵机的算术密钥混淆方案 (Arithmetic Key Garbling Scheme, AKGS) 和一个函数隐藏的多槽内积函数加密作为底层构建块来构建方案。另外, 因为内积函数加密 (Inner Product Functional Encryption, IPFE) 可视为支持 AWS 的 FE 的一个特例, 所以由文献[2]的工作可推导出第一个达到适应性模拟安全的非受限 IPFE 具体实例。

文献[3]分析了现有静态有界抗合谋 FE 的系统漏洞, 并提出了新的解决思路。在模拟安全的有界抗合谋函数加密中, 如果敌手至多问询 $q=q(\lambda)$ 个解密密钥时仍然可以进行安全模拟, 就称该 FE 系统满足 q 有界抗合谋 (q -Bounded Collusion Resistance); 否则, 如果敌手拿到超过 q 个密钥, 系统将被攻破。有界抗合谋 FE 也被称为静态有界抗合谋 FE (Static Collusion Bounded Functional Encryption)。该类方案的最大缺点就是必须在系统建立时确定合谋界限 q , 并且在系统的整个运行过程中, 界限 q 固定不变。这对权威机构来说十分困难, 因为权威机构需要在系统建立前就预判合谋参与方的最大数量。界限值过低易致使系统不安全, 过高则会产生显著的性能开销。一旦合谋界限被打破, 权威机构唯一的选择便是重启系统并重新分发密钥, 这显然十分低效。

有鉴于此, Garg 等人在文献[3]中定义了动态有界抗合谋函数加密 (Dynamic Collusion Bounded Functional Encryption) 系统的语法和安全模型, 该系统在选择合谋边界方面提供了更多的灵活性, 同时兼顾了静态有界抗合谋 FE 系统的优势 (例如, 抗量子、模拟安全性和一般困难假设)。在方案构造上, 他们基于身份基加密 (Identity-Based Encryption, IBE) 获得了一个支持多项式电路的动态有界抗合谋 FE 方案。具体来说, 他们利用 IBE 和混淆电路构造了一个有界抗合谋的带标签 FE, 并将其转化为静态有界抗合谋 FE, 最后通过聚合多个拥有不同合谋界限的静态有界抗合谋 FE 得到最终方案, 使得用户在加密时可以根据不同的合谋界限生成相应的密文。动态有界抗合谋函数加密方案有以下三方面的优势: ①细粒度个性化选择, 用户可灵活选择合谋界限; ②发展型加密方式, 无须在初始化阶段确定合谋界限; ③简单易用的可更新方式, 解密不同界限的密文时无须进行烦琐的更新操作。

在原本的函数加密中,用户一旦获得解密密钥,就可以获得永久解密权限。文献[4]指出,此类问题出现的根本原因在于人们无法阻止私下复制行为的发生。即使建立可撤销机制也无法有效解决该问题,如该机制不适用于密文数量较大的大数据应用场景,且被撤销用户可以通过事先复制“旧”密文来保留解密权限。因此在文献[4]中,基于 Ananth 等人提出的安全软件租赁 (Secure Software Leasing) 概念^[16],Kitagawa 和 Nishimaki 针对私钥 FE 提出了安全密钥租赁 (Secure Key Leasing) 概念,即在满足安全密钥租赁的私钥 FE 中,权威机构仍然会给用户发放密钥,一旦用户向中心返还密钥并且验证通过 (类似于日常租赁),用户就无法再利用该密钥进行解密。在方案构造方面,他们成功将一个支持多项式电路的有界抗合谋私钥 FE 转化为一个满足安全密钥租赁的私钥 FE (这一过程未导入额外假设)。进一步深究之,该方案其实是将解密密钥转化为量子态以阻止用户私自复制。

此外,文献[4]还定义了安全性更强的单一解密者 FE (Single-Decryptor FE, SDFE),其要求每个解密密钥均是防复制的。文献[4]基于不可区分混淆 (indistinguishability Obfuscation, iO) 和 QLWE 假设给出了一个支持多项式电路的抗合谋公钥 SDFE 方案。具体来说,首先利用一个单一解密者加密 (Single-decryptor Encryption, SDE) 方案^[17]将单密钥安全的公钥 FE 方案转化为单密钥安全的 SDFE 方案,接着使用 iO 技术实现抗合谋性质,由此得到结果方案。

除了上述的函数计算类和安全类问题,文献[5]关注的则是 FE 在多用户环境下的衍生扩展问题。面对多用户环境,Goldwasser 等人首次引入了多输入函数加密 (Multi-Input FE, MIFE) 和多用户函数加密 (Multi-Client FE, MCFE) 的概念^[18],二者都是将原有 FE 加密算法的单输入 x 分化成多输入向量 $(x_1, \dots, x_n)$,其中所有分量相互独立,由不同的用户提供。在 MCFE 中,加密算法输入每个用户的索引 i 和标签 tag_i (通常为时间戳) 生成密文: $c_1 = \text{Enc}(1, x_1, \text{tag}_1), \dots, c_n = \text{Enc}(n, x_n, \text{tag}_n)$,解密密钥 dk_f 则与函数 f (该函数包含 n 个输入) 绑定。当使用 dk_f 解密密文 $(c_1, \dots, c_n)$ (所有标签 tag_i 相等) 时,可计算得到 $f(x_1, \dots, x_n)$,但无法泄露单个 x_i 的信息。MIFE 与 MCFE 类似,区别在于前者并不使用标签,密钥解密直接获得 $f(x_1, \dots, x_n)$,所以 MIFE 极易遭受密文重组攻击,导致隐私泄露风险较高。由于访问控制机制可以有效遏制密文中隐私信息的泄露,Abdalla 等人首次提出了带细粒度访问控制的 FE 和 MIFE^[19],同时将标签作为特定的属性嵌入访问策略中,以此自动获得相应的 MCFE 构造。

然而,在文献[5]中,Nguyen 等人认为在考虑某一具体的函数性时,上述方法并不高效。Abdalla 等人在将单用户的 IPFE 扩展到 MIFE 时在效率上做出了较大牺牲,使得密文尺寸由 $O(n)$ 退化为 $O(n^2)$,由此也致使自动获得的 MCFE 方案效率低下^[19]。针对这一问题,Nguyen 等人首先提出了两个支持线性秘密共享方案 (LSSS) 访问控制的单用户 IPFE^[5],分别实现了选择性安全和适应性安全。其中选择性安全版本的效率接近于文献[19],且适应性安全版本的效率几乎是文献[20]的 3 倍。之后他们提出了“复制-提炼” (Duplicate-and-Compress) 技术

以实现单用户 IPFE 到 MCFE 的转化,最终得到的 MCFE 方案达到了随机谕言机模型下的适应性不可区分安全,其中加密过程需要对标签 tag_i 进行哈希映射,而哈希函数在安全证明中被抽象为随机谕言机。该“复制-提炼”技术的优势在于密文尺寸仍保留在 $O(n)$ 级别,极大降低了通信开销。此外,通过为所有密文固定一个公开的标签,他们得到了一个标准模型下更为有效的 MIFE 方案。

4.2 属性加密

属性加密 (Attribute-Based Encryption, ABE) 和谓词加密 (Predicate Encryption, PE) 一直以来都是密码学领域的研究重点,它们的主要特征为:密文与属性 x 、明文 m 绑定,解密密钥与函数 f 绑定,当且仅当 $f(x)=1$ 时才能恢复明文 m 。ABE 和 PE 之间的唯一区别在于后者额外要求属性 x 是不可见的。虽然密码学界已经发表了不少关于这两种原语的研究成果,但这之中适用于多输入情景的方案极为稀少。另外,多源输入聚合功能对于现代密码学应用(如隐私计算等)具有十分重要的现实意义。

在文献[20]中,Agrawal 等人总结了由现有的 MIFE 实现多输入 ABE 或多输入 PE 时所呈现的一些安全和效率问题,如所依赖的困难假设种类受限,最终方案的算法复杂度继承自 MIFE 方案而 MIFE 方案本身的效率极低。因此,Agrawal 等人首次正式定义了私钥环境下的多输入属性基加密 (Multi-Input ABE, MIABE) 和多输入谓词加密 (Multi-Input PE, MIPE) 的相关语法和安全性(标准的适应性不可区分安全)。相比于传统的单输入 ABE 和 PE,解密密钥绑定函数 f ,该函数包含 $n > 1$ 个输入,而系统中 n 个用户分别提供绑定 (x_i, m_i) 的密文,规定当且仅当 $f(x_1, \dots, x_n) = 1$ 时才能恢复所有明文 $(m_1, \dots, m_n)$ 。除此之外,他们还构造了以下一些方案实例。

(1) 两个支持 NC_1 电路的两方 ABE 方案。在双方 ABE 方案中,不同用户的密文互相独立且解密时需要联合这些密文以恢复明文。为此,Agrawal 等人在第一个两方 ABE 方案中要求每个密文的随机值独立选取且在解密时通过双线性配对运算生成共同随机值,并受到文献[21]的单输入 CP-ABE 技术启发,解决了配对操作导致的运算次数受限问题(只能实现线性运算),从而成功解密。与文献[21]类似,该方案的安全性同样依赖 LWE 假设和通用群模型。之后他们参考文献[22]的技术路线,通过一个 IPFE 和一个非标准双线性群假设 KOALA 得到了第二个两方 ABE 方案,其达到了标准模型下的安全性。

(2) 支持多项式电路的启发式两方 ABE 方案。类似地,利用文献[23]的(基于格的)简洁单输入 CP-ABE 构造技术可以得到第一个支持多项式电路的两方 ABE 方案。

(3) 支持 NC_1 电路的启发式三方 ABE 方案。结合文献[23]和第一个支持 NC_1 电路的两方 ABE 方案, Agrawal 等人得到了一个支持 NC_1 电路的三方 ABE 方案, 其安全性依赖格和通用群模型。

(4) 从 MIABE 到 MIPE 的通用转化框架。利用可锁定混淆 (Lockable Obfuscation) [24] 提供的隐藏机制, Agrawal 等人成功实现了单输入 ABE 到单输入 PE 的通用转化并将之扩展至多输入环境下。该框架可以应用至前面提到的两/三方 ABE 方案上, 从而得到对应的两/三方 PE 方案。

上述工作专注于直接构造高效安全的多输入 ABE, 因为多输入 ABE 可用于实现证据加密 (Witness Encryption, WE) [25], 所以该工作也将影响 WE 相关研究。

4.3 广播加密

在广播加密 (Broadcast Encryption, BE) 中, 规定只有在发送者指定集合中的接收者时才能正确解密。本质上, 广播加密可以被看作密文策略属性基加密 (Ciphertext-Policy ABE, CP-ABE) 的一种特殊形式。广播加密自 20 世纪 90 年代以来一直是密码学界活跃的研究领域, 主要目标在于构造具有更短密文、公钥、私钥等参数的 BE 方案。当前学术界的研究聚焦于最佳广播加密 (Optimal BE), 其参数长度都控制在 $\text{poly}(\log_2 N)$ 级别, 但目前已提出的方案都是基于 iO 和多线性映射[26]构造的。同时, 虽然格密码学在过去数十年得到了长足的发展, 但目前并没有基于 LWE 假设且密文和密钥参数长度较短的广播加密方案出现, 这也使得当前的广播加密方案不具有足够的抗量子安全性。

在文献[27]中, WEE 提出了第一个 (可能) 达到后量子安全性的最佳广播加密方案, 其构造基于 Agrawal 等人的方案[28-29]。Agrawal 等人的方案安全性依赖于 LWE 假设, 以及一个双线性群困难问题, 因此其不能在单一格困难问题下证明安全性。而 WEE 提出的广播加密方案仅依赖于 evasive LWE 假设, 该假设是 LWE 假设的一个非标准变体。在技术上, WEE 首先构造一个基于 LWE 假设的单密钥安全的 CP-ABE 方案 (支持通用电路), 将该方案的密钥进行随机化后, 使得该方案具有抗腐化安全性; 接着证明通过选取合适的电路族, 该 CP-ABE 方案可以转化为最佳广播加密。在构造该最佳广播加密方案的同时, WEE 提出了一个支持先验有界多项式深度电路的 CP-ABE 方案, 其密文和密钥长度均与电路深度无关。

本节作者: 陈洁、李宇 (华东师范大学)

参考文献

- [1] ANANTH P, CHUNG K M, FAN X, et al. Collusion-Resistant Functional Encryption for RAMs[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 160-194.
- [2] DATTA P, PAL T, TAKASHIMA K. Compact FE for Unbounded Attribute-Weighted Sums for Logspace from SXDH[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 126-159.
- [3] GARG R, GOYAL R, LU G, et al. Dynamic collusion bounded functional encryption from identity-based encryption[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 736-763.
- [4] KITAGAWA F, NISHIMAKI R. Functional encryption with secure key leasing[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 569-598.
- [5] NGUYEN K, PHAN D H, POINTCHEVAL D. Multi-client functional encryption with fine-grained access control[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 95-125.
- [6] CANETTI R, HOLMGREN J, RICHELSON S. Towards doubly efficient private information retrieval[C]. Theory of Cryptography: TCC 2017, Baltimore, MD, USA, November 12-15, 2017. Springer International Publishing, 2017: 694-726.
- [7] GENTRY C, HALEVI S, LU S, et al. Garbled RAM revisited[C]. Advances in Cryptology-EUROCRYPT 2014: EUROCRYPT 2014, Copenhagen, Denmark, May 11-15, 2014. Cham: Springer Berlin Heidelberg, 2014: 405-422.
- [8] GARG S, OSTROVSKY R, SRINIVASAN A. Adaptive garbled RAM from laconic oblivious transfer[C]. Advances in Cryptology-CRYPTO 2018: CRYPTO 2018, Santa Barbara, CA, USA, August 19-23, 2018. Cham: Springer International Publishing, 2018: 515-544.
- [9] HAMLIN A, HOLMGREN J, WEISS M, et al. On the plausibility of fully homomorphic encryption for RAMs[C]. Advances in Cryptology-CRYPTO 2019: CRYPTO 2019, Santa Barbara, CA, USA, August 18-22, 2019. Cham: Springer International Publishing, 2019: 589-619.
- [10] GARG S, GENTRY C, HALEVI S, et al. Two-round secure MPC from indistinguishability obfuscation[C]. Theory of Cryptography: TCC 2014, San Diego, CA, USA, February 24-26, 2014. Springer Berlin Heidelberg, 2014: 74-94.

[11] DE CARO A, IOVINO V, JAIN A, et al. On the achievability of simulation-based security for functional encryption[C]. Advances in Cryptology-CRYPTO 2013: CRYPTO 2013, Santa Barbara, CA, USA, August 18-22, 2013. Cham: Springer Berlin Heidelberg, 2013: 519-535.

[12] ANANTH P, BRAKERSKI Z, SEGEV G, et al. From selective to adaptive security in functional encryption[C]. Advances in Cryptology-CRYPTO 2015: CRYPTO 2015, Santa Barbara, CA, USA, August 16-20, 2015. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015: 657-677.

[13] ABDALLA M, GONG J, WEE H. Functional encryption for attribute-weighted sums from k-Lin[C]. Advances in Cryptology-CRYPTO 2020: CRYPTO 2020, Santa Barbara, CA, USA, August 17-21, 2020. Cham: Springer International Publishing, 2020: 685-716.

[14] DATTA P, PAL T. (Compact) Adaptively Secure FE for Attribute-Weighted Sums from k-Lin[C]. Advances in Cryptology-ASIACRYPT 2021: ASIACRYPT 2021, Singapore, December 6-10, 2021. Cham: Springer International Publishing, 2021: 434-467.

[15] LIN H, LUO J. Compact Adaptively Secure ABE from k-Lin: Beyond and Towards[C]. Advances in Cryptology-EUROCRYPT 2020: EUROCRYPT 2020, Zagreb, Croatia, May 10-14, 2020. Cham: Springer International Publishing, 2020: 247-277.

[16] ANANTH P, LA PLACA R L. Secure software leasing[C]. Advances in Cryptology-EUROCRYPT 2021: EUROCRYPT 2021, Zagreb, Croatia, October 17-21, 2021. Cham: Springer International Publishing, 2021: 501-530.

[17] COLADANGELO A, LIU J, LIU Q, et al. Hidden cosets and applications to unclonable cryptography[C]. Advances in Cryptology-CRYPTO 2021: CRYPTO 2021, Virtual Event, August 16-20, 2021. Cham: Springer International Publishing, 2021: 556-584.

[18] GOLDWASSER S, GORDON S D, GOYAL V, et al. Multi-input functional encryption[C]. Advances in Cryptology-EUROCRYPT 2014: EUROCRYPT 2014, Copenhagen, Denmark, May 11-15, 2014. Cham: Springer Berlin Heidelberg, 2014: 578-602.

[19] ABDALLA M, CATALANO D, GAY R, et al. Inner-product functional encryption with fine-grained access control[C]. Advances in Cryptology-ASIACRYPT 2020: ASIACRYPT 2020, Daejeon, the Republic Korea, December 7-11, 2020. Cham: Springer International Publishing, 2020: 467-497.

[20] AGRAWAL S, YADAV A, YAMADA S. Multi-input attribute based encryption and predicate encryption[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 590-621.

[21] AGRAWAL S, YAMADA S. Optimal broadcast encryption from pairings and LWE[C]. Advances in Cryptology-EUROCRYPT 2020: EUROCRYPT 2020, Zagreb, Croatia, May 10-14, 2020. Cham: Springer International Publishing, 2020: 13-43.

[22] AGRAWAL S, WICHS D, YAMADA S. Optimal broadcast encryption from LWE and pairings in the standard model[C]. Theory of Cryptography: TCC 2020, Durham, NC, USA, November 16-19, 2020. Springer International Publishing, 2020: 149-178.

[23] BRAKERSKI Z, VAIKUNTANATHAN V. Lattice-inspired broadcast encryption and succinct ciphertext-policy ABE[C]. 13th Innovations in Theoretical Computer Science Conference (ITCS 2022). Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[24] GOYAL R, KOPPULA V, WATERS B. Lockable obfuscation[C]. 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS 2017). IEEE, 2017: 612-621.

[25] BRAKERSKI Z, JAIN A, KOMARGODSKI I, et al. Non-trivial witness encryption and null-iO from standard assumptions[C]. Security and Cryptography for Networks: SCN 2018, Amalfi, Italy, September 5-7, 2018. Cham: Springer International Publishing, 2018: 425-441.

[26] BONEH D, WATERS B, ZHANDRY M. Low overhead broadcast encryption from multilinear maps[C]. Advances in Cryptology-CRYPTO 2014: CRYPTO 2014, Santa Barbara, CA, USA, August 17-21, 2014. Cham: Springer Berlin Heidelberg, 2014: 206-223.

[27] WEE H. Optimal broadcast encryption and CP-ABE from evasive lattice assumptions[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 217-241.

[28] AGRAWAL S, WICHS D, YAMADA S. Optimal broadcast encryption from LWE and pairings in the standard model[C]. Theory of Cryptography: TCC 2020, Durham, NC, USA, November 16-19, 2020. Springer International Publishing, 2020: 149-178.

[29] AGRAWAL S, YAMADA S. Optimal broadcast encryption from pairings and LWE[C]. Advances in Cryptology-EUROCRYPT 2020: EUROCRYPT 2020, Zagreb, Croatia, May 10-14, 2020. Cham: Springer International Publishing, 2020: 13-43.

5 全同态加密

全同态加密 (Fully Homomorphic Encryption, FHE) 支持在无须解密的情况下对加密数据进行任意函数求值, 这一挑战最初由 Rivest 等人在 1978 年提出, 直到 2009 年才在 Gentry 的突破性工作中被解决。经过十余年的深入研究, 一系列实用解决方案已经出现并开始推动标准化。现有 FHE 方案的密文必须包含一定程度的噪声以确保加密的安全性。然而, 同态计

算会导致密文中的噪声增加,并且当噪声溢出时可能会导致解密失败。为了实现真正的全同态加密,Gentry 提出了一种称为自举(Bootstrapping)的特殊降噪操作来刷新密文。

基于 RLWE 的 TFHE 方案是一种支持快速自举的全同态加密方案。文献[1]综述性地介绍了 TFHE 的内部工作原理,描述了在离散化版本的环面(Torus)上实现该方案的过程。接着,该文章详细解释了可编程自举技术,该技术可以在自举的同时计算单变量函数,且不会增加额外开销。

文献[2]基于 TFHE 自举算法,提出了一种高精度同态计算 Sign 函数的方法。首先,使用基于 TFHE 的低精度算法计算 Sign 函数与恒等函数,并由此构造计算 Floor 函数的算法,用于清除被加密明文的一些低有效比特位。然后通过迭代计算 Floor 函数,以高精度进行 Sign 函数的同态计算,并且基于 PALISADE 库实现了该算法,发现它能以至少 21 比特精度计算 Sign 函数,且计算时间仅随 $\log_2 q$ 线性增长(q 是输入密文的模数)。这项技术可以与 CKKS 方案结合使用,相较于目前最优的基于 CKKS 方案的 Sign 函数计算算法,能够以更低的计算代价支持高精度计算。

除常见的 RLWE 等问题之外,NTRU 问题还是构建高效率 FHE 的候选方案之一。然而,近年来,人们发现利用 NTRU 格内稠密子格的存在性可以实施攻击,导致具有大模数的“过度拉伸”的 NTRU 实例易于被攻破。在 2021 年的亚密会上,Ducas 和 van Woerden 给出了一个严格的 NTRU “疲劳点” $q = n^{2.484+O(1)}$,并且发现在大于该“疲劳点”的模数下,稠密子格发现攻击比以往的私钥恢复攻击更加有效。因此,控制 NTRU 型方案的模数小于该“疲劳点”可以有效避免上述子格攻击。

文献[3]遵循 TFHE 的主要框架构建了两个 NTRU 参数在“过度拉伸”区域之外的 FHE 方案。TFHE 自举的核心技术是外部积,允许 RLWE 密文与 RGSW 密文进行同态乘法产生 RLWE 密文。为了将其相应地推广到 NTRU 设置下,文献[3]给出了基于 NTRU 的加密方案 NGS,并以标量形式的 NGS 和向量形式的 NGS 分别取代 RLWE 与 RGSW,从而节省内存需求和运行时间。该工作的第一个方案仅基于 NTRU 假设,表现出与 TFHE 等最先进的 FHE 方案相当的性能;第二个方案同时基于 LWE 和 NTRU 假设,使得自举速度上比 TFHE 快 28%,而且自举和密钥切换时所需的密钥尺寸比 TFHE 小 45%。

CKKS 方案是一种支持复数上近似同态计算的层次型 FHE 方案,每进行一次同态乘法就会消耗一个层级(消耗一定的密文模数比特)。当所有的层级都被用完时,需要使用自举技术来支持进一步的乘法操作。CKKS 自举方法由以下 4 个步骤组成:ModRaise、CoeffToSlot、EvalMod 和 SlotToCoeff。其中,ModRaise 用于放大输入密文模数 q ,但对应明文 m 会变为 $m+qI$ (I 是整系数多项式),因此需要执行 EvalMod 来同态近似模 q 操作将明文恢复成 m 。CoeffToSlot 和 SlotToCoeff 用于明文多项式系数与明文槽之间的映射,包括用缩放因子 Δ 将

DFT/iDFT 复数矩阵扩大并取整,再执行同态乘法,之后按比例缩小结果,它们的层级消耗大约占自举消耗的层级数的一半。

文献[4]提出了一种新的 CKKS 自举方法,在 CcoeffToSlot 中使用较小的缩放因子 $\tilde{\Delta} \ll \Delta$ 减少模数的消耗,再用取整操作 EvalRound 代替 EvalMod 消除小缩放因子造成的误差。该方法的计算成本与传统 CKKS 自举方法相当,同时支持在 CcoeffToSlot 中使用较小的 32 比特缩放因子,并保持自举精度。而且在实际参数中,该工作的模数消耗比传统方法减少了 84 比特,相当于节省了大约两个层级。

CKKS 的解密结果含有噪声,其自举的精度可表示为 $-\log_2 |m-m'|/|m|$,其中 m 与 m' 分别是自举前、后的解密结果。如何提高 CKKS 方案的自举精度也是目前的研究热点之一。文献[5]提出用 Sine 级数逼近模函数,进而提高 CKKS 方案的自举精度。首先,考虑在 q 的倍数附近小区间的模函数,推导逼近模函数的 n 阶 Sine 级数,使该级数有绝对值小于 2 的系数,以及形如 $x+x^{2n+1}p(x)$ 的 Taylor 级数展开 [$p(x)$ 是某个多项式]。接着,通过计算 e^{ix} 的 Taylor 级数再提取虚部,来计算 $\sin x$ 的 Taylor 级数,然后通过同态平方,计算 e^{ikx} 进而计算 $\sin kx$ 。其中,利用交错级数检验与预期的模函数近似误差来确定 $\sin x$ 的 Taylor 级数展开的阶数。该方法能以任意精度逼近模函数,而且基于 HEAAN 库实现了精度达 100 比特的 CKKS 自举,比先前最好方法的精度高 33 比特。

文献[6]则是通过最小化近似误差方差推导出模函数的近似多项式,进而提出了高精度的 CKKS 方案自举算法。该工作使用切比雪夫多项式,分析近似多项式与模函数之间的误差、CKKS 方案带来的误差与解密结果的统计特性,给出了最小化自举误差方差的近似多项式的解析解。在同态计算近似多项式时,利用该解析解的奇偶性,并且通过尽可能地推迟执行重线性化与重缩放,改进了 Baby-Step Giant-Step 算法,降低了自举的计算复杂度与误差。实验表明,该方法仅需 11 层同态乘法(损失 1495 比特密文模数)就可以达到 93 比特自举精度,而先前的自举方法用 11~12 层同态乘法只能达到约 32 比特自举精度,且文献[5]为了达到 100 比特自举精度,需要损失 2157 比特密文模数。

针对 CKKS 方案的安全性, Li 和 Micciancio 在 2021 年的欧密会上指出传统的 IND-CPA 安全不足以充分定义近似同态加密方案在被动敌手下的安全性,同时提出了一种增强的 IND-CPA 安全性定义,称为 IND-CPAD 或带解密谕言机的 IND-CPA,作为近似加密方案的安全性目标。该工作中提出了一些实用的对策以避免这种攻击,随后主要的开源库也都采用了类似的对策来实现 CKKS 方案。然而,这些工作都未能提供一个可证明达到 IND-CPAD 安全的解决方案。

文献[7]展示了如何将任意 IND-CPA 安全的近似同态加密方案转换为可证明 IND-CPAD 安全的方案。该方法使用差分隐私机制来处理解密函数的输出,添加一定量的噪声以匹配同

态计算中最坏情况的误差增长。接着基于静态和动态噪声估计给出了两种近似 FHE 的形式化定义,证明了对静态噪声估计下的 CKKS 方案采用基于高斯噪声的差分隐私机制足以实现 IND-CPAD 安全,并精确量化了所添加噪声的量。而通过动态误差估计并应用上述方法来校准差分隐私噪声,并不能得到安全的方案。此外,文献[7]还提出一种更细粒度的比特安全性定义来区分计算安全性参数和统计安全性参数。

SIMD (Single Instruction Multiple Data) 是基于 RLWE 的 FHE 中常用的打包技术,能够将 $(\mathbb{F}_{t^d})^\ell$ 中的向量打包成 $\mathbb{F}_{t^d}$ 中的元素,从而支持数据的批处理,但要求 t^d 很大。而 RMFE (Reverse Multiplication Friendly Embedding) 技术能够将小域 $\mathbb{F}_t$ 上的 k 维向量空间 $(\mathbb{F}_t)^k$ 嵌入 $\mathbb{F}_w$ (w 与 RMFE 具体采用的嵌入方法有关),使得 $(\mathbb{F}_t)^k$ 中的逐元素乘法等价于 $\mathbb{F}_w$ 中的乘法。

文献[8]通过结合 SIMD 与 RMFE,提出了一种打包更多小域数据的方法——FIMD (Field Instruction Multiple Data)。FIMD 先用 RMFE 将 $k\ell$ 个小域 $\mathbb{F}_t$ 的元素编码成 ℓ 个 $\mathbb{F}_{t^w}$ 的元素,再将它们用 SIMD 打包成 1 个 $\mathbb{F}_{t^d}$ 的元素 ($d \geq w$)。但是先前的 RMFE 要求编码后的数据在每次同态乘法后进行重新编码(同态译码再同态编码的过程),才能继续做同态乘法。对此,文献[8]基于代数函数域改进了 RMFE,使得编码后的数据在重新编码前可以进行多次乘法,从而减少了高代价的重新编码的次数。该工作还通过引入介于 $\mathbb{F}_t$ 与 $\mathbb{F}_{t^d}$ 之间的扩域,优化了重新编码过程。最后基于 HELib 库实现了 FIMD,发现与 SIMD 相比,在 80 比特安全强度下,对于小明文模数 t (如 $t=7$),基于 FIMD 的同态乘法的摊销时间能减少 1/2 以上。

虽然 FHE 具有丰富的计算功能,但是其加密过程与密文大小要比常用的对称密码大得多,而在物联网等应用中客户端的计算与存储能力非常有限。Transciphering 框架支持将对称方案的密文转换成 FHE 密文。利用该框架,客户端仅需用对称加密方案加密明文得到小密文,之后拥有强大算力的服务器可以将该密文转换成 FHE 密文以执行计算任务,从而大幅降低客户端的成本。对于一个对称方案,若其密文能被高效转换成 FHE 密文,则称为同态加密友好的。

文献[9]通过结合传统对称密码与 LWE 加密,提出了一种新的同态加密友好的对称密码——Rubato。Rubato 密码的加密过程包括对私钥与临时随机数 nonce 迭代计算 Rubato 函数、添加高斯噪声以及添加编码后的消息。其中,每轮迭代又包括列混合与行混合、用 Feistel 密码加密以及用可扩展输出函数 (Extendable Output Function) 进行密钥编排。该工作还通过应用对称密码分析与 LWE 密码分析,综合评估了 Rubato 的安全强度,进而提供了推荐参数并且进行了实现。实验结果表明,在 128 比特安全强度下,将 Rubato 密码应用于支持对称密码与 CKKS 方案之间转换的 RtF (Real-to-Finite-Field) Transciphering 框架,能使客户

端与服务器拥有较大的吞吐量（分别为 31.04 MB/s 和 6.712 KB/s），同时保持低密文膨胀率（1.26）与低客户端时延（4.585 μ s）。

文献[10]也设计了一种同态加密友好型对称密码——Elisabeth。首先将定义在 $\mathbb{F}_2$ 上的改进版滤波置换器 IFP 推广到了任意群 $\mathbb{G}$ 上，提出了群滤波置换器（GFP）。其次，用 GFP 设计了流密码 Elisabeth-4（ $\mathbb{G} = \mathbb{Z}_{16}$ ），其同态解密过程非常高效，只需计算 $\mathbb{Z}_{16}$ 上的加法，以及用 TFHE 自举进行两次同态查表。再次，将针对 IFP 的攻击扩展到了群上，讨论了群的使用导致的新攻击，分析了 GFP 和 Elisabeth-4 的安全性。最后，将 Elisabeth-4 与 TFHE 用于 Transciphering 框架，再基于 Concrete 库做了实现，与目前最好的结合 TFHE 的 Transciphering 方案相比，密钥大小缩小了 1/20000~1/200000，效率提升了 5~27.5 倍。而且用该 Transciphering 方案同态计算卷积神经网络以执行图像分类任务，与只用 TFHE 相比，服务器的计算时间虽然稍有增长，但是客户端的带宽消耗节省了 93.62%。

相比于 FHE，仅限于线性变换的协议也非常实用且效率更高。在门限线性同态加密方案中，解密密钥被分布于多个服务器，只有服务器数量达到门限值才能联合执行完整的解密过程。另外，由于现代 CPU 计算在环 $\mathbb{Z}/2^k\mathbb{Z}$ （ k 是某个正整数）中执行，因此在模 2^k 下实现安全多方计算更加简单且允许直接应用 CPU 优化。

为此，文献[11]提出了首个明文空间为 $\mathbb{Z}/2^k\mathbb{Z}$ 的门限线性同态加密方案，其构造参考了 Castagnos 和 Laguillaumie 在 CT-RSA 2015 提出的方案中使用虚二次域类群的做法。后者具备 Elgamal 加密结构，但要求明文空间的阶是已知因式分解的奇素数。文献[11]则使用一个阶为 2^k 的元素加密指数中的消息，解密时在一个未知阶数群中对私钥进行幂运算，同时使用线性整数加密共享方案分发私钥，允许该方案设置任意解密策略的访问结构。该方案的安全性依赖于困难子群成员（Hard Subgroup Membership）假设，该工作可应用于安全多方计算、同态秘密分享和有损陷门函数等领域。

本节作者：王丽萍、徐妍、王庆峰（中国科学院信息工程研究所）

参考文献

- [1] JOYE M. SoK: Fully homomorphic encryption over the [discretized] torus [J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 661-692.
- [2] LIU Z, MICCIANCIO D, POLYAKOV Y. Large-precision homomorphic sign evaluation using FHEW/TFHE bootstrapping[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 130-160.

[3] BONTE C, ILIASHENKO I, PARK J, et al. Final: Faster fhe instantiated with ntru and lwe[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 188-215.

[4] KIM S, PARK M, KIM J, et al. EvalRound Algorithm in CKKS Bootstrapping[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 161-187.

[5] JUTLA C S, MANOHAR N. Sine series approximation of the mod function for bootstrapping of approximate HE[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 491-520.

[6] LEE Y, LEE JW, KIM YS, et al. High-precision bootstrapping for approximate homomorphic encryption by error variance minimization[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 551-580.

[7] LI B, MICCIANCIO D, SCHULTZ M, et al. Securing approximate homomorphic encryption using differential privacy[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 560-589.

[8] AUNG K M M, LIM E, SIM J J, et al. Field Instruction Multiple Data[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 611-641.

[9] HA J, KIM S, LEE B, et al. Rubato: Noisy Ciphers for Approximate Homomorphic Encryption[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 581-610.

[10] COSSERON O, HOFFMANN C, MEAUX P, et al. Towards Case-Optimized Hybrid Homomorphic Encryption: Featuring the Elisabeth Stream Cipher[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 32-67.

[11] CASTAGNOS G, LAGUILLAUMIE F, TUCKER I. Threshold linearly homomorphic encryption on $\mathbb{Z}/2^k\mathbb{Z}$ [C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 99-129.

(后) 量子密码

1 格理论与格密码

数域和函数域有很多相似的地方，数域的一些结果可以类比到函数域，函数域的一些结果也可以类比到数域。但也有例外，如函数域的黎曼猜想（Weil 猜想）已经得到解决，但数域的黎曼猜想仍然遥遥无期。自 2005 年 Regev 提出 LWE 问题并给出 LWE 问题困难性的量子算法归约以来，以 LWE 问题为代表的格密码学已经取得了巨大的成功。在美国 NIST 后量子密码算法征集活动第 3 轮选出的 4 个标准算法中，有 3 个是基于格问题构造的（其中 2 个是 LWE 问题）。此外，基于代数编码的后量子密码算法还在 NIST 第 4 轮评估中。对于基于编码的后量子算法，为了降低密钥规模，人们引入了更多的代数结构。一个长期的问题是，对于这种有较多代数结构的线性码的译码问题，是否存在搜索版问题到判定版问题的归约（Search to Decision Reduction）。在文献[1]中，几位作者定义了函数域上的 LWE 问题，并成功给出了搜索版问题到判定版问题的归约。由于 Carlitz 扩张可以看成是分圆域的类比，文献[1]也考虑了 Carlitz 扩张上的 LWE 问题。与格密码学相比，文献[1]丰富了基于编码的后量子密码学的理论基础。沿着这篇论文的思路，还有很多工作需要进一步开展。

与一般的格相比，理想格（Ideal Lattice）具有更多的代数结构，可以用来设计效率更高的密码算法和协议。另外，理想格的代数结构，有可能会降低这些密码算法和协议的安全性。2021 年，Pan 等学者发现有些素理想具有很好的对称性，导致与这些素理想相关的 SVP 问题存在多项式时间的经典算法。同年，Porter 等学者将这个结果推广到更多的理想。在文献[2]中，Boudgoust 等 3 位作者进一步推广了上述工作，找到了一个简单的条件来判断一个理想是否为密码学上不好的理想。这个条件可以用来判断任意数域中的任意理想（素因子不是分歧的）。相较于上述工作，对于密码学上不好的理想，Boudgoust 等人还设计了通用的密码分析算法，适用于所有数域。此外，文献[2]还将上述结果应用到了 PV-Knap 问题，在多项式时间内解决了 PV-Knap 问题的某些特定实例。于是，这导致与这些实例相关的 PASS Sign 和 PASS Encrypt 两个算法不再安全，它们的密钥可以被恢复出来。

经过数学家和密码学家的长期研究，现在人们普遍相信，一般格上的逼近 SVP 问题（Approx-SVP）是很难的。但是对于理想格上的 SVP 问题（Approx-id-SVP），人们没有足够的信心，因为理想格具有丰富的代数结构。针对这个疑虑，人们最早研究了主理想的情况。

研究思路分为两步：首先计算主理想的生成元；然后通过解决 log-unit 格中的 CVP 问题去计算最短向量。2016 年，对于导子（Conductor）为素数方幂的分圆数域，Cramer 等人证明了主理想上的 Approx-id-SVP 问题存在多项式时间的量子算法，但逼近因子为 $2^{\tilde{O}(\sqrt{n})}$ 。后来，人们沿着两个思路来推广这个结果。第一条思路是 2017 年提出的 CDW 算法及其后续结果：将 Cramer 等人的方法推广到任意分圆域上的任意理想，方法是把任意理想约化到主理想的情况，但逼近因子仍然是 $2^{\tilde{O}(\sqrt{n})}$ ；第二条思路是由 2019 年 Pellet-Mary 等人提出的，通过引入预计算，将整个算法分为预计算和询问两个过程。2020 年，Bernard 和 Roux-Langlois 又提出了 Twisted-PHS 算法。在文献[3]中，通过引入一个独立 S-unit 的满秩家族（Full-Rank Family of Independent S-Units），无论是理论还是实验，Bernard 等人都改进了上述结果。文献[3]去掉了 CWD 算法的量子步骤，而且通过实验发现，Twisted-PHS 算法超过了 CWD 算法的性能。

使用多签名协议，多个签名者可以合作产生某个特定消息的签名。多签名协议可以应用到加密数字货币领域。一个安全的多签名协议需要满足如下 4 个要求：①在 PPK（Plain Public-Key）模型下安全；②并发安全（Concurrent Security）；③低的在线轮复杂度；④密钥聚合（Key Aggregation）。到目前为止，公开文献中已经有 4 个基于格构造的后量子多签名协议，但它们都没能同时满足上述 4 个安全要求。在文献[4]中，Boschini、Takahashi 和 Tibouchi 构造了第一个基于格的后量子多签名协议 MuSig-L，达到了上述 4 个要求。MuSig-L 允许所有签名者在收到需要签名的消息之前离线预计算一些必要的数，这使得在线阶段只需要一轮就行，达到了最优的在线轮数。由于在线阶段只有一轮，因此没法使用承诺方案，这个缺点通过一个广义的拒绝采样方案得到了弥补。MuSig-L 的安全性证明可以在两个标准密码学假设 MSIS 和 MLWE 之下给出。

针对格中计算困难问题，寻找更快的算法是人们孜孜以求的目标。1996 年，Ajtai 将逼近 SVP 问题（Approximate SVP）归约到 SIS 问题。2002 年，通过量子算法，Regev 将逼近 SVP 问题归约到 DCP 问题（Dihedral Coset Problem）。2005 年，同样通过量子算法，Regev 将逼近 SVP 问题归约到 LWE 问题。因此，如果能够找到这 3 个问题的更快算法，我们就找到了逼近 SVP 问题的更快算法。在文献[5]中，几位作者定义了两个新的问题：①创建量子 LWE 状态（C|LWE>）；②给定 LWE 型（LWE-Like）量子状态的情况下，解决 LWE 问题（S|LWE>）。C|LWE>问题已经蕴含在 2009 年 Stehlé、Steinfeld、Tanaka 和 Xagawa 的工作中，S|LWE>问题已经蕴含在 2018 年 Brakerski、Kirshanova、Stehlé 和 Wen 的工作中。在文献[5]中，几位作者设计了一种量子滤波技术（Quantum Filtering Technique）。然后，利用该技术，他们设计了一个量子算法，将 LWE 问题归约到 S|LWE>问题。接着，他们将 SIS 问题（无穷范数情况）和 EDCP 问题（Extrapolated Dihedral Coset Problem）归约到 LWE 问题。最终，

在某些特定的参数情况下,文献[5]构造出了如下 3 个问题的多项式时间量子算法: SIS 问题(无穷范数情况)、LWE 问题、EDCP 问题。需要说明的是,这些参数并没有对应格中问题的最困难情况,因此并没有给出格困难问题的多项式时间量子算法。

内积计算是密码学中的一个常用操作。在文献[6]中,基于 LPN 问题和 LWE 问题两个后量子计算困难性假设, Couteau 和 Zarezhadeh 设计出非交互的安全内积计算协议。假设 Alice 和 Bob 是参与计算的两方, Alice 的秘密向量是 $\mathbf{x}$, Bob 的秘密向量是 $\mathbf{y}$ 。Alice 将 $\mathbf{x}$ 进行编码,然后公开这个编码,并存储一个秘密的状态; Bob 将 $\mathbf{y}$ 进行编码,然后公开这个编码,也存储一个秘密的状态。Alice 获得 Bob 的公开编码,然后利用她自己的秘密状态,计算出 z_1 ; Bob 获得 Alice 的公开编码,然后利用他自己的秘密状态,计算出 z_2 。那么, $z_1 + z_2 = \langle \mathbf{x}, \mathbf{y} \rangle$ 。在文献[6]中,上述协议被称为 NIAIP 协议(Non-Interactive Approximate Inner-Product Protocol)。在半诚实(Semi-Honest)模型下,文献[6]分别基于 LPN 问题和 LWE 问题构造了 NIAIP 协议,并将其应用于两方安全计算内积。由于 NIAIP 协议存在出错的概率,利用预计算,文献[6]又进一步去除了出错的概率。最后,利用零知识证明,文献[6]给出了恶意模型下的协议。文献[6]还提出了一个公开问题:是否存在这样的码,该码和它的对偶码都是可以在线性时间内完成编码的,而且该码和它的对偶码都正好给出计算上困难的 LPN 问题变体(LPN Variant)?

1982 年, Chaum 提出了盲签名的概念。盲签名有很多应用,包括电子现金、电子选举、区块链隐私保护等场景。基于格问题来构造盲签名方案,具有抵抗量子计算攻击的潜力,无疑成为人们关心的一个问题。2010 年,借鉴 Schnorr 盲签名方案的思路, Rükert 构造了第一个基于格的盲签名方案。沿着 Rükert 的想法,人们又给出了很多构造方法,其中 Alkadri 等人于 2020 年构造的 BLAZE+是效率最高的。遗憾的是,这些工作的安全证明存在瑕疵。在文献[7]中, del Pino 和 Katsumata 构造了第一个基于格问题的两轮盲签名方案,签名只有 100KB,而且支持任意多个签名和恶意的签名者。这个方案的安全基础是标准的密码学假设,而且在随机谰言模型(ROM)和量子随机谰言模型(QROM)之下都可以给出安全证明。因为这个方案是两轮的,所以轮数已经达到最优。文献[7]的构造是建立在 2006 年 Fischlin 的工作和 2018 年 del Pino 等人的工作之上的。这个构造需要一种具有特别性质的非交互零知识证明(NIZK),称为多证明直线可提取非交互零知识证明(Multi-Proof Straight-Line Extractable NIZK)。由于文献[7]的构造方法具有一般性,因此改进起来比较方便,即改进其中的某个构件,就可以改进整个方案。

在文献[8]中,基于格同构问题(Lattice Isomorphism Problem, LIP), Ducas 等人提出了 Module LIP 问题。然后,基于 Module LIP 问题,并借鉴 Falcon 和 NTRUSign 两个算法的想法, Ducas 等人设计了一个数字签名算法,称为 Hawk。Module LIP 问题与 LIP 问题的联系,类似 MLWE 问题与 LWE 问题的联系。基于 Module LIP 问题的密码算法,比基于 LIP 问题的密码算法效率更高。Hawk 算法不需要浮点数运算,也不需要拒绝采样技术。Hawk 算法

的签名速度几乎是 Falcon 算法签名速度的 4 倍, 而且签名长度更短, 但公钥长度比 Falcon 算法要长一些。在没有浮点数运算支持的情况下, Hawk 算法的签名速度可以达到 Falcon 算法的 15 倍。此外, 文献[8]还研究了 Module LIP 问题的困难性, 给出了从最困难情况到平均情况的归约。

人们一直在寻找利用格来设计密码算法的新途径。在文献[9]中, Ducas 和 van Woerden 提出了格同构问题 (Lattice Isomorphism Problem, LIP)。两个实数域上的格 L_1 和 L_2 是同构的, 如果存在实数域上的标准正交方阵 O 使得 $L_1 = O \cdot L_2$ 。格同构问题是, 给定同构的两个格 L_1 和 L_2 , 计算出 O 。Ducas 和 van Woerden 研究了 LIP 问题的性质, 并用它来构造密码算法和协议。给定 L_1 的基 B_1 , L_2 的基 B_2 , 存在幺模矩阵 U 使得 $B_1 = O \cdot B_2 \cdot U$ 。由于 U 的存在且不确定, 导致计算 O 是困难的。对于搜索版 LIP 问题和判定版 LIP 问题, Ducas 和 van Woerden 都给出了从最难情况到平均情况的归约。基于 LIP 问题, 文献[9]设计了一个零知识证明协议, 该协议可以用作身份识别方案。此外, 文献[9]还设计了一个密钥封装算法和一个数字签名算法, 该数字签名算法沿用了“哈希然后签名”(Hash-then-Sign)的框架。

Falcon 是一个基于 NTRU 格设计的数字签名算法, 是美国 NIST 公布的 4 个后量子标准密码算法之一。在文献[10]中, Espitau 等人提出了 Mitaka 算法, 改进了 Falcon 算法。2015 年, Prest 在博士论文里提出了一种混杂采样器 (Hybrid Sampler), 应用到秩有限的自由 R 模 (Free R Module of Finite Rank) 时, 性能比离散高斯采样器更好, 其中 R 是 d 次分圆域的代数整数环, 而 d 是 2 的方幂。但是, 直接将这个混杂采样器应用到 Falcon 算法, 会降低安全性。在文献[10]中, Espitau 等人通过设计一种更好的密钥生成算法, 产生质量更好的格陷门, 来弥补这种安全性的损失。此外, 文献[10]还设计了另一种采样器, 可以支持整数运算, 弥补了 Falcon 算法需要浮点运算的缺点。Mitaka 算法还有一个优点: 与 Falcon 算法相比, 抵抗侧信道攻击的能力更强。在 Mitaka 算法签名时, 首先要生成一个随机数, 然后采样的中心是一个由该随机数决定的随机格点, 因此可以设计高阶的掩码来抵抗侧信道攻击。

2008 年, Gentry、Peikert 和 Vaikuntanathan 提出了“哈希并签名”(Hash-and-Sign)的签名方案。Falcon 算法就是沿着这个思路来设计的, 并成为美国 NIST 公布的 4 个后量子标准密码算法之一。在文献[11]中, 为了降低 Falcon 算法的签名长度, Espitau 等人重新研究了“哈希并签名”的签名框架, 得到了很好的结果。文献[11]的作者发现, 在做高斯采样生成签名(s_1, s_2)时, 可以只返回 s_1, s_2 等到验证的时候再计算出来, 这样可以减少签名长度。为了保证新方法的安全性, 新的采样方案是椭球离散高斯采样 (Ellipsoidal Discrete Gaussian), 而不再是球面离散高斯采样 (Spherical Discrete Gaussian)。作者通过引入一个采样的参数来控制新方法的安全水平, 在适当降低安全水平的情况下, 可以明显减少 Falcon 算法的签名长度。此外, 通过安全性分析, 文献[11]的作者又发现, 使用小的模数 q , 如 $q = 257$, 在适当降低安全水

平的情况下,也可以明显减少 Falcon 算法的签名长度。文献[11]中的方法也适用于 Mikata 算法, Mikata 算法是 Falcon 算法的另一种改进,由 Espitau 等人于 2022 年欧密会提出。

基于格的密码算法和协议在后量子密码学领域具有非常强的竞争力。但是,与 RSA 算法和 ECC 算法相比,基于格的方案往往具有过大的公钥长度、签名长度和密文长度。因此,降低格密码方案的公钥长度、签名长度或密文长度,无疑成为人们的一个目标。在文献[12]中,基于 NTRU 格, Fouque 等人构造了一个 CCA 安全的密钥封装方案(KEM)。这个 KEM 方案被取名为 BAT,具有迷人陷门的格基(Basis with Attractive Trapdoor)。与其他格密码方案相比,在相同的安全强度之下, BAT 达到了最小的公钥长度与密文长度之和。在传统的 NTRU 公钥加密算法中,私钥是一个短向量。受到 Falcon 算法和 2001 年 Micciancio 工作的启发,文献[12]改进了这种传统的思路:私钥不再是一个短向量,而是由两个短向量构成的格基和一个额外的元素。在使用私钥来解密时,不再需要专门的模数 p 来计算消息 m ,而是通过两个方程来计算 m 。在计算的过程中,需要用到 Babai 的最近平面算法(Nearest Plane Algorithm)。此外, BAT 不依赖于浮点运算,全部可以采用整数运算。

从计算复杂度理论的角度去研究 NTRU 问题的困难性是一个很有趣的问题,尤其是 NTRU 问题与其他格问题的归约关系,这有助于我们更好地理解 NTRU 问题和相关的格困难问题。在文献[13]中, Felderhoff 等人用 mod-uSVP 表示模中唯一最短向量问题(Module Unique Shortest Vector Problem),并证明了最难情况的 NTRU 问题与最难情况的 mod-uSVP 问题在计算上是等价的。具体来说,文献[13]得到了如下新结果:①最难情况的 mod-uSVP 问题可以归约到最难情况的 NTRU 问题;②最难情况的 mod-uSVPmod 问题可以归约到平均情况的 mod-uSVPmod 问题;③在 id-SVP 预言机的帮助下,平均情况的 mod-uSVPmod 问题可以归约到平均情况的 NTRUmod 问题。结合 2021 年 Pellet-Mary 和 Stehlé 的研究结果和其他一些相关结果,我们就可以得到文献[13]的主要结论。此外,如下两点需要指出:①文献[13]考虑的参数范围与 2013 年 Stehlé 和 Steinfeld 关于 NTRU 可证安全结果考虑的参数范围不同;②文献[13]考虑的参数范围与 Overstretched-NTRU 攻击考虑的参数范围也不同。文献[13]也遗留了一个公开问题:文献[13]考虑的参数范围与 Overstretched-NTRU 攻击考虑的参数范围是否存在交集?

在文献[14]中,针对量子攻击者, Kitagawa 和 Nishimaki 构造了第一个带水印的伪随机函数。文献[14]的贡献包括 3 个方面:一个定义和两个构造。针对伪随机函数, Kitagawa 和 Nishimaki 给出了量子攻击下安全的水印的定义。这个定义借鉴了 2021 年 Goyal 等人的工作。基于 LWE 问题,文献[14]构造了一个可以抵抗量子攻击的带水印的伪随机函数,支持公开添加水印和秘密提取水印。公开添加水印的含义是,水印添加算法只需要输入公开参数;秘密提取水印的含义是,水印提取算法需要输入一个秘密的密钥。基于不可区分混淆(indistinguishability Obfuscation, iO)和 LWE 问题,文献[14]构造了一个可以抵抗量子攻击

的带水印的伪随机函数，支持公开添加水印和公开提取水印。这两种水印都是无法去掉的，哪怕在第一种构造中攻击者可以访问提取水印的谕言机，在第二种构造中攻击者拥有水印提取密钥。除了伪随机函数，文献[14]的方法还可以用到公钥加密算法。

2013 年，Garg 等 4 位学者提出了证据加密（Witness Encryption）的概念。对于证据加密方案，密文是与某个 NP 关系的具体实例绑定的，只有知道这个具体实例对应的证据（Witness），才能成功解密密文。人们一直想从标准的密码学假设来构造证据加密，但直到目前为止，唯一的构造是基于多线性映射的，由 Gentry、Lewko 和 Waters 于 2014 年给出。直接通过不可区分混淆，也可以构造出证据加密。在文献[15]中，通过 LWE 问题，Tsabary 给出了一种新的构造方法。这个构造不是基于原始的 LWE 问题，而是基于与 LWE 问题有关的另一个困难性假设：对于 LWE 问题，给定公开矩阵 A 和一个陷门样本（Trapdoor Sample） $K \leftarrow \text{ATD}(B)$ ，与给定公开矩阵 $[A|B]$ 的计算难度是一样的。Tsabary^[15]用到的构造方法与以前基于 iO 的方法有本质不同，因为使用文献[15]中的方法来构造 iO 是不安全的。Tsabary 认为，证据加密与 iO 之间的区别，类似于 ABE 与功能加密（Functional Encryption, FE）之间的区别。

2013 年，Garg 等人提出了证据加密的概念。到目前为止，几乎所有的证据加密方案都是通过 iO 来构造的，导致这些方案很复杂、很难实用。证据加密与 Null iO 有密切的关系。Null iO 是一类特殊的 iO，是针对零电路（Null Circuit）构造的 iO。因此，在密码学意义下，证据加密看起来比 iO 要弱。2018 年，Chen、Vaikuntanathan 和 Wee 在 GGH15 编码方法的基础上构造了简单的证据加密和 Null iO 方案。在文献[16]中，Vaikuntanathan、Wee 和 Wichs 提出了一个与 LWE 问题相关的计算困难问题：Evasive LWE 问题。然后，基于 Evasive LWE 问题，Vaikuntanathan、Wee 和 Wichs 给出了上述的证据加密和 Null iO 两个方案的安全性证明。该证明的关键一步是要证明 GGH15 编码方法是伪随机的。此外，他们还研究了 Evasive LWE 问题的困难性。

如果攻击者在拥有解密谕言机的情况下，无法从密文中获取所使用公钥的任何信息，则该公钥加密算法是匿名的（Anonymous）。如果只有合法的接收者才能解密密文并获得有意义的明文，则该公钥加密算法是稳健的（Robust）。这两个概念是 CCA 安全之上的概念，对公钥加密算法提出了更高的要求。2010 年，Mohassel 研究了公钥加密方案的匿名性和稳健性，发现即使匿名的 KEM 和 DEM 方案也有可能不匿名的公钥加密方案。2022 年，在量子随机谕言模型之下，Grubbs、Maram 和 Paterson 研究了 NIST 第 3 轮几个 KEM 方案的匿名性与稳健性，并遗留了一个公开问题——NTRU 算法的匿名性和稳健性，Classic McEliece 算法的匿名性。在文献[17]中，Xagawa 全面研究了 NIST 第 3 轮的所有 KEM 方案，回答了上述公开问题。文献[17]的工作建立在 2018 年 Saito 等人的工作之上，通过不相交-模拟能力（Disjoint-Simulatability）、强伪随机性（Strong Pseudorandomness）、光滑性（Smoothness）、稀疏性（Sparseness）等概念来研究匿名性和稳健性。

本节作者：胡红钢（中国科学技术大学）

参考文献

- [1] BOMBAR M, COUVREUR A, DEBRIS-ALAZARD T. On Codes and Learning with Errors over Function Fields[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 513-540.
- [2] BOUDGOUST K, GACHON E, PELLET-MARY A. Some easy instances of ideal-SVP and implications on the partial Vandermonde Knapsack problem[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 480-509.
- [3] BERNARD O, LESAVOUREY A, NGUYEN TH, et al. Log-S-unit lattices using explicit Stickelberger generators to solve approx ideal-SVP[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 677-708.
- [4] BOSCHINI C, TAKAHASHI A, TIBOUCHI M. MuSig-L: Lattice-Based Multi-Signature With Single-Round Online Phase[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 276-305.
- [5] CHEN Y, LIU Q, ZHANDRY M. Quantum Algorithms for Variants of Average-Case Lattice Problems via Filtering[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 372-401.
- [6] COUTEAU G, ZAREZADEH M. Non-Interactive Secure Computation of Inner-Product from LPN and LWE[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 474-503.
- [7] DEL PINO R, KATSUMATA S. A New Framework for More Efficient Round-Optimal Lattice-Based (Partially) Blind Signature via Trapdoor Sampling[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 306-336.
- [8] DUCAS L, POSTLETHWAITE E W, PULLES L N, et al. Hawk: Module LIP makes lattice signatures fast, compact and simple[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 65-94.

- [9] DUCAS L, VAN WOERDEN W. On the lattice isomorphism problem, quadratic forms, remarkable lattices, and cryptography[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 643-673.
- [10] ESPITAU T, FOUQUE P A, GERARD F, et al. Mitaka: A Simpler, Parallelizable, Maskable Variant of Falcon[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 222-253.
- [11] ESPITAU T, TIBOUCHI M, WALLET A, et al. Shorter Hash-and-Sign Lattice-Based Signatures[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 245-275.
- [12] FOUQUE PA, KIRCHNER P, PORNIN T, et al. Bat: small and fast KEM over NTRU lattices[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 240-265.
- [13] FELDERHOFF J, PELLET-MARY A, STEHLE D. On Module Unique-SVP and NTRU[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 709-740.
- [14] KITAGAWA F, NISHIMAKI R. Watermarking PRFs Against Quantum Adversaries[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 488-518.
- [15] TSABARY R. Candidate Witness Encryption from Lattice Techniques[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 535-559.
- [16] VAIKUNTANATHAN V, WEE H, WICHES D. Witness Encryption and Null-IO from Evasive LWE[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 195-221.
- [17] XAGAWA K. Anonymity of NIST PQC Round 3 KEMs[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 551-581.

2 同源密码

相较于格密码等其他几种抗量子密码体系，基于同源的抗量子密码体系发展较晚，但由于其具有最为紧致的密钥及密文大小，在最近 10 年发展十分迅速。2022 年亚密会及欧密会上同源方面的文章共 4 篇。具体成果介绍如下。

SIDH (Supersingular Isogeny Diffie-Hellman) 是基于同源的密钥交换协议，SIDH 的安全性基于同源映射问题的难度，该问题涉及计算两个给定椭圆曲线之间的同源映射。由于在同源计算上，SIDH 引入了椭圆曲线点作为辅助信息，致使 SIDH 的安全性一直存在较大争议，也导致了許多针对该辅助信息的攻击。Galbraith 等人提出的自适应攻击表明，在多个交换中重复使用静态 SIDH 密钥会导致信息泄露。SIKE (Supersingular Isogeny Key Encapsulation) 采用使用一次性临时密钥的方式，并结合 FO 转换设计了安全的密钥封装机制，但该方案在双方都期望使用长期密钥的情况下仍不实际。Azarderakhash 等人提出采用并行 SIDH 的方式，该方案被称为 k-SIDH，但 SIDH 实例增长与 k^2 呈正相关，该方案的效率十分低下。

应对自适应攻击的另一种方法是设计安全的非交互式零知识证明 (Non-Interactive Zero-Knowledge Proof, NIZK)，De Feo 等人在 De Feo-Jao-Plût 方案中提出了一种基于 SIDH 的 NIZK，但文献[1]指出该协议中使用了一种无用的假设，并提供了该假设的反例，从而证明该协议可靠性的证明是无效的。文献[1]提出了两种新的 sigma 协议：其一用于证明在超奇异同源图中特定长度的同源；其二则针对 SIDH 中额外扭曲点映射问题，并提供了新的 NIZK 用于验证 SIDH 的公钥以防止自适应攻击。

同源的表示以及有效计算是同源密码学中一个基础的、重要的问题。传统标准同源的计算常用 Vélú 范式，该范式提供了一种基于核计算同源的方法。截至目前，最有效的算法需要

$\tilde{O}(\sqrt{D'})$ 的操作数，其中 D' 为所计算同源度数的最大因数。该方法仅在同源度数光滑下是有效的。最近对 Deuring Correspondence 及其应用的研究提供了一种有效表示和计算任意度数同源的方法。Kohel 等人了解决四元数代数上的同源路径问题建设性地提出了 KLPT 算法，Eisenträger 等人基于 Kohel 等人的研究启发式地表明四元数代数可以作为一种新的同源表示。De Feo 等人在其设计的 SQISign 签名算法中，给出了 KLPT 算法的完整有效实现，该协议是首个利用大素数同源的实例。文献[2]基于上述成果提出了一种新的通用同源表示方法，提出了子阶与理想转换问题 (Suborder to Ideal Problem, SOIP) 以及子阶与自同态环转换问题 (Suborder to Endomorphism Ring Problem, SOERP)，并基于此设计了基于同源的交互式密钥交换 (Non-Interactive Key Exchange, NIKE) 协议。

计算超奇异椭圆曲线的自同态环以及理想类群在超奇异椭圆曲线作用的反演是同源密码学中十分重要的两类计算性困难问题。其中计算超奇异椭圆曲线自同态环的困难性是同源密码学的基石，截至目前，最有效的解决方案仍是指数级的。另外，理想类群在椭圆曲线上的作用的反演问题引出了许多著名的同源密码算法，如早期 Couveignes 的工作以及 CSIDH。在 Kohel 等人提出定向（Orientation）的概念后，反演问题在同源密码学中所扮演的角色逐渐清晰，即使是 SIDH 这类不包含类群作用的方案，也可归约为反演问题。文献[3]证明上述两类问题在广义黎曼假设下是多项式相互归约的。在文献[3]的结果下，CSIDH 的安全性可归约到次指数时间的自同态环问题。

同源密码学是后量子密码学的重要分支之一，相较于格密码等其他几种后量子密码体系，同源密码学拥有最紧凑的密钥及密文大小，但存在实现效率较差的问题。有限域上固定度数的同源链的计算是同源密码学中的核心操作，该操作的快速实现是解决同源密码学效率问题的核心。传统的同源计算涉及选取核子群的生成元并利用 Vélú 范式，但选取核子群的生成元只能采用随机选取的方式。因此，研究者提出了采用模多项式以及除子多项式来避免扭曲点的选取。针对度数为 2 和 3 的同源，研究者提出了采用选取特殊的蒙哥马利曲线或 Hessian 曲线，像曲线的系数可被起始曲线的系数以及其简单的代数式（2 次根式或 3 次根式）表示。Castryck 等人提出了“radical isogeny”的概念，将计算度数为 2 和 3 的同源的方法推广到了任意度数 N 的同源，即像曲线可被起始曲线的系数以及其系数的 N 次根表示。但 Castryck 等人的工作遗留了 3 个主要问题。首先 Castryck 等人虽然论证了“radical isogeny”公式的存在性，但产生具体的实例十分困难，随着 N 的增长，除子多项式的度数以及函数域的复杂性增长迅速，实际上 Castryck 等人并没有给出 $N > 11$ 以上的“radical isogeny”实例；其次对于固定度数 N ，“radical isogeny”实例并不唯一，且不同的实例在性能上差异巨大，如何确定最有效的版本至关重要；最后对于 N 次根的选取，Castryck 等人也没有给出具体的解释。文献[4]针对 Castryck 等人遗留的问题，对其方法进行了改进。文献[4]给出了一种完全不同的寻找“radical isogeny”公式的方法，避免了函数域上除子多项式的因式分解，并成功生成了 $N=37$ 的“radical isogeny”公式，同时该文献论证其生成方法所产生的公式版本较 Castryck 等人更为有效。

本节作者：彭聪、黄伟晗、白野（武汉大学）

参考文献

[1] DE FEO L, DOBSON S, GALBRAITH S D, et al. SIDH proof of knowledge[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 310-339.

[2] LEROUX A. A new isogeny representation and applications to cryptography[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 3-35.

[3] WESOLOWSKI B. Orientations and the supersingular endomorphism ring problem[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 345-371.

[4] CASTRYCK W, DECRU T, HOUBEN M, et al. Horizontal racewalking using radical isogenies[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 67-96.

3 对后量子密码的攻击

随着 NIST 后量子密码标准征集工作进入第 3 轮，入围算法受到广泛关注。

不平衡油醋（Unbalanced Oil and Vinegar，UOV）体制和隐藏域方程（Hidden Field Equation，HFE）体制一直是多变量公钥密码研究的两大热点。Rainbow 体制是 UOV 类的数字签名算法，GeMSS 是 HFE 类数字签名体制。Rainbow 入选了 NIST 后量子密码标准征集第 3 轮数字签名算法，而 GeMSS 入选为 NIST 后量子密码标准征集第 3 轮的数字签名替补算法。

Rainbow 作为 UOV 体制的改进，使用多层油醋结构。自 1999 年以来，其选定参数还未被证明是不安全的，有望入围 NIST 后量子最终标准。Beullens 提出了一种针对 Rainbow 的新型攻击^[1]，攻击通过公钥的差分寻找第二层油空间 O_2 的核向量，代入公钥后使得两层的 Rainbow 变成 UOV 体制再进行后续攻击。对于第 2 轮满足 SL1 安全等级的 Rainbow 算法，该攻击可在 53 小时内恢复密钥。文章中还估计针对第 3 轮 SL1 参数选择，该攻击的时间预计增大 28 倍。文献[1]还指出，对于 SL3 和 SL5 的参数，将新型攻击和长方形最小秩攻击（Rectangular MinRank Attack）结合使用可以降低攻击复杂度，具体攻击复杂度分别为 2^{61} 、 2^{131} 、 2^{164} ，比已有攻击复杂度分别降低了 2^{62} 、 2^{20} 、 2^{27} 。

GeMSS 是一类 HFE 结合了减方法和醋变量方法的变体（HFEv-）。现有的对于 HFE 类体制安全性分析的最小秩攻击都是在中心映射所在的扩域上进行分析的。Baena 等人^[2]在支持余子式（Support-Minors，SM）攻击的基础上做出改进，可以在扩域上进行求解，使得通过简单增加参数大小或用投影技术改进 GeMSS 方案已无法确保其安全性。针对 GeMSS128，该攻击的时间复杂度为 272。在针对 Rainbow 的 MinRank 攻击中，需要大量的矩阵向量乘积运算，这增加了攻击查询内存所需成本，进一步提高了攻击的总成本。因此，文献[2]还提出

了一种内存优化策略,使用 Block Wiedemann 算法作为子程序降低了 XL 算法求解大型 SM 系统的内存访问成本,使得 Rainbow 在 MinRank 攻击下的安全性强度降低到安全性标准下。

NIST 后量子密码标准征集第 4 轮候选算法中有 3 个是基于格的公钥加密体制或密钥封装机制的。基于格的机制分为错误学习和舍入学习两类,后者包含了基于 N 次截断多项式环 (NTRU) 问题的方案。NTRU 和 NTRU Prime 是 NIST 后量子密码标准征集的候选算法。在选择密文模型下 (IND-CCA) 安全的基于 NTRU 的密钥封装机制已被研究了很长时间,现有的 IND-CCA 安全的方案以非常高的概率检测畸形密文,并且在检测时返回失败或伪随机输出,从而提供针对 CCA 的具体保护。

Ravi 等人^[3]提出了一种针对 IND-CCA 安全的密钥封装机制的基于侧信道的选择密文攻击。攻击构造了畸形密文,解封时攻击确保中间变量和私钥的联系十分紧密。随后攻击者通过侧信道获取变量信息,进而恢复完整的私钥。此外,还提出了几种新的 CCA 攻击方法,利用来自解封装过程的侧信道泄露,来实例化 3 种不同类型的谕言机,即明文检查谕言机、解密失败谕言机和完全解密谕言机。文献[3]对开源库中基于 NTRU 算法的优化进行了实验验证,只需对目标设备进行几千次选择密文查询,就能恢复完整的密钥,并且适用于 NTRU 和 NTRU Prime 的所有参数集。

后量子密码方案通常依赖拒绝抽样技术来从特定分布获取伪随机性。HQC 和 BIKE 是两种密钥交换方案,都使用了拒绝抽样算法来生成密钥,但是算法存在侧信道泄露的风险,可以被攻击者利用进行时序攻击。

Guo 等人^[4]提出了一种针对 HQC 和 BIKE 的新型密钥恢复时序攻击。通过采用非恒定时间拒绝抽样过程来生成具有特定权重的随机向量。无论 IND-CCA 变化中输入信息是否一致,由于拒绝抽样产生的时间差都会泄露,泄露的信息足以完成 HQC 和 BIKE 密钥恢复。攻击允许完全私钥恢复或高效消息恢复,并需要静态密钥设置和主动选择密文攻击者。文献[4]也提供了几种抵抗新攻击的对策,如增加最初生成的字节数,这虽然增加了攻击所需的工作量,但此对策仍允许攻击在低噪声环境中恢复密钥。文献[4]提出了消除所有时序变化的对策,这些变化可能会从分析的函数中检测到,并且强调了在后量子密码学中保护侧信道的重要性。

Kyber 是基于格的密钥封装算法,使用了一种称为 Frobenius 扭曲的技术来增强安全性。Kyber 已经被选为 NIST 的后量子密码标准之一。Delvaux 提出了一种新的可行的故障攻击家族 Roulette^[5],可以用于从 Kyber 实现中提取私钥。Roulette 利用掩蔽和混淆方法来随机化中间变量,从而进一步恢复密钥。攻击对于 Kyber-512 和 Kyber-768,成功率为 96%,对于 Kyber-1024,成功率为 96%,注入故障数量相对较少。文章中对运行 ARM-Cortex-M4 上的 Kyber 软件进行攻击实验,使用 ChipWhisperer 板进行故障注入,对于优化后的 Kyber 软件可在几小时到一天内恢复私钥,反而对于未优化的版本大约需要 5 天。文献[5]认为该攻击可能也适用于另一种基于格的 KEM——SABER。

基于同源密码学的超奇异同源问题是后量子密码学领域的一个重要问题，因为传统椭圆曲线在量子计算机的攻击下容易被破解。目前已有的解决此类问题的算法，如 Delfs-Galbraith 算法，复杂度仍然很高，需要进一步优化。Santos 等人^[6]提出了一种新算法 SuperSolver，用于解决基于同源密码学的基础问题——超奇异同源问题。SuperSolver 利用了一种新的子域根检测算法和模多项式的组合，可以有效地判定一个多项式是否在子域中具有根，而不需要显式地找到任何根。SuperSolver 可以找到子域 j 不变量，这是原始算法的瓶颈。虽然新算法和 Delfs-Galbraith 算法的渐进复杂度相同，但比传统的 Delfs-Galbraith 算法执行速度更快，并为超奇异同源问题的具体难度提供了新的见解。文献[6]使用 p 域上的平均乘法次数作为具体复杂度估计的度量标准，SuperSolver 所需平均乘法次数约是 Solver 算法的十分之一。

输出反馈计数器（COFB）是基于分组密码的轻量级认证加密与相关数据（AEAD）模式，是 GIFT-COFB 的基础，后者是 NIST 轻量级标准化项目的入围者之一。Khairallah 提出了一种针对 COFB 可能存在的伪造和隐私攻击^[7]，并表明伪造和隐私的安全性都受到伪造尝试次数的限制。COFB 虽然生成了 128 位标签，但其行为与具有 64 位标签的 AEAD 方案非常相似。该分析提供了对先前论文的主要定理的矛盾，并讨论了在可证明的结果和攻击之间存在安全差距的原因。具体来说，文献[7]讨论了 COFB 对选择密文攻击的安全性，提出了 3 种针对 COFB 的伪造敌手，需要大约 $2^{n/2}$ 次伪造尝试。第一个敌手是基于之前提出的分析框架的应用；第二个敌手是针对 COFB 的优化，称为优化块丢弃（OBD）敌手；第三个敌手是一个掩码碰撞（MC）敌手。文献[7]还讨论了并非每个对 NAE 方案的 INT-CTXT 敌手都可以用来构建 IND-CCA 敌手，即使后者伪造成功，也可能无法从 INT-CTXT 敌手获得任何有效信息。

本节作者：聂旭云、何晨宁（电子科技大学）

参考文献

- [1] BEULLENS W. Breaking Rainbow Takes a Weekend on a Laptop[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 464-479.
- [2] BAENA J, BRIAUD P, CABARCAS D, et al. Improving Support-Minors Rank Attacks: Applications to GeMSS and Rainbow[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 376-405.
- [3] RAVI P V, EZERMAN M F, BHASIN S, et al. Will You Cross the Threshold for Me? Generic Side-Channel Assisted Chosen-Ciphertext Attacks on NTRU-based KEMs[J]. IACR Transactions Cryptographic Hardware Embedded Systems, 2022(3): 722-761.

[4] GUO Q, HLAUSCHEK C, JOHANSSON T, et al. Don't Reject This: Key-Recovery Timing Attacks Due to Rejection-Sampling in HQC and BIKE[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(3): 223-263.

[5] DELVAUX J. Roulette: A Diverse Family of Feasible Fault Attacks on Masked Kyber[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 637-660.

[6] CORTE-REAL SANTOS M, COSTELLO, SHI J. Accelerating the Delfs-Galbraith Algorithm with Fast Subfield Root Detection[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 285-314.

[7] KHAIRALLAH M. Security of COFB against Chosen Ciphertext Attacks[J]. IACR Transactions on Symmetric Cryptology, 2022(1): 138-157.

4 其他

在 2022 年欧密会上, Grubbs 等人对后量子公钥加密协议的匿名性和健壮性进行了系统的研究^[1]。首先, Grubbs 等人将隐式拒绝确定为大多数后量子密钥封装机制共享的关键设计选择, 表明隐式拒绝对基于密钥封装机制-数据封装机制的公钥加密方案的匿名性和健壮性的先前结论不适用, 并尽可能将先前结论转移到隐式拒绝的情况下。其次, 由于 Fujisaki-Okamoto 变换被广泛用于构建后量子公钥加密, Grubbs 等人也研究了它们如何赋予公钥加密的健壮性和增强弱匿名性。最后, 基于上述得到的结论对 Saber、Kyber、McEliece 等密码的健壮性和匿名性进行了分析。

在文献[2]中, Tang 等人提出了一种基于交替三线性形式等价问题的实用签名方案。该方案受到 Goldreich-Micali-Wigderson 的图同构零知识协议的启发, 同时可以作为 NIST 后量子数字签名的备选方案。其中, Tang 等人提供了诸多理论证据, 如隐藏子群问题、多元密码学、基于群作用的密码学等问题的最新进展, 来支持其安全性, 并讨论了该方案的实际用途。

在 2022 年亚密会上, Yan 从标准量子比特承诺的角度研究了计算安全量子比特承诺的一般性质^[3]。该研究表明任意计算安全的量子比特承诺方案都可以转换为标准(非交互)形式。同时, 该研究证明了两种标准量子比特承诺是等价的, 即当且仅当标准统计隐藏计算绑定量子比特承诺存在时, 标准计算隐藏统计绑定量子比特承诺才存在。该结果暗示了计算安全的量子比特承诺是对称的。

在文献[4]中, Morimae 和 Yamakawa 表明若存在伪随机量子态, 则对经典消息而言存在具有计算隐藏和统计绑定的非交互式量子承诺。众所周知, 即使不存在量子安全的经典密码

原语, 伪随机量子态也可以存在。这暗示了即使不存在量子安全的经典密码原语, 量子承诺也可以存在。对于数字签名, Morimae 和 Yamakawa 表明若存在伪随机量子态, 则存在具有量子公钥的一次性安全数字签名。这也暗示了即使不存在量子安全经典密码原语 (包括量子安全单向函数), 量子签名也可以存在。在 2022 年欧密会上, Bitansky 等人在标准困难假设下构造了第一个抗量子攻击的不可延展的承诺, 该承诺在统计上具有约束力^[5]。该工作表明假设存在后量子单向函数, 则可获得对应的 $\log_2^*(\lambda)$ 轮 (其中 λ 为安全参数) 后量子承诺协议。此外, Bitansky 等人还提出一种新的通用技术, 可从任何可提取的承诺协议中模块化地提取不可延展的承诺, 无须考虑底层的提取策略或轮数规模。

在文献[6]中, Bitansky 和 Freizeit 基于 LPN (Learning Parity with Noise) 问题假设与标准的 Nisan-Wigderson 风格去随机化假设提出了具有统计发送者隐私的双消息不经意传输协议。具体地, Bitansky 和 Freizeit 先基于 LPN 设计了公共随机串模型下的具有统计发送者隐私的双消息不经意传输协议, 再对公共随机串进行去随机化。分析结果表明, 该协议可最大限度地减少各种密码学应用程序的轮复杂度。

在文献[7]中, Chia 等人从后量子半诚实不经意传输的最小假设出发, 构建了第一个 ϵ 可模拟的两方计算协议来对抗量子多项式时间敌手。该协议既是恒定轮又是黑盒, 其放宽了基于标准模拟的安全性, 即允许任意小的模拟误差 ϵ 。此外, 在允许量子通信的情况下, 可进一步削弱对后量子安全单向函数的假设, 同时保持恒定轮和黑盒属性。

在 2022 年美密会上, Bitansky 等人以构造性的方式将大部分经典归约转化为后量子归约^[8]。结果证明, 在多项式解空间的假设前提下 (如决策假设), 任何非交互式非自适应的经典归约都可以转化为后量子版本; 但是在超多项式解空间的假设前提下 (如一般搜索假设), 经典归约无法被转化。此外, Bitansky 等人还提出了一个针对有状态求解器的计算问题进行归约的框架, 并侧面验证了 Chiesa 等人的新型倒带技术是紧致的。

塌缩是一种抗碰撞性的后量子强化, 需要将许多经典结果增加到量子设定中^[9]。Zhandry 在 2022 年美密会上基于 4 种情形从抗碰撞哈希函数构造塌缩哈希函数, 进而说明半正则性是构造后量子哈希函数的重要考虑因素^[10]。在文献[11]中, 曹书蛟和薛锐给出两个关于量子抗碰撞性与塌缩性关系的结论: ①当哈希值原像的大小被某个多项式限时, 塌缩性和量子抗碰撞性肯定可以同时成立。②当原像的大小小于期望值时, 量子抗碰撞哈希函数存在表明坍塌哈希函数也存在, 进而说明, 两者之间的差距易受原像大小的影响。

目前对于不可克隆不可区分性的加密方案可行性的分析仍然是困难的。在文献[12]中, Ananth 等人给出 3 点关于不可克隆不可区分性的加密方案可行性的结论: ①满足不可克隆不可区分性的加密方案在量子随机黑盒模型中是无条件存在; ②给出大量的加密方案不能满足不可克隆不可区分性结果; ③确立了原语对单比特输出点函数复制保护的可行性。

Don 等人在欧密会 2022 证明只要量子随机黑盒模型中的量子查询算法输出一个与 $H(x)$ 有某种紧密关系的 t , 则 x 都可以被有效提取^[13]。同时, Don 等人给出通用在线可提取性结论的两个应用, 并证明了“承诺-和-开放”(Commit-and-Open) Σ -协议在量子环境下紧致的在线可提取性, 为典型的 Fujisaki-Okamoto 变换提供了第一个完整的后量子安全证明。而在已知的 Fujisaki-Okamoto 变换的安全缩减中, 解密失败是通过在给定私钥情况下简化查找失败明文的过程来进行处理。并且, 该处理需要隐式拒绝无效密文机制, 以便在量子可访问随机黑盒模型 (Quantum-accessible Random Oracle Model, QROM) 中实现合理的安全界限。Hovellmanns 等人在通过引入两个与查找解密相关的安全游戏提出不存在这些缺陷的缩减, 进而紧致了 QROM 中的安全边界^[14]。此外, Hovellmanns 还证明了几个表征 QROM 中的原像提取和某些搜索任务的技术结果。

针对 Hosoyamada 和 Yamakawa 提出的拥有量子计算能力与经典通信的双方 A 和 B 是否可以只依靠随机黑盒来确定对窃听者保密的密钥这样一个问题^[15], 在文献[16]中, Austrin 等人给出 3 种情形下的证明: ①当仅 A 具有经典能力, 而 B 具有量子能力时, 只要他们总共进行 d 次黑盒查询, 并以 1 的概率确定一个密钥, 那么总有办法可以通过 $O(d^2)$ 次经典黑盒查询攻破该密钥协商协议。②当双方都可以对随机黑盒进行量子查询时, 可以对任意密钥协商协议进行经典查询攻击。③在量子随机黑盒模型中, 找到对不完全完备的密钥协商协议的经典攻击是困难的。

量子标记化签名方案允许发送方生成和分发量子不可克隆态, 这授予持有者以发送方的名义进行一次性签名。而一个更强的原语是半量子标记化签名, 其中发送方是经典的且可将标记的生成委托给(可能是恶意的)量子接收方。在文献[17]中, Shmueli 基于量子安全不可区分混淆和带误差学习问题, 构造了一个半量子标记化签名方案。在该过程中, Shmueli 得到了量子陪集态的新性质和经典子空间隶属电路不可区分混淆的新结果。

2018 年 Mahadev 构建了一个量子计算的经典验证 (a Classical Verification of Quantum Computation, CVQC) 协议, 将有界误差量子多项式时间 (Bounded-error Quantum Polynomial time, BQP) 中的决策问题委托给计算假设下的不可信量子证明者。在文献[18]中, Chung 等人进一步探讨了 CVQC 在 BQP 中更一般的抽样问题, 并应用所提编译器产生 CVQC 在量子采样问题上的首个常轮盲 CVQC 协议。该协议具有可忽略不计的和逆多项式的稳健性误差, 以及可忽略不计的完整性误差。在 2022 年美密会上, Bartusek 等人为量子计算构造了一个经典可验证的简洁交互式论证, 其通信复杂度和验证器在 BQP 计算时运行时间是多项式对数级 (安全参数为多项式级)^[19]。基于不可区分混淆和容错学习的后量子安全性, 可保证该协议的安全性。不同于先前的工作, 既需要一个很长的公共引用字符串, 又需要非黑盒使用一个随机预言机模型的哈希函数, 这是普通模型上量子计算的首个简洁论证。

为了给基于编码密码学的 McEliece 密码和基于准循环方案的 BIKE、HQC 密码提供比特安全评估, Esser 等人给出了基于表现的信息集解码算法^[20], 如 May-Meurer-Thomae 算法和 Becker-Joux-May-Meurer 算法的具体实现。并且展示了 McEliece、BIKE 和 HQC 在中等规模情况下的实例, 即大约 60 比特安全级, 所给出信息解码算法的具体实现可以显著提高实际密码分析的速度。此外, Esser 等人发现 McEliece 在 Category 3 和 Category 5 情况下的实例并没有达到对应的安全级别, 而是分别还差 20 比特和 10 比特的安全距离。

在 2022 年美密会上, Hülsing 等人利用 EasyCrypt 工具集对 Saber 公钥加密方案 (NIST 后量子密码标准化的候选密码之一) 的选择明文攻击下的不可区分安全性和 δ -正确性进行了正式验证^[21], 即将该公钥加密方案转换为选择密文攻击下的自适应不可区分安全且 δ -正确的密钥封装机制所需的属性。为此, Hülsing 等人设计了对应的手写证明, 并将这些手写证明作为正式验证的指南。结合计算机验证, 最后结果证明了 Saber 公钥加密方案确实满足所需的安全性和正确性。

在可调哈希函数结构提供的抗目标碰撞、抗决定性第二原像、抗原像和不可检测性的假设下, Hülsing 等人在 2022 年亚密会上给出 SPHINCS+结构的严格安全性证明^[22]。此外, Hülsing 等人还分析了可调哈希函数结构在不可检测性和抗原像假设下的安全性, 并评估了对哈希函数族各种性质进行一般攻击时的破解困难性。

在 2022 年亚密会上, Liu 等人通过调整 Wagner 提出的广义生日问题^[23], 并将该技术应用于 c 元树结构, 以此提出 BKW 算法的变体算法^[24]。该变体算法优势在于: ①获得了与 Esser 等人工作^[25]相同的时空权衡, 但无须任何启发式算法。②相比于原始 BKW 算法和 Devadas 等人提出的 BKW 变体算法, 在时间/样本优化方面减少了 $2^{\Omega(n^{\frac{1}{1+\epsilon}})}$ ($\epsilon \rightarrow 0^+$) 因子。③针对有限样本下的 LPN 问题, 相比于 Lyubashevsky 提出的 BKW 变体算法, Liu 等人所提算法在时间复杂度方面减少了 $2^{\Omega(n)/\log_2(1-n)^{1-k}}$ ($k \rightarrow 1^-$) 因子。

在 2006 年欧密会上, Kelsey 和 Kohno 提出占卜者攻击 (Nostradamus Attack)^[26], 即敌手必须提交一个迭代哈希函数 H 的哈希值 y , 如此一来当稍后给定消息前缀 P 时, 敌手能够找到一个合适的“后缀解释” S , 使得 $H(P \parallel S) = y$ 。随后, Kelsey 和 Kohno 给出一种对哈希函数 H 进行 $2^{n/3}$ (其中 n 为函数输出长度) 次访问的任意选定前缀原像攻击 (Herding Attack)。在 2022 年亚密会上, Benedikt 等人考虑了量子敌手实施占卜者攻击的安全性, 提出一种求解占卜者问题的量子原像攻击算法^[27]。该算法对哈希函数 H 的访问次数约为 $\sqrt[3]{n} \cdot 2^{3n/7}$, 显著提升经典攻击所需的访问次数。并且, Benedikt 等人证明了对于随机压缩函数而言量子原像攻击算法最少需要进行 $2^{3n/7}$ 次的函数访问, 这表明其所提算法 (本质上) 是最优的。此外, 他们还讨论得出针对随机压缩函数的一般占卜者攻击的一个接近紧致的界 $2^{3n/7-s}$ (其中 s 是所选后缀 S 的最大块长度)。

Even-Mansour 结构是一种从 n 比特公共随机置换 P 构造伪随机置换 E 的简单方法。在经典攻击模型下, 最优攻击需要对 E 进行 qE 次查询, 对 P 进行 qP 次查询, 使得 $qE \cdot qP \approx 2^n$ 。然而, 若攻击者被授予对 E 和 P 的量子查询, 则该算法是完全不安全的, 攻击者仅需要 $qE \approx qP = O(n)$ 次量子查询。然而, 在常见的现实环境中, 量子攻击者一般只能对带密钥的 E 进行经典查询, 而对公开置换 P 可进行量子叠加查询。在这种情况下, 已知的攻击复杂度满足 $qE \cdot qP^2 \approx 2^n$, 该结果表明与经典模型相比其安全性有所下降, 但在该后量子模型 (Q1 模型) 下 Even-Mansour 密码结构是否可以安全性证明仍是公开问题。在文献[28]中, Alagic 等人解决了该问题, 表明在 Q1 模型下对于 Even-Mansour 结构的任何攻击查询复杂度都需要满足 $qE \cdot qP^2 + qP \cdot qE^2 \approx 2^n$, 该结论适用于 Even-Mansour 的双密钥和单密钥变体。基于该结论, Alagic 等人对之前关于量子查询下界工作结果进行了推广。

在 Simon、Shor 等量子周期搜索算法中, 需要通过对周期函数 f 进行量子叠加访问产生量子叠加态 $\sum_x |x\rangle |f(x)\rangle$, 这要求用一定数量的输出量子比特来表示 $|f(x)\rangle$ 。在 2022 年 FSE 会议上, May 和 Schlieper 通过在 Oracle 中应用 Hash 技术将该输出量子比特个数降为单量子比特, 显著减少所需量子比特的总数^[29]。在预期情况下, 该方法只会使算法所需测量次数增加一倍, 同时显著减少所需量子比特的总数。

本节作者: 高飞 (北京邮电大学), 罗宜元 (惠州学院)

参考文献

[1] GRUBBS P, MARAM V, PATERSON K G. Anonymous, robust post-quantum public key encryption[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 402-432.

[2] TANG G, DUONG D H, JOUX A, et al. Practical post-quantum signature schemes from isomorphism problems of trilinear forms[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 582-612.

[3] YAN J. General properties of quantum bit commitments[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 628-657.

[4] MORIMAE T, YAMAKAWA T. Quantum commitments and signatures without one-way functions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 269-295.

[5] BITANSKY N, LIN H, SHMUELI O. Non-malleable Commitments Against Quantum Attacks[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 519-550.

[6] BITANSKY N, FREIZEIT S. Statistically sender-private OT from LPN and derandomization[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 625-653.

[7] CHIA N H, CHUNG K M, LIANG X, et al. Post-quantum simulatable extraction with minimal assumptions: Black-box and constant-round[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 533-563.

[8] BITANSKY N, BRAKERSKI Z, KALAI Y-T. Constructive Post-Quantum Reductions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 654-683.

[9] UNRUH D. Computationally binding quantum commitments[C]. Advances in Cryptology-EUROCRYPT 2016: EUROCRYPT 2016, Vienna, Austria, May 8-12, 2016. Springer Berlin Heidelberg, 2016: 497-527.

[10] ZHANDRY M. New constructions of collapsing hashes[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 596-624.

[11] CAO S, XUE R. The gap is sensitive to size of preimages: Collapsing property doesn't go beyond quantum collision-resistance for preimages bounded hash functions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 564-595.

[12] ANANTH P, KALEOGLU F, LI X, et al. On the feasibility of unclonable encryption, and more[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 212-241.

[13] DON J, FEHR S, MAJENZ C, et al. Online-extractability in the quantum random-oracle model[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 677-706.

[14] HOVELMANN K, HULSING A, MAJENZ C. Failing gracefully: decryption failures and the Fujisaki-Okamoto transform[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 414-443.

[15] HOSOYAMADA A, YAMAKAWA T. Finding collisions in a quantum world: quantum black-box separation of collision-resistance and one-wayness[C]. Advances in Cryptology-ASIACRYPT 2020: ASIACRYPT 2020, Daejeon, the Republic Korea, December 7-11, 2020. Cham: Springer Nature Switzerland, 2020: 3-32.

[16] AUSTRIN P, CHUNG H, CHUNG K M, et al. On the impossibility of key agreements from quantum random oracles[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 165-194.

[17] SHMUELI O. Semi-quantum Tokenized Signatures[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 296-319.

[18] CHUNG K, LEE Y, LIN H, et al. Constant-Round Blind Classical Verification of Quantum Sampling[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 707-736.

[19] BARTUSEK J, KALAI Y T, LOMBARDI A, et al. Succinct Classical Verification of Quantum Computation[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 195-211.

[20] ESSER A, MAY A, ZWEYDINGER F. McEliece Needs a Break - Solving McEliece-1284 and Quasi-Cyclic-2918 with Modern ISD[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 433-457.

[21] HÜLSING A, MEIJERS M, STRUB P-V. Formal Verification of Saber's Public-Key Encryption Scheme in EasyCrypt[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 622-653.

[22] HÜLSING A, KUDINOV M. Recovering the Tight Security Proof of SPHINCS+[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 3-33.

[23] WAGNER D. A Generalized Birthday Problem[C]. Advances in Cryptology-CRYPTO 2002: CRYPTO 2002, California, USA, August 18-22, 2002. Cham: Springer Berlin Heidelberg, 2002: 288-304.

[24] LIU H, Yu Y. A Non-heuristic Approach to Time-Space Tradeoffs and Optimizations for BKW[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 741-770.

[25] ESSER A, HEUER F, KUBLER R, et al. Dissection-BKW[C]. Advances in Cryptology-CRYPTO 2018: CRYPTO 2018. Santa Barbara, CA, USA, August 19-23, 2018. Cham: Springer Nature Switzerland, 2018: 638-666.

[26] KELSEY J, KOHNO T. Herding Hash Functions and the Nostradamus Attack[C]. Advances in Cryptology-EUROCRYPT 2006: EUROCRYPT 2006, St. Petersburg, Russia, May 28-June 1, 2006. Cham: Springer Berlin, Heidelberg, 2006: 183-200.

[27] BENEDIKT B J, FISCHLIN M, HUPPERT M. Nostradamus Goes Quantum[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 583-613.

[28] ALAGIC G, BAI C, KATZ J, et al. (2022). Post-Quantum Security of the Even-Mansour Cipher[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 458-487.

[29] MAY A, SCHLIEPER L. Quantum Period Finding is Compression Robust[J]. IACR Transactions on Symmetric Cryptology, 2022(1): 183-211.

安全协议

1 混淆/不可区分性

程序混淆来源于软件工程领域反编译的需求，同时在计算复杂性和密码学领域有着很高的理论价值和应用前景。在程序混淆领域得到最广泛研究的当属不可区分混淆（indistinguishability Obfuscation, iO）。在 iO 的构造方面，2021 年，Jain、Lin 和 Sahai 提出了首个基于标准模型的构造，这一构造用到了包含带噪声学习（Learning With Errors, LWE）假设在内的 4 个假设。在文献[1]中，作者进一步去掉了其中的 LWE 假设。LWE 假设是很多功能强大的密码学应用所依赖的假设，如全同态加密，这一结果同时表明其他 3 个假设的组合有着比肩 LWE 假设的强大能力。除了基于标准假设的构造，iO 还有许多启发式的候选方案，如 Bartusek、Ishai、Jain 等人在 2020 年提出的用于仿射行列式程序的混淆方案。在文献[2]中，作者对该方案进行了密码分析，指出该方案在混淆特定程序时不满足不可区分混淆的安全定义，同时他们提出了新的补救方案。

与此同时，以 iO 为基础进一步构造适用于各种场景的密码学应用也是一个热门的研究方向。在文献[3]中，作者利用 iO 首次构造了分布式采样器和支持任意相关关系的伪随机相关函数。分布式采样器可以让多个参与方非交互地产生公共引用串，而伪随机函数可以让多个参与方通过亚线性的通信非交互地分别得到具有某种相关性的串，此前的伪随机相关函数只支持不经意传输和向量形式的不经意线性估值（Vector-Oblivious Linear Evaluation, VOLE）。由于 iO 的安全性过于强大，当我们得到一个混淆后的程序时，几乎只能对它进行黑盒访问，因此我们往往难以判断原始程序是不是按照期望的方式编写的，或者原始程序是否剽窃并调用了某个第三方的程序。在文献[4]中，作者形式化地定义了选择混淆攻击（Chosen Obfuscation Attack, COA）安全混淆，这类混淆允许验证者验证其原始程序的某个谓词（Predicate），同时两个相似且拥有相同谓词的程序在混淆后是不可区分的，即使敌手能够访问某个可以反编译混淆程序的预言机。作者在标准模型下基于 iO 给出了 COA 安全混淆的构造，并探索了这一定义在密码学领域的应用场景。

除了复杂函数的混淆，简单函数如点函数的混淆可能也有着意想不到的应用。在文献[5]中，作者定义了一个名为不可延展的带关联数据的点函数混淆并给出了相应的构造。以此为

基础再加上使用零知识证明，作者首次在标准模型下构造了不可延展的数字保险箱；将这一构造和校验子译码结合便得到了更强大的可以用于低熵输入的聚合模糊提取器。

混淆的存在也能作为假设去得到计算复杂性领域中一些问题的界。例如，在窃听信道模型中，我们希望通过编码使得窃听者无法获取接收者收到的消息。如果消息接收者的信道是窃听者信道的低配版，即窃听者可以获取消息接收者收到的所有消息，这一目标显然是不可能达成的。排除这一情况是否就足够了呢？在信息论意义上仍然不够，已有的结论指出，当且仅当窃听者的信道噪声更大，窃听信道编码才是可能的。在文献[6]中，作者证明了假设某种形式的混淆存在，窃听信道编码存在当且仅当接收者的信道不是窃听者信道的低配版。

本节作者：郁昱、姚立（上海交通大学）

参考文献

- [1] JAIN A, LIN H, SAHAI A. Indistinguishability Obfuscation from LPN over F_p , DLIN, and PRGs in NC^0 [C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 670-699.
- [2] YAO L, CHEN Y, YU Y. Cryptanalysis of Candidate Obfuscators for Affine Determinant Programs[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 645-669.
- [3] ABRAM D, SCHOLL P, YAKOUBOV S. Distributed (Correlation) Samplers: How to Remove a Trusted Dealer in One Round[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 790-820.
- [4] CANETTI R, CHAKRABORTY S, et al. COA-Secure Obfuscation and Applications[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 731-758.
- [5] APON D, CACHET C, FULLER B, et al. Nonmalleable Digital Lockers and Robust Fuzzy Extractors in the Plain Model[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 353-383.
- [6] ISHAI Y, KORB A, LOU P, et al. Beyond the Csiszár-Korner Bound: Best-Possible Wiretap Coding via Obfuscation[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 573-602.

2 零知识证明

零知识证明是一个基础性的密码学概念，它使得证明者能够向验证者证明某个断言是正确的而不透露任何其他信息。近年来，这一领域无论是在理论还是在应用层面都受到了大量关注和研究，取得了丰硕的成果。下面将介绍零知识证明在 2022 年里取得的一些令人瞩目的成果。

2.1 交互式零知识证明理论

轮复杂度是交互式零知识证明协议的关键效率指标之一，如何构造低轮复杂度的公开抛币黑盒零知识协议是零知识领域十分热门的研究课题。近年来达到 3 轮的零知识协议大多基于不可证伪假设，或者其满足的是相对较弱的安全性定义，Bitansky 等人 2018 年基于拟多项式困难的 LWE 问题以及假设多重抗碰撞无密钥哈希函数的存在性构造了 3 轮零知识协议，而 Kiyoshima 在文献[1]中基于与 Bitansky 等人基本相同的假设（将假设 LWE 的拟多项式困难增强到亚指数困难）构造了 3 轮公开抛币的零知识协议。国内学者邓霖等人通过开发知识加密这一新型原语，在包括多项式困难的 LWE 假设等多种标准假设下实现了包括 3 轮可模拟安全的不经意传输协议，3 轮零知识协议在内的多种密码协议构造^[2]。交互协议的并行执行是常见的在保持轮数的情况下降低可靠性错误（Soundness Error）的方法，但学界对于协议并行执行下知识错误（Knowledge Error）的变化却不够了解，Attema 等人证明了对任意 $(k_1, \dots, k_\mu)$ -特殊健全性的多轮公开抛币交互协议的 t 折并行重复，可以将知识错误从 k 降低到 k' ，且知识错误的缩减程度在 t 折并行重复下是最优的^[3]。除了交互协议的轮复杂度，如何在标准假设下提升协议的通信量和效率也是一个热门的研究话题。Kim 等人在公共随机串模型下针对大小为 N 的算术电路构造了 3 个零知识交互论证协议^[4]，它们都能够在随机谕言机（RO）模型下通过 Fiat-Shamir 启发式被转化为非交互协议。他们所提出的 3 个论证协议的通信量分别为 $O(\sqrt{\log_2 N})$ 、 $O(\log_2 N)$ 、 $O(\log_2 N)$ ，验证者的计算复杂度分别为 $O(N)$ 、 $O(\sqrt{N})$ 、 $O(\sqrt{N} \log_2 N)$ 。这 3 个协议的可靠性都能在标准离散对数和/或至少与 DDH 假设一样可靠的双配对假设（Double Pairing Assumption）下证明。在没有可信初始设置以及只用标准假设的情况下，第一个协议是首个满足亚对数（Sublogarithmic）通信量的交互论证；第二个协议是首个满足 $O(\log_2 N)$ 通信量及 $O(\sqrt{N})$ 验证时间的交互论证；而不同于前两个协议，第三个协议不依赖于配对友好的椭圆曲线。

交互式谕言机证明 (Interactive Oracle Proof, IOP) 是一种交互式证明, 其中验证者无须阅读证明者的全部消息, 而是能够以访问谕言机的形式访问证明者消息。IOP 可被看作是概率可检查证明 (Probabilistically Checkable Proof, PCP) 的交互式版本, 其近年来常用于构造高效的证明系统。文献[5]在仅使用单向函数的条件下, 对于 bounded-space 的 NP 关系, 将 IOP 转化为零知识证明, 这种转化方法可进一步用于构造出当前长度最短的零知识证明。此外, 文章还分析了构造更一般 NP 关系的简洁 IOP 的障碍, 并进一步证明在几个简单而可信的复杂性理论猜想下, 对于 CSAT 关系不存在简洁 IOP。除了证明尺寸的简洁性, 证明者电路的简洁性的研究也很重要。在文献[6]之前, 对于布尔电路的 IOP 要么是基于 Ben-Sasson 等人在 2013 年提出的高效 PCP, 要么是基于 Rothblum 等人在 2022 年提出的高效 PCP, 证明者尺寸分别为 $\text{polylog}(|C|)$ 和 $\text{poly}(\lambda)$ 。对于许多不同种类的布尔电路 $C=C(x, \omega)$, 该文献对断言 $\{x: \exists \omega C(x, \omega)=1\}$ 以 $2^{-\lambda}$ 的可靠性误差构造了简洁论证系统, 证明者尺寸为 $\text{polylog}(\lambda)$ 。

2.2 非交互零知识证明理论

Σ -协议是一种 3 轮公开掷币的交互式协议, 这类协议往往可以通过 Fiat-Shamir 转换在 RO 模型下被转化为非交互的。Zeng 等人针对 k 合取范式关系 (k -Conjunctive Normal Form relations, k -CNF) 构造了 DAG- Σ -协议^[7], 相比于 Cramer 等人 1994 年的经典构造大幅缩减了通信量以及计算开销。Goel 等人构造了一种通用框架^[8], 能将对许多相互独立断言的 Σ -协议编译为对一个新的或断言的 Σ -协议, 新 Σ -协议的通信复杂度取决于原有断言中复杂度最高的一项。

基于“承诺并打开”(C&O) 范式的 Σ -协议与 Fiat-Shamir 转换也被广泛地应用在 NIZK 协议与数字签名方案的构造中。当使用基于哈希的承诺方案时, 这些构造在 RO 模型下具备线上可抽取性 (Online-Extractability)。然而, 由于 RO 模型在量子环境下的安全性并不可靠, 这些构造一般仅有经典意义下的安全性。在文献[9]中, Don 等人通过在量子随机谕言机 (QRO) 下证明了 (一大类) C&O Σ -协议在 Fiat-Shamir 转换下保留紧的线上可抽取性, 证明了这类构造支持抗量子安全性。该文献分别考虑了逐项哈希的承诺方案与基于 Merkle 树的承诺方案的两类情形, 并显著地改善了抗量子数字签名方案 Picnic 的表现。另外, 作为 RO 模型下的一种归约技术, 直线抽取 (Straight-Line Extraction) 的概念是针对可靠性的概念提出的, 直线抽取器可以在不重绕的情况下将一个定理的证据从证明者中抽取出来。直线抽取的技术在理论层面可以被用于安全性的合成以及获得紧安全性 (Tight Security), 在应用中可以被用于解决聚合签名构造中存在的问题。Kondi 等人在计算成本方面推进了 RO 模型中直线抽取的

研究,并优化了此前最先进的 Fischlin 变换的适用性^[10]。相对于在 RO 模型下的转换,近年来对 Fiat-Shamir 转换标准模型安全性的研究也逐渐活跃起来。文献[11]针对与 Paillier 的复合剩余假设相关的语言给出了首个无须重复执行就能保障可靠性的陷门 Σ -协议,解决了标准模型下基于 LWE 假设的转化需要并行重复执行的问题,且该协议与 Schnorr 在 RO 模型下利用 Fiat-Shamir 转换得到的论证长度相当。该协议结合 Correlation Intractable 哈希函数,进一步构造了针对断言“ $\mathbb{Z}_{N_2}^*$ 中的一个元素是复合剩余”的模拟可靠的非交互零知识 (NIZK),在此基础上也给出了标准模型标准假设下具有最短签名长度的环签名方案。

细粒度密码学 (Fine-Grained Cryptography) 主要研究在敌手具有有限计算资源而诚实用户的计算资源少于敌手的情况下的密码方案,旨在研究攻击某些密码方案的确切复杂度,设计更高效的方案并将其安全性建立在较弱的假设上。在文献[12]中,王煜宇等人在 NC^1 细粒度场景中构造了 NIZK 证明系统,这是第一个在细粒度场景中将诚实证明者时间限制在比敌手低的系统,而此前的构造要求证明者在多项式时间运行。紧随其后,王煜宇等人在文献[13]中建立了第一个 AC^0 细粒度设置中无条件安全的 NIZK 系统,其完美可靠性对于所有敌手成立,而对 AC^0 的敌手具有无条件的计算零知识。

对零知识证明的不可延展性分析十分重要,在实际应用中,延展性攻击可能会造成极为恶劣的影响。针对现有标准模型下具有不可延展属性的零知识比一般零知识慢上几个数量级的问题,文献[14]利用基于实例的不可延展承诺,提出首个在普通模型下将高效通用零知识协议转化为高效通用不可延展零知识的方法以及首个不需要公钥假设的高效不可延展承诺方案。著名的、已在电子金融系统中被广泛应用的证明系统——基于 Fiat-Shamir 转换的 Bulletproofs 或其他类似的多轮证明系统,在文献[15]中被证明在代数群模型中具有模拟可抽取性,这也意味着它们具有不可延展性。

此外,还有一些有创意的零知识协议构造被提出。在文献[16]中,Canetti 等人基于 LWE 假设或 LPN+DDH 假设,在 Universally Composable (UC) 场景中构造了第一个三重自适应 (Triply Adaptive) NIZK,为多实例证明提供 UC 安全性、自适应可靠性和自适应零知识。此前的三重自适应 NIZK 基于不可证伪假设 KOE,并且尚不具备 UC 安全性。在文献[17]中,国内学者杨康等人针对非交互零知识证明 (NIZK) 和指定验证者零知识证明 (DVZK) 各自效率特点,提出了多验证者零知识证明 (MVZK),支持一次向多个验证者证明命题的合法性,具有 DVZK 的证明生成速度快、内存开销低的效率优势,同时具有 NIZK 的通信轮数低的效率优势,通信效率显著优于运行多次 DVZK 证明。

随着量子计算的发展,如何证明量子版本的 NP 问题——QMA 问题,成为新的研究热点。针对如何构造经典可验证的零知识证明方案这一问题,在文献[18]中 Morimae 等人构造了 3 个在不同量子预处理模型下的针对 QMA 问题的经典可验证非交互零知识证明/论证系统 (CV-

NIZK): ①在量子秘密参数模型(存在可信第三方分别向证明者与验证者发送量子证明私钥与经典验证私钥)下,构造了满足信息论意义下可靠性与零知识性的 CV-NIZK 方案;②假设 LWE 问题的量子困难性,在一类特殊的预处理模型(存在可信方生成 CRS 的同时,验证方向证明方发送一个断言独立的量子消息作为预处理)下,构造了 CV-NIZK 方案,该方案有两种不同的计算不可区分的 CRS 生成方式,可分别实现信息论意义下的可靠性或零知识性;③在 QRO 模型下,在量子预处理模型下构造了一个 CV-NIZK 方案。该预处理仅要求验证者向证明者发送一个随机 Pauli-basis 态,可替换成由可信第三方向证明者与验证者分发 Bell 态。针对标准模型下对于一般 NP(量子环境下为 QMA)问题无法构造同时满足统计可靠性与统计零知识性的零知识方案这一结论,在文献[19]中,Hiroka 等人提出了一类全新的关于 QMA 问题的零知识证明方案,被称为 certified everlasting 零知识证明方案。这类方案具有统计可靠性与计算零知识性,并且验证方可以发布一个经典的证书说明其删除了其量子信息,当证书有效时,即便是无穷能力的敌手也不能再获取除断言有效之外的任何信息。文献[19]通过提出并构造了具有统计绑定性与 Certified Everlasting 隐藏性的承诺方案,结合 Broadbent 与 Grilo 的量子 Σ 协议,实现了 Certified Everlasting 零知识证明方案构造。

2.3 简洁非交互零知识论证系统

近期,作为一类目前在电子金融中已被广泛应用的论证系统,简洁非交互零知识论证系统/知识的论证系统(zk-SNARG/SNARK)也得到了快速的发展。

在理论层面,具有透明初始设置(Transparent Setup)的 SNARK 虽然具有一个巨大的优势——不需要可信第三方来生成公共参考串,但目前该类 SNARK 的安全性只在随机谕言机模型(Random Oracle Model, ROM)下被证明,而这个谕言机实际应用中只能被启发式实例化并黑盒地使用。事实上,在文献[20]中,Haitner 等人基于指数时间假设证明了当前的 ROM-SNARG 具有最优的证明长度下界 $\Omega(\log_2(t/\varepsilon) \cdot \log_2(t))$,这项结果适用于所有已知的 ROM-SNARG 结构:具有确定性非自适应验证者和 (t, ε) -salted-soundness。而 Chen 等人发现了一个自然的谕言机模型——低阶随机谕言机模型,并在线性码随机谕言机模型(LCROM)下对其进行了分析^[21]。而在低阶随机谕言机模型中存在用于与该谕言机相关的所有 NP 计算的具有透明初始设置的 SNARK,这样的 SNARK 可以直接证明关于其自身验证者算法的计算。这也直接在该谕言机模型下引出了仅仅依赖于标准模型下抗碰撞哈希函数的存在性的携带数据的证明(PCD)。另外,Hulett 等人在文献[22]中对一类非确定性计算实例化了 batch-NP 交互论证系统,通过利用 Correlation Intractable Hash(CIH)等工具,在二次剩余假设和亚指数 DDH 假设的基础上构造了 SNARG,这也是第一个基于标准群假设的 SNARG。

在应用层面, Bootle 等人在文献[23]中, 构造了第一个具有线性复杂度证明者时间和多对数 (Polylogarithmic) 复杂度验证者时间的零知识论证系统, 该构造黑盒调用密码组件 (如抗碰撞哈希函数) 来将概率证明编译为论证系统。该方案首先基于 R1CS 约束系统构造一系列张量交互式谕言机证明 (Tensor IOP), 之后将这些张量交互式谕言机证明转换为一个标准的用于证明 R1CS 问题的交互式谕言机证明, 结合用于证明一致性的 IOPP 与具体的抗碰撞哈希函数构造出论证系统。与之前的方案对比, 该方案优化了通信复杂度和验证者复杂度且增加了零知识性。但这个方案还是无法给出具体的实现应用, 因此仅局限于理论意义。谢天成等人在文献[24]中, 给出了相近渐进复杂度的方案, 该方案还给出了具体的实现。其方案构造分为两部分: 第一部分是提出了一个具有常数相关性距离 (Relative Distance) 的线性时间编码方案的随机化构造, 这样的编码被用于现有的所有具有线性证明者时间复杂度的零知识证明方案里, 关键技术在于安全扩张器采样 (Secure Expander Sampling); 第二部分通过一种码字转换技术 (Code-Switching) 得到了非黑盒的证明组合 (Non-black-box Proof Composition), 从而使得最终方案得到了对数平方级别的证明尺寸, 并给出了具体的实现。Kothapalli 等人在文献[25]中, 构造了一种新的用于证明递归计算 (Recursive Computation) 的零知识论证系统。这种系统用来证明一个非确定的函数原始输入重复执行多次, 每次都有不同的证据输入, 最后输出最终结果的正确性, 其广泛应用于可验证延迟函数 (Verifiable Delay Function)、零知识虚拟机 (zk Virtual Machines) 和零知识堆叠 (zk Rollups)。该方案基于一个类似 R1CS 的约束系统, 利用用于打包两个实例为一个的折叠方案 (Folding Scheme) 设计了一个递增可验证计算方案 (Incrementally Verifiable Computation, IVC), 其保证随着递归层数的增加证明尺寸不变。最后结合一个可以高效证明 IVC 的简洁非交互零知识论证得到完整的方案, 在该最终方案中, 证明者单次递归时间是最优的, 且拥有对数级别的证明尺寸。Bootle 等人在文献[26]中, 构造了一个弹性简洁非交互知识论证系列 (Elastic SNARK), 称为 Gemini。在这类简洁论证系统中, 证明者具有不同的时间和内存权衡的多个配置, 可以根据执行环境和被证明的断言来选择这些配置, 输出的证明与所选择的配置无关。其中在时间效率高的配置中, 证明者为线性时间和存储复杂度; 在存储效率高的配置中, 证明者为拟线性的时间复杂度和对数存储复杂度。相比较其他 SNARK, Gemini 可以更好地用于规模比较大的电路 (对证明者存储要求较低)。另外, 针对目前许多已被广泛应用的零知识证明需要可信初始设置的问题, 文献[27]在离散对数场景下提出了新型高效透明诚实验证者零知识论证 Flashproof。文章中提出了一个 gas-efficient 的范围证明, 在区块链平台上实现了与已知最高效的 zkSNARK 相当的汽油费 (Gas), 随后基于 Bayer 与 Groth 的技术提出了一个计算与通信均高效的多项式评估论证。此外, 群的实现是实现整个 SNARK 的基础, 因此研究 SNARK 常用曲线的各种性质也至关重要。Housni 等人推广了 BLS12 和 BW6 的曲线构造、配对 (Pairing) 公式和在任何 BLS12 曲线顶部定义的任何 BW6 曲线的分组操作, 形成 2-链配对-友好 (Pairing-Friendly)

曲线族^[28]。进一步地，他们研究了在 BLS12 和 BLS24 曲线上形成的其他可能的 2-链家族，并表明对于 BLS12 和 BLS24，与 Cocks-Pinch 曲线相比，BW6 构造总是给出最快的配对和曲线运算。最后，他们基于 BLS12 和 BLS24 的族分别针对 Groth16 和基于 KZG 的 SNARK 递归证明合成进行了定制优化，提出了一个适用于 Groth16 和基于 KZG 的通用 SNARK 的短曲线列表，并给出了这些曲线的优化实现。另外，在抗量子的环境下没有类似配对这样具有良好性质的工具是导致构造抗量子的公开可验证 zkSNARK 十分困难的主要原因之一。在文献[38]中，Albrecht 等人首次提出了将配对结构迁移到格结构的转换方法，并给出了一个基于格的 zkSNARK 构造，该构造具有许多良好性质，包括后量子安全、公开可验证、对数级别的验证时间、易于递归组合。本构造的核心是一个可抽取的（Extractable）基于格的向量承诺（VC）方案，它支持常数度多元多项式映射的打开。

2.4 其他实用非交互零知识证明系统

除了应用于一般 NP 断言的 zkSNARK 协议，还有一些为了解决特定问题而提出的证明系统。近年来，有关可验证延迟函数（Verifiable delay function, VDF）以及基于时间的不可否认性的研究有不少，出现了诸如短期身份认证协议以及受时间限制的加密、签名、承诺方案等。在文献[29]中，Arun 等人提出了 3 个全新的概念——可重随机化的 VDF、零知识 VDF、短期证明（Short-Lived Proof），其中可重随机化的 VDF 和零知识 VDF 可作为组件用于优化短期证明的构造。Arun 等人构造了高效的短期证明，该证明一经发布即可实现在无证明者或任意第三方后期参与的情况下，在一段指定的时间后自动失去其可靠性（Soundness）。该构造的核心思想是允许某个敌手在执行大量顺序计算后任意伪造证明从而提供一种不可否认性。Arun 等人基于短期证明还进一步地构造了一个高效的短期签名（Short-Lived Signatures）方案。而零知识集是一类特殊的“承诺-证明”系统，其允许对一个集合进行承诺，并且能零知识证明其中的（非）成员关系。近年来，现有的只扩展零知识集（append-only Zero-Knowledge Sets, aZKS）和密钥传输（Key Transparency, KT）方案无法在服务器受损后充分恢复隐私保证，相当于妥协后安全性（Post-Compromise Security, PCS），同时保持可审核性。在文献[30]中，Chen 等人提出了可旋转（Rotatable）零知识集的概念，并基于此构造了首个后向安全（Post-Compromise Security）的密钥传输系统。该文献通过一类全新的原语，可旋转的且可验证的随机函数，结合有序累加器与只扩展的向量承诺方案实现了可旋转零知识集的构造。Hoffmann 等人则在文献[31]中，构造了一种具有统计可靠性且对于任意群的群指数证明（Proofs of Exponentiation, PoE）。与之前最好的对于任意群具有统计可靠性的 PoE 方案相比，新的方案在通信复杂度上降低了一个数量级。

此外, 目前已有的零知识证明系统在实际应用的过程中仍然存在一些问题, 针对这些实际问题也有一些研究成果。Cohen 等人^[32]提出了“轮换采样”(Round-Robin Sampling)这一安全概念, 其刻画了如 zkSNARK 中“power of tau”协议的一些特征, 这类协议通常需要 n 轮的顺序广播, 而 Cohen 等人将这类协议所需的广播轮数缩减到 $O(\sqrt{n})$ 。另外, 目前零知识证明系统的设计通常都只针对 $\mathbb{F}_2$ 或 $\mathbb{F}_p$ 上的电路计算, 而不是所有现代计算机 CPU 操作所基于的 $\mathbb{Z}_{2^k}$ 上的计算。而且在素数模下模拟 $\mathbb{Z}_{2^k}$ 算术会带来不可避免的额外开销。尽管此前有直接针对 $\mathbb{Z}_{2^k}$ 的指定验证者的零知识证明系统, 但该构造需要一个目前还无法高效大量产生的能在 $\mathbb{Z}_{2^k}$ 上实例化的预处理随机向量不经意线性赋值 (random Vector Oblivious Linear Evaluation, VOLE)。Baum 等人提出了一个恶意安全的 VOLE 扩展协议, 该协议可以将 $\mathbb{Z}_{2^k}$ 上的短种子 VOLE 转化为同一个环上的更长的伪随机 VOLE^[33]。此外, Baum 等人还展示了 QuickSilver 零知识证明系统所采取的方法可以被拓展成支持 $\mathbb{Z}_{2^k}$ 上运算的版本, 从而得到一个新的比目前已有方案更高效的基于 VOLE 的证明系统。

随着量子计算机的发展, 大量传统的密码学假设 (如 RSA 假设、离散对数假设等) 被证明不具备抗量子的安全性, 因此学界开始寻求具备抗量子特性的与良好特性的密码学假设以实现抗量子密码原语的构造, 而基于格上困难问题的零知识证明协议的构造则是目前的一个热门研究话题。在文献[34]中, Lyubashevsky 等人提出了一种更简单、有效、直接的方法来满足 $As=t \bmod q$ 的短向量 s 进行知识证明, 该方案通过使用一个多项式乘积证明系统, 并隐藏除一个系数之外的所有系数, 来进行对于两个向量 (或一个向量本身) 模 q 的内积的知识证明, 这样就避免了使用 CRT 嵌入的方式, 提升了实际效率。同时, 该方案是对向量 s 的 2-范数进行界定, 而不是对无穷范数进行界定, 这方便了对其他格原语的构造。在文献[35]中, Lyubashevsky 等人利用文献[1]的零知识证明的构造思路, 构造出基于格的高效的 one-out-of-many 证明系统。在文献[36]中, Nguyen 等人给出了一个实用的基于格的 RICS 的零知识证明系统, 其证明大小为亚线性即与证据大小的平方根渐近成比例。此外, 在文献[37]中, Ghosal 等人利用高效的“MPC in the head”技术构造出一个底层的诚实验证者协议, 结合高效的基于 LWE 的 Correlation Intractable 哈希函数, 应用 Fiat-Shamir 范式得到一个全新的、更高效的用于构造针对 NP 问题的基于 LWE 的 NIZK 论证。结合提高效率为导向的自举 (Bootstrapping) 技术, 可以实现更加高效的基于格的非交互零知识论证。

除了格问题外的研究, 模子集和问题是一类特殊的 NP 完全问题, 被认为是密码学中基于数论的困难假设的有趣的替代方案, 特别地, 模子集和问题被认为可提供抗量子安全性。针对该问题, 经典方案与传统的“MPC in the head”方案仅能实现立方 ($\Theta(\lambda^3)$) 通信复杂度的零知识证明。在文献[39]中, Feneuil 等人通过使用对小整数的秘密共享方案, 并在 MPC in the head 中引入“拒绝”的思路, 结合若干新的技术构造了关于模子集和问题的全新零知识论证方案。该方案大幅降低了通信复杂度, 实现了平方 ($\Theta(\lambda^2)$) 通信量, 可以被实用化。此外, 该

文献中提出的技术还可被用来构造另外几个重要协议/方案,包括证明 ISIS (Inhomogeneous Short Integer Solution) 问题;证明全同态加密的明文/私钥知识;构造高效数字签名方案。关于随机线性编码的校验子解码问题被认为是另一类具备抗量子特性的 NP 完全的困难问题,近年来大量的基于编码的密码协议/方案基于该假设得到构造。在文献[40]中,Feneuil 等人针对关于随机线性编码的校验子解码问题构造了一个全新的零知识证明协议。该协议基于 MPC in the head 范式,将证明校验子解码的低汉明重量问题转换成证明特定多项式的特定关系上。通过将该零知识协议转换成签名方案,文献[40]基于校验子解码问题构造了具有优异签名尺寸的数字签名方案。该方案的表现超过了 Picnic3 方案并能与 SPHINC+方案竞争,后两个方案均是正在进行的 NIST 标准化工作中的抗量子签名方案候选。另外,该方案在“签名尺寸+公钥尺寸”这一参数上的表现优于现有所有基于编码的签名方案。

本节作者: 邓焱、朱旭东、张心轩、齐石 (中国科学院信息工程研究所、中国科学院大学网络空间安全学院)

参考文献

- [1] KIYOSHIMA S. Public-Coin 3-Round Zero-Knowledge from Learning with Errors and Keyless Multi-Collision-Resistant Hash[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 444-473.
- [2] DENG Y, ZHANG X. Knowledge Encryption and Its Applications to Simulatable Protocols with Low Round-Complexity[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 334-362.
- [3] ATTEMA T, FEHR S. Parallel repetition of $(k_1, \dots, k_\mu)$ -special-sound multi-round interactive proofs[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 415-443.
- [4] KIM S, LEE H, SEO J H. Efficient Zero-Knowledge Arguments in Discrete Logarithm Setting: Sublogarithmic Proof or Sublinear Verifier[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 403-433.
- [5] NASSAR S, ROTHBLUM R D. Succinct Interactive Oracle Proofs: Applications and Limitations[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 504-532.

[6] HOLMGREN J, ROTHBLUM R D. Faster Sounder Succinct Arguments and IOPs[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 503-574.

[7] ZENG G, LAI J, HUANG Z, et al. DAG- Σ : A DAG-Based Sigma Protocol for Relations in CNF[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 340-370.

[8] GOEL A, GREEN M, HALL-ANDERSEN M, et al. Stacking sigmas: a framework to compose Σ -protocols for disjunctions[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 520-549.

[9] DON J, FEHR S, MAJENZ C, et al. Efficient NIZKs and Signatures from Commit-and-Open Protocols in the QROM[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 729-757.

[10] KONDI Y, SHELAT A. Improved Straight-Line Extraction in the Random Oracle Model With Applications to Signature Aggregation[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 279-309.

[11] LIBERT B, NGUYEN K, PETERS T, et al. One-Shot Fiat-Shamir-Based NIZK Arguments of Composite Residuosity and Logarithmic-Size Ring Signatures in the Standard Model[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 488-519.

[12] WANG Y Y, PAN J. Non-Interactive Zero-Knowledge Proofs with Fine-Grained Security[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 305-335.

[13] WANG Y Y, PAN J. Unconditionally Secure NIZK in the Fine-Grained Setting[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 437-465.

[14] KIM A, LIANG X, PANDEY O. A New Approach to Efficient Non-Malleable Zero-Knowledge[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 389-418.

[15] GANESH C, ORLANDI C, PANCHOLI M, et al. Fiat-Shamir Bulletproofs are Non-Malleable (in the Algebraic Group Model)[C]. Advances in Cryptology-EUROCRYPT 2022:

EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 397-426.

[16] CANETTI R, SARKAR P, WANG X. Triply Adaptive UC NIZK[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 466-495.

[17] YANG K, WANG X. Non-interactive Zero-Knowledge Proofs to Multiple Verifiers[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 517-546.

[18] MORIMAE T, YAMAKAWA T. Classically Verifiable NIZK for QMA with Preprocessing[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 599-627.

[19] HIROKA T, MORIMAE T, NISHIMAKI R, et al. Certified Everlasting Zero-Knowledge Proof for QMA[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 239-268.

[20] HAITNER I, NUKRAI D, YOGEV E. Lower Bound on SNARGs in the Random Oracle Model[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 97-127.

[21] CHEN M, CHIESA A, SPOONER N. On Succinct Non-interactive Arguments in Relativized Worlds[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 336-366.

[22] HULETT J, JAWALE R, KHURANA D, et al. SNARGs for P from Sub-exponential DDH and QR[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 520-549.

[23] BOOTLE J, CHIESA A, LIU S. Zero-knowledge IOPs with linear-time prover and polylogarithmic-time verifier[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 275-304.

[24] XIE T C, ZHANG Y, SONG D. Orion: Zero knowledge proof with linear prover time[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 299-328.

[25] KOTHAPALLI A, SETTY S, TZIALLA I. Nova: Recursive zero-knowledge arguments from folding schemes[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 359-388.

[26] BOOTLE J, CHIESA A, Hu Y, et al. Gemini: Elastic SNARKs for diverse environments[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 427-457.

[27] WANG N, CHAU S CK. Flashproofs: Efficient Zero-Knowledge Arguments of Range and Polynomial Evaluation with Transparent Setup[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 219-248.

[28] HOUSNI Y E, GUILLEVIC A. Families of SNARK-Friendly 2-Chains of Elliptic Curves [C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 367-396.

[29] ARUN A, BONNEAU J, CLARK J. Short-lived zero-knowledge proofs and signatures [C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 487-516.

[30] CHEN B, DODIS Y, GHOSH E, et al. Rotatable Zero Knowledge Sets[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 547-580.

[31] HOFFMANN C, HUBACEK P, KAMATH C, et al. Practical statistically-sound proofs of exponentiation in any group[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 370-399.

[32] COHEN R, DOERNER J, KONDI Y, et al. Guaranteed Output in $O(n)$ Rounds for Round-Robin Sampling Protocols[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 241-271.

[33] BAUM C, BRAUN L, MUNCK-HANSEN A, et al. $\text{Moz}\mathbb{Z}_{2^k}$ arella : Efficient Vector-OLE and Zero-Knowledge Proofs over $\mathbb{Z}_{2^k}$ [C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 329-358.

[34] LYUBASHEVSKY V, NGUYEN N K, PLANCON M. Lattice-Based Zero-Knowledge Proofs and Applications: Shorter, Simpler, and More General[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer International Publishing, 2022: 71-101.

[35] LYUBASHEVSKY V, NGUYEN N K. BLOOM: Bimodal Lattice One-out-of-Many Proofs and Applications[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer International Publishing, 2022: 95-125.

[36] NGUYEN N K, SEILER G. Practical Sublinear Proofs for R1CS from Lattices[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer International Publishing, 2022: 15-42.

[37] GHOSAL R, LOU P, SAHAI A. Efficient NIZKs from LWE via Polynomial Reconstruction and “MPC in the Head”[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Lecture Notes in Computer Science, vol 13792. Springer, Cham: Springer International Publishing, 2022: 496-521.

[38] ALBRECHT M R, CINI V, LAI R W F, et al. Lattice-Based SNARKs: Publicly Verifiable, Preprocessing, and Recursively Composable[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer International Publishing, 2022: 102-132.

[39] FENEUIL T, MAIRE J, RIVAIN M, et al. Zero-Knowledge Protocols for the Subset Sum Problem from MPC-in-the-Head with Rejection[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 371-402.

[40] FENEUIL T, JOUX A, RIVAIN M. Syndrome Decoding in the Head: Shorter Signatures from Zero-Knowledge Proofs[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 541-572.

3 承诺协议

3.1 承诺协议

承诺 (Commitment) 协议是一个包含两个阶段的两方安全协议, 由承诺方 (Committer) 和接收方 (Receiver) 执行。承诺协议分为承诺阶段 (Committing Phase) 和揭示阶段 (Opening Phase)。在承诺阶段, 承诺方对一个秘密信息 m 做出承诺并得到承诺值 c 和揭示值 d , 然后将承诺值 c 并发送给接收方, 接收方在承诺阶段无法获取有关 m 的任何信息, 这个性质也称为承诺协议的隐藏性质 (Hiding); 在揭示阶段, 承诺方将秘密信息 m 以及揭示值 d 发送给接收方, 接收方可利用揭示值 d 验证该消息 m 与承诺方在承诺阶段所发送的承诺值 c 是否绑定, 任何承诺方都不能发送一个与实际承诺不同并且可通过验证的值 m' , 这个性质也称为承诺方案的绑定性质 (Binding)。承诺协议作为一个基本的密码学原语被广泛用于各类密码学高级协议, 如不经意传输协议、安全多方计算协议、零知识证明系统等。

文献[1]在全局随机谕言机 (Global Random Oracle, GRO) 模型下, 研究全局通用可组合 (Generalized Universal Composable, GUC) 安全的承诺协议。许多可证明安全的协议只能保证协议在单独运行一次时是安全的, 而 GUC 安全的承诺协议能保证在和其他协议一起运行, 或者运行很多次的时候也是安全的。因此, GUC 安全的承诺协议在许多场景下有着重要应用。文献[1]的主要成果是: ①证明了在 Canetti 等人的 GRO 模型 (CCS 2014) 和 Camenisch 等人的 GRO 模型 (欧密会 2018) 下, 都不存在 2 轮 GUC 安全的承诺协议; ②在 Canetti 等人的 GRO 模型下给出了 3 轮 (轮数最优)、GUC 安全、仅需单向函数 (One-Way Function, OWF) 假设的承诺协议。Canetti 等人的协议 (CCS 2014) 和 Mohassel 等人的协议 (欧密会 2017) 都需要离散对数假设, 因此他们的协议都不具备抗量子性, 而文献[1]的协议能够抵抗量子计算机的攻击。文献[1]主要使用了著名的 “MPC in the head” 方法 (STOC 2007) 来构造 Sigma 协议, 该方法的优势便是仅需 OWF 假设。通过精妙地使用 Sigma 协议的各种性质, 文献[1]构造了非交互式见证隐藏 (Non-interactive Witness-Hiding) 协议和含糊承诺 (Equivocal Commitment) 协议, 并以此为构造模块, 最终构造了 GUC 安全的承诺协议。文献[1]的不可行性证明还可以被拓展到零知识证明上, 即在 Canetti 等人的 GRO 模型下不存在一轮 GUC 安全的零知识证明协议。

文献[2]提出了线性映射向量承诺 (Linear-map Vector Commitment, LVC) 的概念, 并研究高效、通用的 LVC 协议。在向量承诺协议中, 承诺方会对一个秘密向量 $\mathbf{v}$ 进行承诺, 并且能在揭示阶段揭示 $\mathbf{v}$ 中任意位置的值。向量承诺协议有意思的地方在于, 揭示阶段所需要的通信量应该是要次线性于 $\mathbf{v}$ 的大小。基于 Lai 和 Malavolta 的工作 (美密会 2019), 文献[2]对向量承诺进行拓展, 使承诺方在揭示阶段所揭示的内容是某个函数在秘密向量 $\mathbf{v}$ 上的运算结果。LVC 的定义包含了所有先前的向量承诺协议的定义, 因此 LVC 是一个更加普适的密码学概念。除此之外, LVC 的定义还包含了可更新 (Updatability)、可聚合 (Aggregation) 和同态 (Homomorphism) 等性质。同时, LVC 协议在可验证数据库 (Verifiable Databases)、无状态数字货币 (Stateless Cryptocurrency)、空间证明 (Proof of Space)、缓存优化 (Caching Optimizations) 等方面都有着广泛的应用前景。除了理论贡献, 文献[2]还做出了能用于实际应用的构造。首先, 文献[2]给出了一种 LVC 的黑盒框架, 即使具有同态性质的 LVC 具备可聚合、可更新的性质。文献[2]提出了两种针对内积函数族的基于双线性配对的 LVC 协议构造, 并且证明了这两种构造都满足同态、可聚合等性质。这类 LVC 承诺协议在零知识证明等领域中有着重要应用。最后, 文献[2]还提出了两种可维护的 LVC 协议构造, 分别是 Srinivasan 等人的协议 (USENIX Security 2022) 和 Tomescu 等人的协议 (IEEE SP 2020) 的拓展。

文献[3]同样研究向量承诺协议, 但聚焦于 Gorbunov 等人所提出的 PointProofs (CCS 2020)。对于向量承诺协议, 在揭示阶段所发送的揭示值往往包含了一个短字符串作为证明,

这个证明用于让接收方验证承诺方所发送的内容是否和承诺阶段所发送的承诺值是绑定的。作为一种新颖的向量承诺协议，PointProofs 可以将不同承诺的证明聚合在一起形成一个新的短字符串作为证明，这能够大大减少存储空间开销，在去中心化场景（如区块链）中有着重要应用。PointProofs 的构造基于 Libert 和 Yung 所提出的向量承诺协议（TCC 2010），并且可以被看作是 Libert 等人所提出的函数承诺协议（ICALP 2016）的重要应用。Gorbunov 等人在证明 PointProofs 的安全性时，用到了 AGM 模型（Fuchsbaauer 等人，美密会 2018）、RO 模型（Bellare 等人，CCS 1993）和 n -wBDHE 假设。然而，AGM 模型作为一种高度理想化的模型，不仅被 Zhang 等人指出在模型构建时存在问题（亚密会 2022），而且难以贴合实际。文献[3]的主要贡献便是绕开了 AGM 模型，在安全性分析中用到了 Bagherzandi 等人所提出的 Generalized Folking 引理（CCS 2008）和 Bellare 等人所提出的 Local Folking 引理（ASIACRYPT 2019），仅使用了 RO 模型和 n -DHE 假设（该假设要弱于 n -wBDHE 假设）就证明了 PointProofs 的安全性。这消除了使用者对 PointProofs 的安全顾虑，极大地增强了人们对于使用 PointProofs 的信心，同时具有现实意义和理论意义。

文献[4]提出并研究了加法同态函数承诺协议，并展示了如何使用加法同态函数承诺协议来构造同态数字签名。函数承诺协议由 Libert 等人提出（ICALP 2016），函数承诺协议同样是让承诺方对秘密向量 $\mathbf{v}$ 进行承诺，但不同于向量承诺协议的是，在揭示阶段承诺方所揭示的是某个函数 f 在该秘密向量 $\mathbf{v}$ 上的计算值 y ，即 $y=f(\mathbf{v})$ 。函数承诺协议有很简单的构造方式，即用一个普通的向量承诺协议加上一个可以证明“ $y=f(\mathbf{v})$ 并且 $\mathbf{v}$ 是所承诺的秘密向量”的简洁非交互式零知识证明（zero-knowledge Succinct Non-interactive ARgument of Knowledge, zkSNARK）系统。但这种构造方式有着致命的缺点：能证明 NP 关系的 zkSNARK 依赖于不可证伪假设（Gentry 等人，STOC 2011），因此用这种方式所构造出的函数承诺协议也将依赖于不可证伪假设。文献[4]的主要贡献便是使用可证伪假设（DHE 假设）来构造具备加法同态性质的函数承诺协议，该协议支持的函数族是常数次数的多变量多项式。而先前的工作并不具备这样的优势：Lai 和 Malavolta 使用了 GGM 模型（美密会 2019），Lipmaa 和 Pavlyk 使用了一种新提出的数学假设（亚密会 2020）。具备加法同态性质的函数承诺协议有许多用途，如构造加法同态的数字签名、构造可更新的函数承诺协议、构造可验证的数据库等。值得注意的是，文献[4]所提出的基于加法同态函数承诺的同态数字签名具备“紧”的特点，即最终的数字签名的大小和函数的定义域/值域的大小无关。而先前的工作，如 Boneh 和 Freeman 的工作（欧密会 2011）、Catalano 等人的工作（美密会 2014）、Gorbunov 等人的工作（STOC 2015）都不具备这样的特点。同时，文献[4]还提出了一个基于可证伪假设、能证明线性关系的 zkSNARK。由于线性关系并不属于 NP 关系，因此这个结果绕开了 Gentry 和 Wichs 的不可行性结果（STOC 2011）。

3.2 承诺协议在零知识证明系统的应用

零知识证明 (Zero-Knowledge Proof, ZKP) 是一种两方安全协议, 证明方可以用其向验证方证明某件事情, 而不泄露其他额外信息。零知识证明的概念由 Goldwasser、Micali 和 Rackoff 提出 (STOC 1985)。早期的零知识证明系统一般都是交互式的, 非交互式零知识证明 (Non-Interactive Zero-Knowledge proof, NIZK) 由 Blum 等人提出 (STOC 1988)。与交互式零知识证明 (Interactive Zero-Knowledge proof, IZK) 相比, NIZK 有许多不同的特性: ①可扩展性不同, IZK 仅支持一对证明方和验证方, 而 NIZK 可支持一个证明方和多个验证方。②易用性不同, IZK 往往要求证明方和验证方同时在线, 而 NIZK 支持证明方可以在验证方离线时发送证明, 验证方上线时再进行验证。③初始化条件不同, NIZK 必须有一个第三方设置 (Setup), 而在交互式中这一条件并不必要。由于 IZK 和 NIZK 的特性不同, 因此导致两类系统有着不同的应用范围。IZK 可用于身份认证 (如 Schnorr 身份验证协议) 等, 而 NIZK 可用于大规模的密码学应用上, 如区块链系统。

NIZK 有许多种构造途径: ①基于同态公钥密码体制, 如 Pinocchio (Gennaro 等人, 欧密会 2013)、Groth16 (Groth, 欧密会 2016)、GM17 (Groth 等人, 美密会 2017) 等。②基于离散对数问题, 如 BCC+16 (欧密会 2016)、BulletProofs (Bunz 等人, IEEE SP 2018) 等。③基于交互式证明, 如 Hyrax (Wahby 等人, IEEE SP 2018)、Libra (Xie 等人, 美密会 2019)、Virgo (Zhang 等人, IEEE SP 2020) 等。④基于安全多方计算, 如 ZKBoo (Giacomelli 等人, USENIX Security 2016)、ZKB++ (Chase 等人, CCS 2017)、Ligero (Ames 等人, CCS 2017) 等。⑤基于信息论层面的证明系统 (Information-Theoretic Proof), 如 LPZK (Dittmer 等人, ITC 2021)、LPZKv2 (Dittmer 等人, CCS 2022) 等。文献[5]的研究对象便是一类新兴的信息论层面的证明系统——交互式谕言机证明 (Interactive Oracle Proof, IOP), IOP 由 Ben-Sasson 等人提出 (TCC 2016)。传统的信息论层面的证明系统, 即概率可检查证明 (Probabilistically Checkable Proof, PCP), 已经被研究者所深刻理解了。Arora 和 Safra 等人发现了著名的 PCP 定理 (FOCS 1992)。简单来说, PCP 定理是指任何属于 NP 关系的语言都能用一个 PCP 系统来证明。这个定理可以由一个简洁的数学公式来表达: $NP = PCP[O(\log_2 n), O(1)]$, 其中 $PCP[O(\log_2 n), O(1)]$ 表示验证方在判断过程中使用至多 $O(\log_2 n)$ 比特的随机数, 并且查询字符串上至多 $O(1)$ 比特的值。研究者自然会对 IOP 的能力感兴趣。Drucker 证明了能用一轮的公开随机数 (Public Coin) 交互式证明系统所判断的语言能有一个一轮的 IOP (ICALP 2011), 但是这个结果无法被拓展到多轮。而文献[5]发现, 任何可以由 k 轮交互式证明系统所判断的语言都能用一个 k 轮、公开随机数 IOP 来证明。这个结果具有崇高的理论价值, 因为这个结果说明了不属于 NP 关系的语言也能够用 IOP 来证明, 而且没有被前人的研究所覆盖。文

献[5]还证明了随机可满足性问题（Stochastic SATisfiability, SSAT）的实例逼近值的困难性。同时,文献[5]提出了一种新的 PCP 变种——位置可解码 PCP(index-decodable PCP, idPCP), 并且展示了如何使用 idPCP 来构造基于 RO 模型的 zkSNARK。Arnon 等人认为可以把 idPCP 看作是经恶意编码的 PCP, 即证明方想要证明的内容包含了若干个经过（恶意）编码的数据点。这种 PCP 可以用于构造一种特殊的 zkSNARK, 即先承诺再证明的 zkSNARK (Commit-and-Prove zkSNARK, CaP-SNARK)。CaP-SNARK 能用于某些之前就被承诺的数据, 或者是被重复使用的承诺。文献[5]的结果同时具备理论意义和实际意义。

文献[6]提出了一种基于双线性配对的、使用通用可更新公共参考字符串（Universal and Updatable Common Reference String）模型的 zkSNARK, Lipmaa 将其称为 Vampire。Vampire 的通信量非常低, 其仅需要 4 个群元素和 2 个域元素, 在通信量的维度上, Vampire 是目前所有使用通用可更新公共参考字符串的 zkSNARK 中最优秀的方案。Vampire 优秀的通信量得益于其使用了一种全新的基于内积证明的 Sumcheck 协议作为构造模块, 而先前的工作 Aurora (Ben-Sasson 等人, 欧密会 2019) 使用的是基于多项式承诺的 Sumcheck 协议作为构造模块。Sumcheck 协议是 Lund 等人在 1992 年引入的交互式证明, Sumcheck 协议可以用于证明某个给定多项式在遍历某个给定域上的计算值之和等于某个给定值, 其在复杂性的概率证明理论以及最近的零知识证明技术中发挥了重要作用。文献[6]将所提出的 Sumcheck 协议称为 Count。Count 的优势有两个方面: ①证明方的计算量是线性输入大小的。②证明大小仅需一个群元素。基于 Count 的优势, Vampire 的计算成本和通信成本都能被降低。值得注意的是, 由于 Sumcheck 协议在密码学中有着广泛的应用, 因此 Count 并不局限于 Vampire, 反而有着宽阔的应用场景。

本节作者: 张秉晟、周哲磊 (浙江大学)

参考文献

- [1] ZHELEI Z, BINGSHENG Z, HONGSHENG Z. GUC-Secure Commitments via Random Oracles: New Impossibility and Feasibility[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 129-158.
- [2] CAMPANELLI M, NITULESCU A, RAFOLS C, et al. Linear-map Vector Commitments and their Practical Applications[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 189-219.

[3] LIBERT B, PASSELEGUE A, RIAHINIA M. PointProofs, Revisited[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 220-246.

[4] CATALANO D, FIORE D, TUCKER I. Additive-Homomorphic Functional Commitments and Applications to Homomorphic Signatures[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 159-188.

[5] ARNON G, CHIESA A, YOGEV E. A PCP Theorem for Interactive Proofs and Applications[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 64-94.

[6] LIPMAA H, SIIM J, ZAJAC M. Counting Vampires: From Univariate Sumcheck to Updatable ZK-SNARK[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 249-278.

4 不经意伪随机函数/不经意传输/秘密共享

不经意伪随机函数 (Oblivious Pseudo Random Function, OPRF) 是一类重要的密码工具, 常被用于设计具有隐私保护的协议。OPRF 是一种两方协议, 其中一方 (发送方/服务器) 拥有一个伪随机函数 (Pseudorandom Function, PRF) 的密钥 k , 另一方 (接收方/客户端) 持有一个输入 x , 最终实现的目标是接收方/客户端得到 PRF 关于 k 和 x 上的输出 $f(k, x)$, 而发送方得不到任何信息。OPRF 通常需要为每个公共输入数据 (或元数据标签) t 使用单独的密钥实例, 这增加了密钥管理的复杂性, 而部分不经意伪随机函数 (Partially-Oblivious PRF, POPRF) 直接提供了降低这个复杂度功能。然而, 最初由 Everspaugh 等人在文献[1]提出的 POPRF 依赖于双线性对, 导致性能显著下降并增加了部署的复杂性。为了克服这些缺点, TYAGI 等人在文献[2]中结合 2HashDH OPRF 和 Dodis-Yampolskiy PRF, 首次构造了不依赖于双线性对的 POPRF 协议——3HashSDHI。该协议在 2HashDH 的基础上, 仅需要额外的哈希计算、群上运算和标量求逆即可在一个 Chaum-Pedersen 证明中计算多个相同的公共元数据标签 t 。特别是文献[2]引入了新的 one-more 类型强 DH 逆假设 (one-more gap Strong Diffie-Hellman Inversion Assumption), 并在代数群模型中证明了该假设的困难性。此外, 文献[2]为 POPRF 提出了新的形式化语法描述和安全性定义, 并在随机谕言机模型中通过将其安全性归约到 one-more 类型强 DH 逆假设上, 证明了该协议的安全性。

不经意传输 (Oblivious Transfer, OT) 协议是隐私计算领域重要技术之一。在许多应用中, 为实现隐私保护目标通常需要执行多个 OT 操作。已有的文献构造了多种批处理 OT 方

法, 但就通信复杂度而言, 即使在现阶段, 接收方和发送方的固有成本也仍然分别需要 n 个比特、 $2n$ 个比特。Brakerski 等人^[3]以及 Gentry 和 Halevi^[4]在文献中表明, 全同态加密 (FHE) 可以实现最优的通信复杂度, 这意味着在 LWE 假设下可以实现最优速率的批处理 OT 设计。Brakerski 等人在文献[5]中证明了在不牺牲计算效率情况下, 通过采用具有 small-inverse-polynomial 噪声的 LPN 假设、DDH 与 QR 假设以及 LWE 假设三者之一的假设, 可以实现最优速率的批处理 OT 的设计。在半诚实模型下, 该协议不需要执行繁重的自举 (Bootstrapping) 操作, 且仅依赖于 $\mathbb{Z}_2$ 上线性同态性质, 即在加密数据上计算线性函数。总体来说, 该协议的效率甚至超过了基于 random-OT 的方法。此外, 文献[5]还将证明结果推广到不经意线性计算 (Oblivious Linear Evaluation, OLE), 其中发送者持有环上的线性函数, 接收者持有该函数的输入, 并且接收者仅仅知道该输入所对应的输出。给定少量的基本不经意传输 (Oblivious Transfer, OT), IKNP 协议 (Ishai 等人, CRYPTO 2003) 可以有效地生成大量扩展 OT。然而, IKNP 协议在每个扩展 OT 上需要传输 λ 比特, 导致通信成本较高。SilentOT 协议^[6]改进了 IKNP 协议, 虽然在计算量上与 IKNP 相当, 但其基于 LPN 假设并没有经过充分的密码分析验证。文献[7]提出了一个 minicrypt 模型下的 SoftSpokenOT 协议。该协议的核心是基于一种称为 $\mathbb{F}_{2^k}$ -VOLE 的 SilentOT 协议, 用于在多项式大小的域上进行向量的不经意线性计算。

相较于 IKNP 协议, SoftSpokenOT 协议在计算成本方面仅需要 λ/k 比特即可实现 $\binom{2}{1}$ -OT 的恶意安全扩展。然而, 在通信成本方面, SoftSpokenOT 协议的发送方需要生成 $\lambda \cdot 2^k/k$ 个伪随机比特, 而 IKNP 协议只需要生成 2λ 比特。渐近地, 当设置 $k = \Theta(\log_2 \ell)$ 时, SoftSpoken-OT 协议可以在多项式时间内生成具有亚线性通信 $\Theta\left(\frac{\lambda \cdot \ell}{\log_2 \ell}\right)$ 的 ℓ 个 OT。这是一个通信-计算的权衡, 对于较小的 k 值, 这种折中是有利的。在 1Gbps 带宽的局域网中, SoftSpokenOT 协议的效率比 IKNP 协议提高了近 5 倍。

分布式点函数 (Distributed Point Function, DPF) 是一种加密原语, 它允许两方或多方之间进行秘密单位向量的压缩加法共享。然而, 现有的 DPF 方案存在一些问题, 如需要两个或更多非共谋方接收秘密函数份额, 以及高计算、通信和协调成本, 因此在单服务器场景下不适用。为解决上述问题, Boyle 等人^[8]提出了一种新的两方 DPF 框架, 其具体贡献如下: ①构造了一种名为可编程 DPF (Programmable DPF, PDPF) 的“1.5-服务器” DPF, 其中离线密钥是标准的伪随机函数 (PRF) 密钥, 而在线密钥仅是长度为域大小 N 的字符串, 并与 PRF 的全域扩展一起形成所需真值表的附加秘密部分。②基于最小的单向函数假设, 提出了适用于多项式大小可行域的两方 2-PDPF, 其中密钥生成、单点计算和全域计算的运行时间相

对于域大小 N 是准线性的, 并且在线密钥大小为 $\text{poly}(\lambda, \log_2 N)$ 。相对于标准 DPF 方案, 2-PDPF 方案具有以下优点: ①在线交互仅涉及从客户端到在线服务器的单个消息, 避免了对两个在线同步服务器的依赖, 减少了网络等待时间, 并且隐藏了离线服务器的身份, 使非共谋假设得以现实。②PDPF 只需要一次通信来传递离线密钥, 并且该密钥可以多次重复使用, 因此在“可信离线 PCG”等应用场景下, 总通信成本得到了接近指数级的优化。该方案还有一些改进的方面: 首先, 它无法提供具有良好具体效率的可忽略隐私误差的 ϵ ; 其次, Gen 和 (单点) Eval 的运行时间与域大小成线性比例 (而不是对数比例); 最后, 它对有效载荷 β 的大小有较差的依赖性 (乘法而非加法)。为了现实环境中实际部署诸多应用程序, 必须解决一个关键障碍是实现某种形式的恶意安全性。可验证分布式点函数 (Verifiable Distributed Point Function, VDPF) 是指支持验证客户端输入格式的密码方案, 以确保客户端无法获取未经授权的服务器数据库信息或以未经授权的方式修改数据库。在已有的方案中, Boyle 等人^[9]构造了半诚实服务器模型下的 VDPF, Boneh 等人^[10]实现了 $\beta \in \{0, 1\}$ 时恶意安全的 VDPF 设计, 但这些方案都需要在服务器之间运行固定大小的安全多方计算协议 (Secure Multiparty Computation, MPC)。另外, 对于具有多个非零点的函数, 服务器需要对每个非零点分别执行 DPF 计算, 这导致计算成本显著增加。为解决上述问题, Castro 等人^[11]提出了一种适用于点函数和多点函数的轻量级可验证两方函数秘密共享方案 (Verifiable Two-Party Function Secret Sharing, FSS)。该方案具有以下贡献: ①该方案仅使用对称加密, 避免了使用公钥加密和多方安全计算技术, 从而在性能、计算和通信复杂度方面优于现有的 VDPF 方案; ②该方案约束条件较少, 仅需验证 DPF 份额是否对应于某个点函数, 而与输出组大小、DPF 输出的结构或需要计算的 DPF 点集无关; ③所构造的 DPF 验证协议具有很强的扩展性, 不仅适用于具有二进制非零输出的 DPF, 还适用于一般的 DPF; ④该方案验证过程可以进行批处理, 从而使得验证多个 DPF 份额所需的通信量与验证一对 DPF 份额的通信量相同。这些优势使得该方案在实际应用中更加高效和灵活。

非交互式可公开验证的秘密共享 (Publicly Verifiable Secret Sharing, PVSS) 方案允许在存在恶意参与方的分布式环境中进行秘密共享。然而, 现有的 PVSS 方案在参与者数量众多 (成千上万) 的情况下效率不高, 需要大量的证明和/或在大型群上进行多次幂运算。为解决这个问题, 文献[12]提出了一种基于 LWE 问题的非交互式 PVSS 方案, 相较于基于格的其他 PVSS 方案, 具有更短的密文和公钥。该方案首先将 Peikert-Vaikuntanathan-Waters (PVW) 加密方案应用于多接收者情形, 使得各参与方的大部分密钥是公共随机串。从而, 产生的摊销明文/密文比率随着参与方数量的增加趋近于 $1/2$, 实现了该比率为 $\Omega(1)$ 的性能。其次该方案在阶为 LWE 模长 (约为 256 比特) 的 DL 群上使用 bulletproof 证明, 紧凑地证明了份额的正确加密/解密。此外, 文献[12]还通过应用类似 Johnson-Lindenstrauss 压缩的技术, 降低了向量的维度, 从而减少了 bulletproof 证明中的指数运算次数。理论上, 在安全参数为 λ 、参

与方数量为 k 的情况下, 每个参与者只需在 $\mathbb{Z}_p$ 群上执行 $O(\lambda+k)$ 次幂运算, 广播 $O(\lambda+k)$ 次标量和 $O(\log_2(\lambda+k))$ 个群元素。这使得方案在性能方面具有很好的可扩展性和效率。

考虑可验证关系共享 (Verifiable Relation Sharing, VRS) 的应用场景, 即客户端 (证明者) 希望在 k 个服务器 (验证者) 之间共享秘密数据向量, 并在零知识证明中证明共享数据满足某些属性。Applebaum 等人^[13]从理论角度对 VRS 进行了研究, 重点关注方案的轮复杂度。Applebaum 等人在文献[14]中指出, 每个可有效计算的关系都可以通过具有两轮最佳轮复杂度的 VRS 方案来实现, 其中第一轮是独立于输入的 (离线轮)。该协议实现了针对活跃敌手的完整 UC 安全性, 敌手被允许破坏包括客户端和一些验证者的任何 t 子集。对于少量 (对数级) 参与方, 该协议实现了最佳弹性门限值 $t < 0.5(k+1)$; 对于大量 (多项式级) 参与方, 协议实现了几乎最优的弹性门限值 $t < 0.5(k+1)(1-\epsilon)$, 其中 $\epsilon > 0$ 是任意小的常量。需要注意的是, 上述两轮解决方案实现了更小的弹性门限值和更弱的安全性概念, 这是在不考虑基本假设前提下的结果。特别地, 文献[14]提出了一个多验证者零知识证明 (Multi-Verifier Zero-Knowledge proof, MVZK) 的 2 轮离线/在线构造, 并表明在诚实多数存在情况下, MVZK 协议可以用更保守的 Minicrypt 假设代替 NIZK 所基于的公钥假设或随机谰言机模型假设。因此, MVZK 协议在诚实多数存在的环境中为 NIZK 提供了全面有效的替代方案。

由于敌手可以在协议执行过程中动态地腐蚀各参与方, 加密协议中实现自适应 (或主动) 安全非常困难。Gentry 等人^[15]在 Benhamouda 等人^[16]研究的基础上尝试在大规模分布式环境 (如区块链) 中构建自适应安全协议的解决方案, 提出 YOSO (You Only Speak Once) 模型, 该模型不是让所有参与方执行整个协议, 而是随机选择较小的匿名委员会来执行协议的每轮。然而, 这仍存在一个巨大的挑战: 如何在不需知道匿名参与方的身份时有效地将消息加密并发送到随机选定的匿名委员, 同时证明加密后的消息对于协议是有效的? 因此, Cascudo 等人^[17]提出了一种新的 YOSO 模型——YOLO YOSO, 用于区块链这样的大规模分布式环境下的密码协议实现自适应安全性。在该模型中, 匿名委员会被随机选择来执行协议的每个单独轮次, 并将协议消息加密发送到下一个匿名委员会, 在发布密文之前删除内部状态。Cascudo 等人^[17]提出了加密协议的秘密状态有效方法, 并引入了两种不同的 PVSS (Publicly Verifiable Secret Sharing) 构造方法。第一种方法是基于具有线性加密和解密性质的加密方案, 被称为 HEPVSS; 第二种方法是名为 DHPVSS 的新型 PVSS 方法, 其中使用了额外的密钥对进行共享分配, 并结合了 SCRAPE 方案中使用的技术, 证明了加密的共享份额在适合次数的多项式上。

抗泄露秘密共享 (Leakage Resilient Secret Sharing, LRSS) 保证在给定未授权的集合和泄露其他诚实份额的情况下, 敌手仍无法获取关于秘密的任何信息。而不可延展秘密共享 (Non-Malleable Secret Sharing, NMSS) 保证在消息被篡改的情况下, 敌手恢复出的消息仍然是随机的。这两种方案中最重要的参数是份额的大小。Srinivasam 和 Vasudevan^[18]提出了具

有门限访问结构的 LRSS 方案, 其份额大小约为 $(3 \cdot \text{message length} + \mu)$ (其中 μ 是每个份额可容忍的泄露比特数), 以及份额大小为 $(11 \cdot \text{message length})$ 的 NMSS 方案。然而, 该方案^[19]显著提高了 LRSS 和 NMSS 方案的份额大小, 并实现了最佳的 LRSS 份额和泄露大小。特别地, Chandran 等人^[19]进一步构造了一个门限访问结构的信息论 LRSS 方案, 其份额大小为 $(\text{message length} + \mu)$ 。随后, 结合了 NMSS 编译器中的 1/3-rate non-malleable 代码, 提出了一个份额大小为 $(4 \cdot \text{message length})$ 的 NMSS 方案。对于共享随机消息的特殊情况, 该方案份额大小仅为 $(2 \cdot \text{message length} + \mu)$ 。

本节作者: 林昌露、周鑫、黄可可 (福建师范大学)

参考文献

[1] EVERSIPAUGH A, CHATERJEE R, SCOTT S, et al. The Pythia PRF Service[C]. 24th USENIX Security Symposium (USENIX Security 15). 2015: 547-562.

[2] TYAGI N, CELI S, RISTENPART T, et al. A fast and simple partially oblivious PRF, with applications. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 674-705.

[3] BRAKERSKI Z, DOTTLING N, GARG S, et al. Leveraging linear decryption: Rate-1 fully-homomorphic encryption and time-lock puzzles[C]. Theory of Cryptography: 17th International Conference, TCC 2019, Nuremberg, Germany, December 1-5, 2019, Proceedings, Part II. Cham: Springer International Publishing, 2019: 407-437.

[4] GENTRY C, HALEVI S. Compressible FHE with applications to PIR[C]. Theory of Cryptography: 17th International Conference, TCC 2019, Nuremberg, Germany, December 1-5, 2019, Proceedings, Part II. Cham: Springer International Publishing, 2019: 438-464.

[5] BRAKERSKI Z, BRANCO P, DOTTLING N, et al. Batch-OT with Optimal Rate[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 157-186.

[6] BOYLE E, COUTEAU G, GILBOA N, et al. Efficient pseudorandom correlation generators: Silent OT extension and more[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 489-518.

[7] ROY L. SoftSpokenOT: Quieter OT extension from small-field silent VOLE in the minicrypt model[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 464-479.

- [8] BOYLE E, GILBOA N, ISHAI Y, et al. Programmable distributed point functions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 121-151.
- [9] BOYLE E, GILBOA N, ISHAI Y. Function secret sharing: Improvements and extensions [C]. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. 2016: 1292-1303.
- [10] BONEH D, BOYLE E, CORRIGAN-GIBBS H, et al. Lightweight techniques for private heavy hitters[C]. 2021 IEEE Symposium on Security and Privacy (SP). IEEE, 2021: 762-776.
- [11] CASTRO L, POLYCHRONIADOU A. Lightweight, maliciously secure verifiable function secret sharing[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 150-179.
- [12] GENTRY C, HALEVI S, LYUBASHEVSKY V. Practical non-interactive publicly verifiable secret sharing with thousands of parties[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 458-487.
- [13] APPLEBAUM B, KACHLON E, PATRA A. The resiliency of MPC with low interaction: the benefit of making errors[C]. Theory of Cryptography: 18th International Conference, TCC 2020, Durham, NC, USA, November 16-19, 2020, Proceedings, Part II 18. Springer International Publishing, 2020: 562-594.
- [14] APPLEBAUM B, KACHLON E, PATRA A. Verifiable relation sharing and multi-verifier zero-knowledge in two rounds: trading NIZKs with honest majority[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 33-56.
- [15] GENTRY C, HALEVI S, KRAWCZYK H, et al. YOSO: You Only Speak Once: Secure MPC with Stateless Ephemeral Roles[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 64-93.
- [16] BENHAMOUDA F, GENTRY C, GORBUNOV S, et al. Can a public blockchain keep a secret[C]. Theory of Cryptography: 18th International Conference, TCC 2020, Durham, NC, USA, November 16-19, 2020, Proceedings, Part I 18. Springer International Publishing, 2020: 260-290.
- [17] CASCUDO I, DAVID B, GARMS L, et al. YOLO YOSO: fast and simple encryption and secret sharing in the YOSO model[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 651-680.

[18] SRINIVASAM A, VASUDEVAN P N. Leakage resilient secret sharing and applications[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 480-509.

[19] CHANDRAN N, KANUKURTHI B, OBBATTU S L B, et al. Short Leakage Resilient and Non-malleable Secret Sharing Schemes[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 178-207.

5 安全多方计算

在 2022 年国际三大密码年会中,安全多方计算(Secure Multi-Party Computation, SMPC)仍然是研究的热点内容,其中 CRYPTO 包含 4 个相关专题,收录 10 篇论文;EUROCRYPT 包含 1 个相关专题,收录 8 篇论文;ASIACRYPT 包含 1 个相关专题,收录 4 篇论文。从研究内容看,主要涉及 5 个方面:安全多方计算基础理论;电路优化;轮复杂度优化;特定协议;其他研究。

5.1 安全多方计算基础理论

针对安全多方计算基础理论的研究主要包括安全模型、安全性定义、可行性结论及通用编译器构造等。

针对安全多方计算中非诚实方大多数安全性的研究,Escudero 等人在 $\mathbb{Z}_{2^k}$ 类型的环上,提出了一个非诚实方大多数场景下实现主动安全的多方计算协议^[1]。他们没有考虑 Cramer 等人在 2018 年美密会上提出的 $\text{SPD}\mathbb{Z}_{2^k}$ 及其衍生的 $\mathbb{Z}_{2^{k+s}}$ 形式的“扩展环”^[2],而是使用伽罗瓦环的扩展 $\mathbb{Z}_{p^k}[X](h(X))$,该扩展具有足够大的阶,以保证敌手仅能以可忽略的概率作弊。具体来说,他们利用 Cramer 等人在 2021 年美密会上提出的在 $\mathbb{Z}_{p^k}$ 上的反向乘法友好嵌入 (Reverse Multiplication-Friendly Embeddings, RMFE)^[3],以及 Abspoel 等人在诚实方大多数场景中引入的对 RMFE 的优化^[4],实现了一个高效的协议,该协议只需要在其在线阶段分摊 $12.4k(n-1)$ 比特的通信量和每个乘法门的一轮通信。此外,他们还通过推广 CCS 2016 上 Keller 等人在 MASCOT 中提出的基于茫然传输 (Oblivious Transfer, OT) 的方法,使用茫然线性计算 (Oblivious Linear Evaluation, OLE) 实例化了必要的离线阶段。与文献[1]不同,文献[5]在非诚实方大多数的预处理模型中构建了第一个 SMPC 协议,该协议每个门的通信复杂度与参

与方数 n 呈亚线性关系。他们将打包秘密共享 (Packed Secret Sharing) 的使用扩展到非诚实方大多数场景。对于被腐化方的比例, 如果 99% 的参与方是被腐化的, 仍然可以实现所有参与方的每个乘法门的通信复杂度为 $O(1)$ 个域元素。

针对安全多方计算中的动态性问题, Choudhuri 等人在 2021 年美密会上提出的 Fluid MPC 模型允许各方在资源空闲时灵活地参与协议^[6], 但是其仅考虑诚实方大多数场景。文献[7]将其扩展到大多数参与者在计算中的任何一点都可能被腐化这种场景, 并通过提供支持动态参与者的 SPDZ 协议的变体来实现上述结论。

Applebaum 等人在 2021 年提出的多方随机编码 (Multiparty Randomized Encoding) 将安全计算复杂的多方功能函数 f 的任务归约到安全计算更简单的功能函数 g 的任务。作为一个特例, 2 阶多方随机编码 (Quadratic Multiparty Randomized Encoding, 2MPRE) 在 SMPC 研究中被发现有诸多应用, 然而其考虑的安全模型均为诚实方大多数。文献[8]打破了诚实方大多数假设, 提出了第一个在非诚实方大多数场景下安全的通用 2MPRE 构造。其构造通过 2MPRE 编码每个 n 方功能函数 f , 且该 2MPRE 最多允许被动腐化 $t = \lfloor 2n/3 \rfloor$ 个参与方。此外, 他们得到以下结果: ①一个非交互式客户端-服务器 SMPC 协议, 该协议具有完美的隐私性, 可以对抗少数服务器与 n 个客户端中最多 t 个的合谋。②非交互式 t -private 归约下三方功能函数的完备性。③从通用 SMPC 到理想 OT 的单轮 t -private 归约。

针对安全多方计算中的分支计算问题, 已有的技术会导致通信成本与整个电路尺寸呈线性关系。为了缓解这个问题, 最近的一系列工作研究了减少 SMPC 内部分支执行的通信成本的问题, 其中堆叠混乱范式 (Stacked Garbling Paradigm)^[9]提出分支电路的混乱电路, 其大小仅取决于最大分支的大小。然而, 目前还不知道如何为涉及两方以上的安全计算获得类似的通信改进。文献[10]为分支多方计算提供了一个通用框架, 支持任意数量的参与方, 其方案的通信复杂度与最大分支的大小成正比, 整个电路的计算是线性的。

在安全多方计算中, 随机复杂度的最小化问题一直以来都是一个热点话题, Kushilevitz 和 Mansour 针对异或函数研究了满足安全阈值的多方安全计算, 表明在该情况下, $O(t^2 \log_2(n/t))$ 个随机比特足以满足安全性要求, 并表明 $\Omega(t)$ 为随机复杂度的下界。文献[11]通过证明异或随机复杂度为 $\Omega(t^2)$ 的下界来重新审视了这个问题, 将之前的上界匹配到一个对数因子 (或当 $t = \Omega(n)$ 时常数因子)。此外, 还获得了一个使用 $O(t^2 \log_2 n)$ 个随机比特的显式协议, 将下界匹配到一个多对数因子。他们将这些结果从异或推广到一般对称布尔函数和有限阿贝尔群上的加法, 并展示了如何摊销多重加法的随机性复杂度。最后, 他们将技术与最近的隐私电路的随机高效结构相结合, 通过利用不提供任何输入的额外“辅助参与方”, 获得了仅使用 $O(t^2 \log_2 |C|)$ 个随机比特计算通用电路 C 的显式协议。

针对公平性和输出可达性(Guaranteed Output Delivery, GOD)的研究,文献[12]针对 Alon 等人^[13]在 2020 年美密会上提出的 FaF (Friends and Foes) 安全的概念,证明了半诚实茫然传输对具有最优弹性的 FaF 安全协议的必要性。在实际应用方面,提出了一种基于环的 4PC (四方) 协议 QuadSquad, 该协议在 FaF 模型中实现了公平性和输出可达性,允许最优腐化 1 个恶意方和 1 个半诚实方。

在通用编译器研究中,针对大规模并行计算 (Massive Parallel Computing), 已知的编译器不能满足恶意安全性,即攻击者可以任意偏离规定的协议。文献[14]在非诚实方大多数场景中为恶意攻击者设计第一个通用编译器,由此产生的协议可以实现除一个之外 (all-but-one) 的所有腐化。该编译器依赖于一个简单的 PKI 和一个 (可编程的) 随机谕言机,并且在假设 LWE 和 SNARK 的情况下被证明是安全的。

此外,最小化 SMPC 轮复杂度的典型方法是以增加通信复杂度或依赖假设为代价的。一个值得注意的例外是, Ananth 等人最近的工作,它使用 FE (Functional Encryption) 组合器在一般模型中构造了轮最优 (2 轮) 半诚实安全的 SMPC 协议,其通信复杂度与正在计算的电路的深度和输入输出长度成比例。然而,如何在一般模型设计抵抗恶意敌手安全的通信高效协议,是一个开放性问题。文献[15]提供了一个轮最优的黑盒编译器,假设 FE 组合器存在,其可以将各种各样的 SMPC 协议在普通模型中编译成电路可扩展的恶意安全 SMPC 协议。他们还提供了一个轮最优的黑盒编译器,假设多密钥全同态加密方案存在,其可以在一般模型中将各种各样的 SMPC 协议编译成电路无关的恶意安全 SMPC 协议。

5.2 电路优化

任意功能函数的安全两方计算均可以通过使用混淆电路 (Garbled Circuit, GC) 来实现,然而,计算开销和通信开销一直以来是混淆电路的瓶颈,如何提升混淆电路的通信效率和计算效率一直备受关注。对于包含条件分支的功能函数,使用堆叠混淆电路 (Stacked Garbled Circuit, SGC) 可以突破通信瓶颈,达到与电路尺寸呈亚线性关系的通信复杂度。这种技术虽然能优化通信开销,但是会导致计算开销增大。因此,文献[16]提出了一种名为 GCWise 方法可以将电路计算方 E 的计算开销降到亚线性,并将 GCWise 应用到一个非常有吸引力的应用场景即混淆隐私信息检索 (Garbled Private Information Retrieval, Garbled PIR), Garbled PIR 允许混淆电路从一个公开的数据库中检索一个隐私元素,然后继续在混淆电路中使用这个元素,利用本文方法可以实现与元素个数呈亚线性关系的通信复杂度和计算复杂度。

此外,对于满足恶意安全性的常数轮两方计算问题,目前最先进的协议是 Katz 等人在 2018 年 CRYPTO 会议上提出的认证混淆功能函数 (Authenticated Garbling Functionality) 的一种变体^[17]。然而,实现认证混淆需要高昂的通信成本,即便是最先进的工作也需要相当于

10 个混淆电路的通信成本和两轮通信。为了解决这个问题,文献[18]利用向量茫然线性计算 (Vector Oblivious Linear Evaluation, VOLE) 的变体、乘法三元组的变体等简单随机关联大幅降低了通信成本。具体而言,通过同时使用 VOLE 的变体和乘法三元组的变体,该工作可以将通信成本降到与 1.31 个混淆电路的成本相当;仅使用 VOLE 的变体,该工作可以将通信成本降到与 2.25 个混淆电路的成本相当;仅使用乘法三元组的变体,该工作实现了非交互协议,该协议的通信成本与 8 个混淆电路的成本相当。此外,本文指出利用 Boyle 等人^[19-20]提出的简单相关伪随机数生成器可以安全地生成该协议所需的简单随机关联,并且不会产生效率瓶颈。

5.3 轮复杂度优化

近些年,安全多方计算的轮复杂度研究成为密码学术界的主流方向之一。为了使安全多方计算的轮复杂度最小化,文献[21]在随机预言机模型 (ROM) 中和 1-out-of-2 OT correlations 模型中,给出了几种轮数最优的黑盒编译器,利用这些编译器,该文献首次提出了一个能在随机预言机模型中抵抗恶意敌手的两轮两方协议,在此前,没有任何工作能做到这一点。在该模型下,该文献还提出了一种抵抗恶意敌手的 3 轮多方安全协议。该协议基于最小的密码学困难假设,且达到了目前已知的轮复杂度下限。除了随机预言机模型,该文献提出了一个能在 1-out-of-2 OT correlations 模型中抵抗恶意敌手的两轮多方安全计算协议,这为算术分支程序提供了一种新的轮最优的协议。

除上述工作对两方安全计算的轮复杂度的研究之外,在非诚实方大多数场景中,最近的工作在构造轮数最优的两轮多方计算协议方面也取得了令人兴奋的进展。然而,到目前为止,大部分的方案仍然是复杂和低效的。文献[22]利用随机预言机和张量不经意线性计算 (Tensor Oblivious Linear Evaluation, Tensor OLE) 提出了一个两轮 SMPC 协议,该协议满足非诚实方大多数场景的恶意安全性。该工作的主要技术贡献在于使用了简单的、轻量级的检查来确保恶意安全,因此该协议的计算开销与 2020 年 TCC 会议上 Lin 等人半诚实安全的协议接近,相比之前的工作,该协议优化了轮数和计算成本,更加简单和高效。利用该协议,该文献为布尔电路的安全计算提出了一个满足恶意安全性的两轮 SMPC 协议,该协议计算复杂度为 $O(s \cdot n^3 \cdot \log_2 |F|)$, 其中 F 是一个特征二 (Characteristic Two) 的域,同时为算术分支程序的安全计算提出了一个满足恶意安全的两轮 SMPC 协议,该协议计算复杂度为 $O(s^{1.5} \cdot n^3 \cdot \log_2 |F|)$, 其中 F 是一个一般域。

在诚实方大多数场景中,可识别中止安全是一种重要的增强型概念,即敌手的恶意行为会引发协议中止,同时,所有的诚实参与方能确认引起中止的参与方的身份。文献[23]在标准和通用的多项式时间假设下,实现了4轮通信的可识别中止安全性。为此,作者引入了一种名为 Bounded-rewind secure MPC 的新概念,在这种安全性定义下,即使敌手可以进行一种“轻度重置攻击”(Mild Form of Reset Attacks),也能保证安全性。此外,还展示了如何在陷门置换假设下,通过任意 SMPC 协议来实例化 Bounded-rewind secure MPC 这个原语。这一概念使 SMPC 协议与其他(交互式)加密原语的并行组合更加容易,因此作者指出,在其他场景中也会有 Bounded-rewind secure MPC 的用武之地,尤其是那些需要结合多种原语且对轮复杂度有要求的 SMPC 协议。

在非交互式安全多方计算研究方面,文献[24]在满足超多项式时间模拟(Super-Polynomialtime Simulation, SPS)概念的 plain model 中研究了非交互式安全两方计算的可行性,并首次提出了可并发组合的满足超多项式时间模拟的非交互式安全两方计算,并且在 plain model 中具有完美的正确性。

5.4 特定协议

在 SMPC 中,掷硬币协议是一种重要的协议,它可以让多个互不信任的参与方在没有可信方的情况下,通过交互的方式生成一个公共的随机比特。然而,1986年,著名的 Cleve 下限指出,在双方掷币协议中不可能达到“事实强公平性”的概念。也就是说,腐化方总是有一种策略,使诚实方的结果向敌手偏好的方向出现严重偏移。不过,早在1981年美密会上,Blum 提出的掷币协议实现了严格意义上较弱的“博弈论”公平概念。该概念要求协议能够形成一个抗合谋的纳什均衡,也就是说,任何合谋联盟都不应该有违反诚实行为的动机。在2018年 TCC 会议上,Chung 等人将这一博弈论概念表述为“合作策略保障”(Cooperative-Strategy-Proofness),简称“CSP 公平”。文献[25]探讨了在非诚实方大多数场景下,满足博弈论公平性的 n 方掷币协议的可行性,并指出当合谋联盟的规模多于半数但不是 $n-1$ 时,在一些有意义的参数条件下这种协议确实是可行的,还提供了一种在这些条件下实现 CSP 公平的掷币协议。此外,作者提供了与之相关的上界和下界,并以此给出了满足 CSP 公平掷币的条件的完整表征定理,该定理显示了事实上的强公平和博弈论公平概念之间的数学结构差异。

乘法协议是许多安全多方计算协议的基石,研究者一直没有停止对两方安全乘法协议的研究,文献[26]提出了一个新的基于 OT 的两方乘法协议,其效率几乎与1999年美密会上 Gilboa^[27]的半诚实协议相同,但在无须进一步编译的情况下,对恶意对手也具有高度的安全性,所取得的安全性足以满足许多应用的需求,而且基于 DDH 假设可以廉价地编译成完全安全的抗恶意敌手的协议。

隐私集合运算 (Private Set Intersection, PSI) 在 2022 年取得了突破性进展, 文献[28]利用函数秘密分享 (Function Secret Sharing, FSS) 技术提出了一种新型 PSI 定义并设计了协议, 称为 Structure-Aware PSI。在这种 PSI 场景中, Alice 的隐私集合元素具有某种特定的结构, 且这种结构是公开的; Bob 的隐私集合没有结构, 最终双方获得交集元素。与传统 PSI 相比, 这种新型 PSI 可以利用 Alice 的公开结构来降低通信成本。此外, 作者还利用 Structure-Aware PSI 协议设计了一种高效的模糊 PSI 协议。

5.5 其他研究

Garbled RAM (GRAM) 是 Ostrovsky 等人提出的一项强大的技术, 它在不增加交互轮数的情况下将一个亚线性成本的 RAM 并入 Garbled Circuit (GC) 中。尽管目前已经有了很多 GRAM 构造, 但由于成本较高且扩展性差, 因此没有一种适用于现实场景。文献[29]提出了第一个实用的 GRAM, 该方法优于基于线性扫描的 RAM。对于计算安全参数 κ 和一个存储尺寸为 $w = \Omega(\log_2 n)$ 比特的块的 size- n RAM, 该 GRAM 每次访问都会产生摊销的 $O(w \cdot \log_2 n \cdot \kappa)$ 的通信和计算复杂度。

信息论安全的多方隐私计算 (Multi-party Information-theoretic Private Computation) 本质上需要各参与方使用局部随机性, 而最小化隐私计算所需的随机比特总数的问题受到广泛关注。文献[30]主要研究了在一个隐私计算中有多少个参与者需要获得随机源, 以及有多少人无法访问随机源的问题。此外, 还讨论了对系统中随机源的数量与计算所需的随机比特总数之间可能存在的相互作用。本文指出, 当 n 个参与方计算任何确定性函数时, 具有单个随机源访问权限的 t ($t < n/2$) 个参与方足以进行信息论安全的多方计算。但是, 对于随机函数, 至少需要 $t+1$ 个随机源。本文还提供了一种改进的协议, 仅使用一个随机源和该源随机抛出的 6 比特, 以实现多方之间的 1-private AND 函数计算。这比以前已知的需要两个源和 8 比特的方法要好。

针对拓扑隐藏计算 (Topology-Hiding Computation, THC), 文献[31]主要考虑无向环上的 THC, 但也给出了通用图上的 THC 的两个结果, 其所有的结果都是在被动敌手静态地腐化任意数量的参与方的情况下得出的。具体来说, 在无向环中, 最先进的拓扑隐藏广播 (Topology-Hiding Broadcast, THB) 协议是 Akavia 等人在 2017 年欧密会上提出的 Akavia-Moran (AM) 协议^[32]。他们对 AM 协议进行了优化, 使 $O(\kappa)$ 比特广播的通信成本从 $O(n^2 \kappa^2)$ 比特减少到 $O(n^2 \kappa)$ 比特。此外, 他们还考虑求和及通用计算功能, 并给出更为高效的 THC 协议, 其通信成本降低了 $1/O(n\kappa)$ 。对于通用图的 THC, 最先进的 THB 协议是 Akavia 等人的 ALM (Akavia-LaVigne-Moran) 协议^[33]。文献对 AM 协议的优化也适用于 ALM 协议, 并且可以将其通信成本降低了 $1/O(\kappa)$ 。此外, 他们还优化了 Lavigne 等人在 2018 年的 TCC 中提出的基

于全同态加密 (Fully-Homomorphic Encryption, FHE) 的通用拓扑隐藏计算 (General Topology-Hiding Computation, GTHC) 协议, 并将其通信成本从 $O(n^8\kappa^2)$ 个 FHE 密文和 $O(n^5\kappa)$ 个 FHE 公钥降低到 $O(n^6\kappa)$ 个 FHE 密文和 $O(n^5\kappa)$ 个 FHE 公钥。

此外, 文献[34]提出了一个通用的群论框架, 统一了此前的同态秘密共享 (Homomorphic Secret Sharing, HSS) 和伪随机相关函数 (Pseudorandom Correlation Functions, PCF) 构造在不同群中计算分布式离散对数的方法, 还将现有的构造转换到他们提出的框架中, 并提出了第一个基于类群假设的分支程序的两方同态秘密共享的构造。通过这个框架, 还可以得到基于数论假设生成 OT 和 VOLE 相关性的 PCF。使用合适的零知识证明, 可以改进所提出的方案并实现主动安全, 由此构造了在随机预言机模型中用于建立 PCF 的一轮主动安全协议, 以及一个基于 HSS 的 3 轮主动安全协议, 该协议是针对分支程序的安全两方计算, 且具有一套简洁的通信系统。

本节作者: 王皓、魏晓超、李智、牛梓宇 (山东师范大学)

参考文献

- [1] ESCUDERO D, XING C, YUAN C. More Efficient Dishonest Majority Secure Computation over $\mathbb{Z}_{2^k}$ via Galois Rings[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 383-412.
- [2] CRAMER R, DAMGARD I, ESCUDERO D, et al. SPD $\mathbb{Z}_{2^k}$: Efficient MPC mod 2^k for Dishonest Majority[C]. Advances in Cryptology-CRYPTO 2018: CRYPTO 2018, Santa Barbara, CA, USA, August 18-24, 2018. Cham: Springer Nature Switzerland, 2018: 769-798.
- [3] CRAMER R J F, RAMBAUD M, XING C. Asymptotically-Good Arithmetic Secret Sharing over $\mathbb{Z} / p^l\mathbb{Z}$ with Strong Multiplication and Its Applications to Efficient MPC[C]. Advances in Cryptology-CRYPTO 2021: CRYPTO 2021, Santa Barbara, CA, USA, August 16-20, 2021. Cham: Springer Nature Switzerland, 2021: 656-686.
- [4] ABSPOEL M, CRAMER R, ESCUDERO D, DAMGARD I, XING C. Improved Single-Round Secure Multiplication Using Regenerating Codes[C]. Advances in Cryptology-ASIACRYPT 2021: ASIACRYPT 2021, Taipei, Taiwan, December 6-10, 2021. Cham: Springer Nature Switzerland, 2021: 222-244.
- [5] GOYAL V, POLYCHRONIADOU A, SONG Y. Sharing Transformation and Dishonest Majority MPC with Packed Secret Sharing[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO

2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 3-32.

[6] CHOUDHURI A R, GOEL A, GREEN M, et al. Fluid MPC: Secure Multiparty Computation with Dynamic Participants[C]. Advances in Cryptology-CRYPTO 2021: CRYPTO 2021, Santa Barbara, CA, USA, August 16-20, 2021. Cham: Springer Nature Switzerland, 2021: 94-123.

[7] RACHURI R, SCHOLL P. Le Mans: Dynamic and Fluid MPC for Dishonest Majority[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 719-749.

[8] APPLEBAUM B, ISHAI Y, KARNI O, et al. Quadratic Multiparty Randomized Encodings Beyond Honest Majority and Their Applications[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 453-482.

[9] HEATH D, KOLESNIKOV V. Stacked Garbling-Garbled Circuit Proportional to Longest Execution Path[C]. Advances in Cryptology-CRYPTO 2020: CRYPTO 2020, Santa Barbara, CA, USA, August 17-21, 2020. Cham: Springer Nature Switzerland, 2020: 763-792.

[10] GOEL A, HALL-ANDERSEN M, HEGDE A, et al. Secure Multiparty Computation with Free Branching[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 397-426.

[11] GOYAL V, ISHAI Y, SONG Y. Tight Bounds on the Randomness Complexity of Secure Multiparty Computation[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 483-513.

[12] HEGDE A, KOTI N. Attaining GOD Beyond Honest Majority with Friends and Foes[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 556-587.

[13] ALON B, OMRI E, PASKIN-CHERNIAVSKY A. MPC with Friends and Foes[C]. Advances in Cryptology-CRYPTO 2020: CRYPTO 2020, Santa Barbara, CA, USA, August 17-21, 2020. Cham: Springer Nature Switzerland, 2020: 677-706.

[14] FERNANDO R, GELLES Y, KOMARGODSKI I, et al. Maliciously Secure Massively Parallel Computation for All-but-One Corruptions[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 688-718.

- [15] CIAMPI M, OSTROVSKY R, WALDNER H, et al. Round-Optimal and Communication-Efficient Multiparty Computation[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 65-95.
- [16] HAQUE A, HEATH D. Garbled Circuits with Sublinear Evaluator[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 37-64.
- [17] KATZ J, RANELLUCCI S, ROSULEK M, et al. Optimizing Authenticated Garbling for Faster Secure Two-Party Computation[C]. Advances in Cryptology-CRYPTO 2018: CRYPTO 2018, Santa Barbara, CA, USA, August 18-24, 2018. Cham: Springer Nature Switzerland, 2018: 365-391.
- [18] DITTMER S, ISHAI Y. Authenticated Garbling from Simple Correlations[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 57-87.
- [19] BOYLE E, COUTEAU G. Efficient Pseudorandom Correlation Generators: Silent OT Extension and More[C]. Advances in Cryptology-CRYPTO 2019: CRYPTO 2019, Santa Barbara, CA, USA, August 15-18, 2019. Cham: Springer Nature Switzerland, 2019: 489-518.
- [20] BOYLE E, COUTEAU G. Efficient Pseudorandom Correlation Generators from Ring-LPN[C]. Advances in Cryptology-CRYPTO 2020: CRYPTO 2020, Santa Barbara, CA, USA, August 17-21, 2020. Cham: Springer Nature Switzerland, 2020: 387-416.
- [21] ISHAI Y, KHURANA D. Round-Optimal Black-Box Protocol Compilers[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 210-240.
- [22] LIN H, LIU T. Two-Round MPC Without Round Collapsing Revisited - Towards Efficient Malicious Protocols[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 353-382.
- [23] CIAMPI M, RAVI D. Round-Optimal Multi-party Computation with Identifiable Abort[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 335-364.
- [24] MORGAN A, PASS R. Concurrently Composible Non-interactive Secure Computation[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 526-555.

- [25] WU K, ASHAROV G, SHI E. A Complete Characterization of Game-Theoretically Fair, Multi-Party Coin Toss[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 120-149.
- [26] HAITNER I, MAKRIYANNIS N. Highly Efficient OT-Based Multiplication Protocols [C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 180-209.
- [27] GILBOA N. Two Party RSA Key Generation[C]. Advances in Cryptology-CRYPTO 1999: CRYPTO 1999, Santa Barbara, CA, USA, August 15-19, 1999. Cham: Springer Nature Switzerland, 1999: 116-129.
- [28] GARIMELLA G, ROSULEK M, SINGH J. Structure-Aware Private Set Intersection, with Applications to Fuzzy Matching[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 323-352.
- [29] HEATH D, KOLESNIKOV V. OSTROVSKY R. EpiGRAM: Practical Garbled RAM[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 3-33.
- [30] COUTEAU G, ROSEN A. Random Sources in Private Computation[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 443-473.
- [31] LI S. Towards Practical Topology-Hiding Computation[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 588-617.
- [32] AKAVIA A, MORAN T. Topology-Hiding Computation Beyond Logarithmic Diameter[C]. Advances in Cryptology-EUROCRYPT 2017: EUROCRYPT 2017, Paris, France, April 30 - May 4, 2017. Cham: Springer International Publishing, 2017: 609-637.
- [33] AKAVIA A, LAVIGNE R, MORAN T. Topology-Hiding Computation on All Graphs[C]. Advances in Cryptology-CRYPTO 2017: CRYPTO 2017, Santa Barbara, CA, USA, August 20-24, 2017. Cham: Springer Nature Switzerland, 2017: 447-467.
- [34] ABRAM D, DAMGARD I. An Algebraic Framework for Silent Preprocessing with Trustless Setup and Active Security[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 421-452.

应用密码

1 区块链/数字货币

近年来,区块链与数字货币一直得到密码社区的广泛关注。在 2022 年国际三大密码年会上共有 7 篇相关论文,其中研究方向包括交易支付、共识机制、安全泛洪等,而 FSE、CHES 上暂无相关研究。

区块链交易支付一直是研究的热点话题。Mimblewimble 是著名的数字货币协议,它不仅能够为用户提供隐私保护,而且能够克服区块链可扩展性差的问题。然而,原始 Mimblewimble 协议要求交易发送者必须在线、无法支持非交互式交易,这严重阻碍了其在金融等领域的应用。在文献[1]中,Fuchsbaauer 和 Orru 详细分析了原始面向 Mimblewimble 的非交互式交易方案,提出了新的改进方案,提供了在文献[2]安全模型下的严格安全证明,他们的方案可以视为 MimbleWimbleCoin 的安全实现以及莱特币(Litecoin)中 MimbleWimble Extension Blocks (MWEB)的变体。在文献[3]中,Jia 等人基于原始 Mimblewimble 协议,提出了一种新型非交互式聚合现金系统 NiACS。她们首次提出了面向 NiACS 基于模拟的安全定义和对应的理想函数,构造了通用可组合性(Universal Composition, UC)模型下实现理想函数的 NiACS 协议。另外,他们还提出了一种 ElGamal 加密算法的变体形式,并给出了 NiACS 协议的具体实例。

区块链共识算法的研究主要集中在如何降低存储和能源开销。区块链的成功使得所有全节点需要存储不停增长的数据账本。与此相反,轻节点仅需存储少部分区块链数据,并依赖于全节点的交互进而获得账本信息。然而,如何使这种交互无信任,一直是区块链研究的热点。在文献[4]中,Abusalah 等人从经典的证明系统理论出发,设计了轻量级客户端的区块链协议。他们基于连续工作量证明机制(Proofs of Sequential Work, PoSW),提出了简洁的非交互式链知识论证(Succinct Non-Interactive Argument of Chain Knowledge, SNACK)。即使验证者(轻节点)和不可信的证明者(全节点)之间进行交互时,它依然可以保证系统安全。他们展示了如何使用图标记(Graph-Labeling)的 PoSW 来构造 SNACK 用于区块链系统。接着,他们指出如何使用 SNACK 来构建轻量级客户端协议,并指出了现有设计的一些缺陷以及对应的缓解措施。最后,他们介绍了增量 SNACK,可以为轻量级挖矿(Light Mining)提供一种新的方法。最小化比特币类区块链的能源成本和碳排放也是共识算法发展的一个方

向。使用有用工作量证明 (Proof of Useful Work, PoUW) 代替工作量证明 (Proofs of Work, PoW) 被视为一个理想的解决方案。然而, 该想法缺乏一个可信的安全实现。在文献[5]中, Fitzi 等人提出了一种基于 PoUW 的区块链协议 Ofelimos, 其共识算法实现了一个分布式的优化问题求解器。他们提出了一个名为双并行本地搜索 (Doubly Parallel Local Search, DPLS) 的本地搜索算法, 用于构造 PoUW 的组件。他们对协议进行了全面的安全分析, 并提出了反应系统有用度的指标。总体而言, 他们的方案为使用区块链系统作为通用优化引擎铺平了道路, 以解决各种需要公开验证解决方案的优化问题。

许多公有链方案的共识协议的安全模型通常需要假设不到一半 (或三分之一) 的资源 (如权益份额或计算能力) 由腐坏方控制, 其安全性证明通常需要假设底层网络是可用的, 以保证诚实方发送的区块在一定时间内被所有诚实方接收。然而, 在实践中底层网络通常是通过泛洪协议实现的, 这些协议在总权重的一小部分可能被腐坏的情况下并没有被证明是安全的。这导致了对许多现有协议的日蚀攻击。文献[6]提出了一个实用的泛洪协议, 该协议可在对数步之后将发送的消息传递给所有诚实方, 并证明了在各参与方都被公开分配正权重且敌手可以腐坏各参与方累积到总权重的固定比例情况下的安全性。这可以直接用于权益证明设置。他们对不同数量的参与方、权重分布和腐坏策略进行广泛模拟。结果表明, 无论权重分布如何, 消息都可以快速传递, 而如果权重分布不均匀, 那么忽略各参与方权重的协议将完全失败。然而, 面对可以腐坏所有发送方节点从而阻止消息传递的自适应敌手, 泛洪协议显然会失败。为了给区块链提供安全保证, 文献[7]在通用可组合性安全框架中精确定义了 δ -延迟敌手。这样的敌手可以自适应地腐坏参与方, 但是从敌手决定腐坏参与方到他们成功控制参与方之间存在时间延迟 δ 。他们正式证明了当 δ 至少是从一个参与方向另一个参与方发送消息所花费的时间加上接收方重新发送消息所花费的时间时, 泛洪协议对于 δ -延迟敌手是安全的。他们的实验结果表明, 针对安全参数 κ 、延迟最多为 Δ 的点对点信道和总共 n 个参与方, 其中具有足够延迟的敌手可以腐坏参与方的任何常数部分, 如果所有参与方平均向 $\Omega(\kappa)$ 方发送, 就可以实现最大延迟 $O(\Delta \cdot \log_2 n)$ 的泛洪功能; 如果所有参与方平均向 $\Omega(\sqrt{\kappa n})$ 方发送, 就可以实现最大延迟 $O(\Delta)$ 的泛洪功能。

本节作者: 包子健、安浩杨、何德彪 (武汉大学)

参考文献

[1] FUCHSBAUER G, ORRU M. Non-Interactive Mimblewimble Transactions, Revisited[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 713-744.

[2] FUCHSBAUER G, ORRU M, SEURIN Y. Aggregate Cash Systems: A Crypto-Graphic Investigation of Mimblewimble[C]. Advances in Cryptology-EUROCRYPT 2019: EUROCRYPT 2019, Darmstadt, Germany, May 19-May 23, 2019. Cham: Springer International Publishing, 2019: 657-689.

[3] JIA Y, SUN S F, ZHOU H S, et al. A Universally Composable Non-Interactive Aggregate Cash System[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 745-773.

[4] ABUSALAH H, FUCHSBAUER G, GAZI P, et al. SNACKs: Leveraging Proofs of Sequential Work for Blockchain Light Clients[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 806-836.

[5] FITZI M, KIAYIAS A, PANAGIOTAKOS G, et al. Ofelimos: Combinatorial Optimization via Proof-of-Useful-Work[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 339-369.

[6] LIU-ZHANG C D, MATT C, MAURER U, et al. Practical Provably Secure Flooding for Blockchains[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 774-805.

[7] MATT C, NIELSEN J B, THOMSEN S E. Formalizing Delayed Adaptive Corruptions and the Security of Flooding Networks[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 400-430.

2 可搜索加密/隐私信息获取/不经意随机存储/隐私保护计算

2.1 可搜索加密

在可搜索加密领域，如何高效安全地处理合取检索和布尔检索等多样化检索请求是一个重要问题。2013 年，Cash 等人在美密会提出了不经意交叉标签（Oblivious Cross-Tags, OXT）协议，解决了上述问题。然而，OXT 协议无法在不预计算的前提下实现两个数据表间的联合（Join）查询，这成为 OXT 协议在实用中难以避免的短板。在文献[1]中，Jutla 和 Patranabis 提出了全新的联合交叉标签（Join Cross-Tags, JXT）协议。JXT 协议的核心思路是将对两个

表进行联合查询的请求分割为在两个表上的子查询请求。这些子查询请求只包含常规的查询表达式，能够在两个表上分别执行。JXT 协议最后汇总子查询的结果，得到两个表进行联合查询的结果。JXT 协议采用对称密钥原语进行构造，其软件实现具备较高的效率。同时，JXT 协议与 OXT 协议完全兼容，两者能够结合在一起同时处理复杂的 SQL 查询请求。文献[1]的工作提升了可搜索加密处理复杂查询请求的能力。

2.2 隐私信息获取

隐私信息检索是实现一个客户端在远程服务器上获取指定数据的同时，对服务器隐藏所获得的数据。传统的隐私信息检索方案多是在两个或多个不合谋的半诚实服务器模型下实现的。文献[2]研究了如何在单个服务器上实现隐私信息检索方案。该方案的构造过程分为两步：一是构造一个双服务器模型下的隐私信息检索方案；二是利用密码假设将这两个服务器“编译”为同一个。在第一步中所构造的方案运行在一个离线-在线模式中。在这个模式中，包含一个离线服务器和一个在线服务器。离线服务器保存加密数据库的部分源信息，在线服务器保存加密数据库本身。客户端在执行查询操作时，先从离线服务器获取相关数据的源信息，再根据这些信息去在线服务器上获取数据本身。在第二步中，文献[2]利用全同态加密对离线服务器上的数据进行加密，并保存在在线服务器上，从而实现了离线服务器和在线服务器的“编译”。最终，所构造的方案实现了亚线性级的服务器查询时间，且支持客户端发布适应性的查询请求，最后所占用的服务器存储开销也仅与数据库的大小呈亚线性关系。

在文献[3]中，Liu 等人主要关注如何实现让一个接收者在消息传递系统中获取发送给自己的消息，同时不向服务器泄露出消息本身以及消息的源数据。简单的处理方法是，下载全部消息并扫描找到发送给自己的消息，会带来极大的通信与计算开销。因此，为了高效安全地实现所希望的功能，文献[3]定义了不经意消息获取与探测这一密码学原语。在该原语的安全模型中，包含有 4 个参与方：消息的发送者、消息的公布者、消息的探测者与消息的接收者。其中消息的探测者是消息接收者的服务器，为消息接收者提供消息获取服务。所定义的密码原语包含很强的安全性，如参与方的合谋与腐败无法导致拒绝服务攻击，以及能够抵抗公钥的链接性泄露等问题。最后，文献[3]基于格密码的同态加密技术构造了实例化方案。

2.3 不经意随机存储

不经意随机访问机（Oblivious Random Access Machine, ORAM）是一种能够在单个半可信服务器上实现对加密数据不经意存取的重要密码学原语。ORAM 为了实现不经意特性，需要对服务器上的加密数据进行大量的冗余访问，造成较大的通信开销，这导致 ORAM 很难

实用。在文献[4]中，作者提出了一种新的 ORAM 构造，称为快照不经意的随机访问机（Snapshot Oblivious RAM, SORAM）。SORAM 绕过了传统 ORAM 中 $\Omega(\log_2 n)$ 的通信开销下界，实现了 $O(\log_2 n)$ 的通信开销。相比传统的 ORAM，SORAM 的核心假设是，在真实系统中，攻击者通常来自外部，而这些外部攻击者只能在发起攻击后且被发现前观察到部分的访问模式，也即访问模式的部分“快照”。在这样的假设下，文献[4]改进了传统 ORAM 的安全性定义，将两个完整操作序列的不可区分性退化为两个完整操作序列中任意两个给定长度序列的不可区分性。在这样的安全性要求下，文献[4]突破了传统 ORAM 的通信开销下界。

2.4 隐私保护计算

文献[5]针对保和加密（Sum-Preserving Encryption, SPE）这一密码学原语进行了研究，并提出了两个全新的 SPE 方案。SPE 是一种保格式加密（Format-Preserving Encryption, FPE），它将一个整数向量加密为另一个整数向量，并且保证加密前后整数向量中各分量的和是相同的。SPE 在图像加密中具有很重要的应用。文献[5]首先利用路径耦合技术对已有的 SPE 方案进行了时间复杂度分析，并得到了一个时间上界，填补了已有 SPE 的理论空白；其次文献[5]基于两种顺序规则分别设计了两个不同的向量排序方案，进一步实现了更加实用的 SPE 方案，降低了密文的存储开销；最后文献[5]对所提出的方案以及已有方案分别编程实现并进行了性能对比，证明了所提方案的实用性。

本节作者：徐鹏、王蔚、陈天阳、郑宇博（华中科技大学）

参考文献

- [1] JUTLA C, PATRANABIS S. SSE and Efficient Searchable Symmetric Encryption for Join Queries [C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 304-333.
- [2] CORRIGAN-GIBBS H, HENZINGER A, KOGAN D. Single-Server Private Information Retrieval with Sublinear Amortized Time [C]. Advances in Cryptology - EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 3-33.
- [3] LIU Z, TROMER E. Oblivious Message Retrieval [C]. Advances in Cryptology - CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 753-783.

[4] YANG D, GENKIN D, GRUBBS P. Snapshot-Oblivious RAMs: Sub-logarithmic Efficiency for Short Transcripts [C]. Advances in Cryptology - CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 152-181.

[5] MIRACLE S, YILEK S. New Algorithms and Analyses for Sum-Preserving Encryption [C]. Advances in Cryptology - ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2022: 3-31.

3 侧信道分析、检测与防护

侧信道攻击可以利用密码电路运行时泄露的侧信道信息恢复出密钥等敏感信息，是密码算法物理安全的主要威胁手段，包括经典密码算法以及后量子密码算法。美国国家标准与技术研究院（NIST）表明在后量子密码标准项目中“考虑后量子密码的抗侧信道攻击特性作为评判算法的一项重要参考”，侧信道分析技术作为对于密码安全性的评判标准之一，如何实现密码算法的侧信道安全检测以及侧信道防护也越来越受到密码学家的关注。以下从侧信道分析、侧信道安全检测和侧信道防护 3 个方面总结了密码算法侧信道领域的相关工作。

3.1 侧信道分析

后量子密码大多基于格、编码、超奇异同源等密码体制。基于格的密码体制在 NIST 标准化进程的各个阶段所占比例最大。针对基于格的数字签名算法 Falcon，Guerreau 等人选择 Falcon 的陷门采样器部分作为攻击对象，通过执行实际功耗攻击，以恢复高斯采样得到的部分信息，进而执行隐藏平行六面体攻击（Hidden Parallelepiped Attack）的变体以恢复全部敏感信息^[1]。同时，Guerreau 等人还改进了 Karabulut 和 Aysu 针对 Falcon 预映射计算部分的攻击方法，降低了攻击的复杂度。针对基于格的 KEM 算法 Saber，Li 等人将深度学习与 SASCA 攻击结合以降低模板数量，基于贝叶斯算法将 Toom-Cook 多项式乘法的因子图进行简化，并改进信任传播算法，降低因子图中环路对信任传播算法的性能影响^[2]。这些技术对其他执行 Toom-Cook 乘法的密码方案的安全性研究也具有借鉴意义。对于基于编码的 KEM 算法 Classic McEliece，Guo 等人发现其在解码过程中，如果错误定位多项式是固定的，那么 FFT 评估过程的行为相同，会产生侧信道泄露^[3]。Guo 等人在其软硬件实现上都成功执行了密钥恢复侧信道攻击。基于超奇异同源的 SIKE 也是一种 KEM 算法，De Feo 等人从理论上分析了坐标随机化防护的 SIKE 在零值攻击威胁下的脆弱性，并选择椭圆点标量乘法部分中的零值和同源性评估部分作为攻击对象，成功攻击了 Cortex-M4 上的 SIKE 实现^[4]。

在后量子密码方案设计中,通常使用 FO 变换机制,将 CPA 安全的 PKE 方案提升到 CCA 安全的 KEM 方案。Ueno 等人选择 FO 变换中 KEM 重加密的伪随机发生器为攻击对象,利用其运行阶段产生的功耗和电磁信息构建选择明文预言机,成功恢复了 NIST PQC 第 3 轮大多数 KEM 候选算法的密钥,包括 Kyber、Saber、NTRU、BIKE 和 SIKE 等^[5]。此外,与传统的 RSA 相比,大多数后量子密码算法被认为能够抵抗部分密钥泄露攻击。但是,Esser 等人通过利用密码算法实现固有的密钥冗余信息,成功执行了对于后量子密码算法的部分密钥泄露攻击,包括公钥密码算法 BIKE、NTRU 和数字签名算法 Rainbow,并且从理论上分析了多项式时间内实现该攻击的信息泄露边界^[6]。

近年来,有许多利用机器学习和深度学习技术进行侧信道攻击的相关工作,攻击对象包括轻量级密码算法、传统公钥密码算法和后量子密码算法。Gohr 等人选择轻量级密码算法 Clyde 为攻击对象,结合基于深度学习和基于线性回归的攻击方法,对第一轮 S 盒的输出位置进行侧信道攻击^[7]。攻击结果表明,基于深度学习的攻击方法可以在建模阶段使用更多的泄露信息,而基于线性回归的攻击方法可以间接观测到更多算法时间的特征。RSA 作为一种被广泛研究的公钥密码算法,其开源的 RSA-CRT 实现通常利用窗口取幂和伪加载的方式来防御传统的侧信道攻击。Saito 等人基于深度学习技术提出了一种单迹能耗攻击,利用神经网络区分 RSA-CRT 实现在选择操作数时的真实负载和伪负载,同时采用启发式优先级的双端队列实现了部分密钥泄露攻击算法,用以恢复窗口取幂的误差分布,在单迹能耗攻击获得部分密钥信息的基础上,恢复完整的加密密钥^[8]。已有的防护方法在面临基于深度学习的侧信道攻击时已不再安全,Saito 等人通过负载顺序随机化技术和伪负载值随机化技术对 RSA 进行防护,用于缓解侧信道泄露问题。基于深度学习的侧信道攻击主要优势之一是它可以在没有预处理或特征工程的情况下,对密码电路执行攻击。在后量子密码的侧信道攻击中,深度学习模型也逐渐被用于对核心算子能耗特征的分类问题上^[2,5],进而达到提升分析效率的效果。

为了广泛了解特征选择如何影响基于深度学习的性能,Perin 等人评估了 3 种不同的特征选择场景下对攻击性能的影响,从最小可能数量的兴趣点(最坏情况下的安全假设)到整个可用的功耗曲线^[9]。Wu 等人提出了一种新的距离度量,称为混合距离,同时考虑了嵌入距离和标签距离,显著提高了攻击时特征提取的质量^[10]。Wu 等人还将深度学习和模板攻击的组合,提出了深度学习辅助模板攻击。利用相似性学习的深度学习方法来找到最相关的数据作为模板攻击的输入。在基于深度学习的侧信道攻击的评价指标方面,猜测熵和成功率通常被用来评估攻击的性能,但其与神经网络模型的评估指标(如准确性、损失和召回率)之间的关系是一个有趣的研究点。Ito 等人经过理论分析发现,具有不同的传统交叉熵和感知信息的多个概率分布可以实现相同的成功率,并不足以评估侧信道攻击的性能^[11]。Ito 等人在消除

这种不确定性的基础上,提出了新的度量用来评估攻击的性能,即通过训练阶段的验证损失可以更加精确地评估成功率,并可以在攻击阶段利用成功率来衡量神经网络模型的泛化能力。

3.2 侧信道安全性检测

在安全检测方面,Richter-Brockmann 等人针对对密码设备常用的两种物理攻击:SCA (Side-Channel Analysis,侧信道分析)与 FIA (Fault Injection Analysis,错误注入分析)的组合进行安全推理,着重考虑了交互效应,将 SCA 和 FIA 的孤立安全模型与可组合性概念(CA)整合到一个整体系统中,提出了 CA 下硬件电路正式安全验证的第一个自动化框架以及验证工具 VERICA^[12]。而 Trautmann 等人提出了一种半自动的定位一段波形中密码算法位置的方法,通过半自动的方法兼顾节省人工成本和人工控制遍历引导结果,并实现了上述算法,对在特定场景下的实现进行了评估^[13]。

不同于传统的攻击分析,在泄露模拟器的构建与研究中,一个关键问题在于模拟结果是否有效覆盖了设备上观测到的泄露信息。在现有侧信道研究中,大多从攻击者视角出发,首先选择攻击的目标中间状态,因此无法覆盖可能存在的其他泄露信息,也无法保证泄露模拟器的模拟“质量”。Gao 等人基于 F 检测提供了一种验证泄露模拟器质量的测试方法,测试了 ELMO/ELMO+/MAPS 等多种现存的 ARM 处理器泄露模拟工具。结果表明,ARM 处理器产生的泄露中可能存在大量未知微架构信息,仅从算法或指令层构建的泄露模型无法准确刻画设备上产生的泄露观测。Gao 等人进一步在分析仿射掩码方案泄露中发现了一类内存读取时产生的微架构泄露,并验证了其对掩码方案安全性产生的影响^[14]。另外,由于缺失微架构模拟,现存泄露模拟器往往不能准确刻画观测到的泄露信息。为构建更为完备、模拟质量有保障的泄露模拟工具,Gao 等人进一步提出了利用 F 检测结果查找缺失信息,逆向恢复对观测泄露有贡献的微架构元素。实验表明,除微架构内部逻辑数据流之外,观测泄露中还包含由于电路毛刺产生的分量。Gao 等人进一步用更为完备的泄露模型构建了泄露模拟器 uELMO。与现存泄露模拟器对比,uELMO 能更为准确地找出仅存在于微架构中或者实践中易忽略的泄露点^[15]。

3.3 侧信道防护

在对称密码防护方面,针对当前最主流的对称算法 AES,Belenky 等人将 AES 算法执行的有限域替换为一个更大的交换环,通过将环随机化的方法实现了 2 阶抗侧信道安全,并且通过实验验证该方案对主流的替换攻击也有一定的防护能力^[16]。相比于现有的侧信道防护方案,该方案在嵌入式系统中的运行速度有明显提升。另外,已有可证明安全的侧信道防护需

要大量随机数，随机数的需求随着被防护密码算法的规模而线性增长，这严重制约了在资源受限的环境（如物联网芯片等）中的侧信道防护。Wang 等人提出了随机数均摊的思路，首次在不降低计算复杂度的前提下，构造了随机数需求与密码算法规模无关的防护方法。文章把新的防护方法应用在 AES 密码算法中，同现有的防护方法相比，随机数需求降低了近 $1/20^{[17]}$ 。

掩码技术和洗牌（Shuffling）技术是侧信道防护的两个非常重要的方法。掩码技术可以达到侧信道可证明安全性，但是需要一定的基础噪声。洗牌技术可以高效地增加侧信道攻击的难度，在一定程度上提升噪声。Azouaoui 等人探索了掩码技术和洗牌技术结合之后的侧信道防护效果。文章列举并形式化地给出了 bitslice 的掩码方案和洗牌的多种结合方法。Azouaoui 等人在考虑了洗牌过程的泄露情况下，发现在寄存器允许范围内单纯使用 bitslice 的掩码方案效果反而更好^[18]。另外，侧信道掩码防护虽然具有可证明安全性，但是存在实现复杂、开销过大的问题。Feldtkeller 等人提出了一种随机数重用的方法，即同一个随机数可以使用在带掩码防护的密码算法的各个位置，从而大幅度节省随机数使用的开销。并且把新提出的方法和电子自动化设计工具链相融合，实现了对无防护电路的自动化防护，有效降低了掩码算法实现的复杂度^[19]。而围绕基于编码的掩码方案的高效实现和实际安全性评估问题，Wu 等人从编码数学结构出发，提出了 3 种结构性优化方法，在保证安全性的条件下显著降低实现代价，成倍地提升防护型密码实现的效率。Wu 等人选取 3 类代表性编码实例化方案，利用模板攻击和二阶相关能量/电磁攻击对实例化方案进行基于实际攻击的分析与评估，有效论证了基于编码的掩码方案具有的优良侧信道攻击防御能力，为安全密码芯片设计者提供了构建新颖防护方案的新思路和新方法^[20]。

从应用角度出发，掩码侧信道防护方法需要把密码算法中所有的运算中间值分成 n 份，此时所需的存储单元数量也增长了 n 份，增大了密码电路的开销。Naito 等人从设计认证加密模式考虑，尽可能把传统认证加密中秘密的信息设置为公开，而这些公开的信息无须掩码防护，从而有效降低了由保存中间值所产生的额外开销^[21]。

考虑到大部分密码应用场合中密码使用者并不具备侧信道防护相关知识，Knichel 等人实现了一个自动化工具 AGEMA，它可以把没有加掩码防护的硬件实现，自动化的转化为掩码防护的硬件实现，这为掩码设计经验薄弱的人员提供了便利^[22]。

此外，侧信道防护技术在神经网络的知识产权保护中也起到了一定作用。利用大量数据训练得到的神经网络作为一种知识产权往往是需要被保护的，但是目前已经出现了很多针对训练后神经网络的逆向工程攻击。针对这些攻击，需要设计相应的防护方法来保护神经网络被泄露。Dubey 等人给出了一种基于模运算的神经网络，这种神经网络的基本运算可以很好地匹配掩码防护方法。然后论文把 Domain-Oriented Masking（DOM）掩码技术应用在对这种网络的保护中，并在 Glitch-Extended Probing Model（毛刺扩展探针模型）下进行了安全性证

明。最后论文进行了工程实现并检测其侧信道泄露，结果展示新的防护方法在保证安全的前提下，有效地节省了硬件资源^[23]。

本节作者：王伟嘉（山东大学），李延斌（南京农业大学）

参考文献

- [1] GUERREAU M, MARTINELLI A, RICOSSET T, et al. The Hidden Parallelepiped Is Back Again: Power Analysis Attacks on Falcon[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(3): 141-164.
- [2] LI Y, ZHU J, HUANG Y, et al. Single-Trace Side-Channel Attacks on the Toom-Cook: The Case Study of Saber[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 285-310.
- [3] GUO Q, JOHANSSON A, JOHANSSON T. A Key-Recovery Side-Channel Attack on Classic McEliece[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 800-827.
- [4] DE FEO L, EL MRABET N, GENET A, et al. SIKE Channels[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(3): 264-289.
- [5] UENO R, XAGAWA K, TANAKA Y, et al. Curse of Re-encryption: A Generic Power/EM Analysis on Post-Quantum KEMs[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 296-322.
- [6] ESSER A, MAY A, VERBEL J, et al. Partial Key Exposure Attacks on BIKE, Rainbow and NTRU [C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, March 2-June 24, 2022. Cham: Springer Nature Switzerland, 2022: 346-375.
- [7] GOHR A, LAUS F, SCHINDLER W. Breaking Masked Implementations of the Clyde-Cipher by Means of Side-Channel Analysis: A Report on the CHES Challenge Side-Channel Contest 2020[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 397-437.
- [8] SAITO K, ITO A, UENO R, et al. One Truth Prevails: A Deep-learning Based Single-Trace Power Analysis on RSA-CRT with Windowed Exponentiation[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 490-526.
- [9] PERIN G, WU L, PICEK S. Exploring Feature Selection Scenarios for Deep Learning-based Side-channel Analysis[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 828-861.

[10] WU L, PERIN G, PICEK S. The Best of Two Worlds: Deep Learning-assisted Template Attack[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(3): 413-437.

[11] ITO A, UENO R, HOMMA N. Perceived Information Revisited New Metrics to Evaluate Success Rate of Side-Channel Attacks[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 228-254.

[12] RICHTER-BROCKMANN J, FELDTKELLER J, SASDRICH P, et al. VERICA-Verification of Combined Attacks: Automated formal verification of security against simultaneous information leakage and tampering[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 255-284.

[13] TRAUTMANN J, BECKERS A, WOUTERS L, et al. Semi-automatic locating of cryptographic operations in side-channel traces[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 345-366.

[14] GAO S, OSWALD E. A Novel Completeness Test for Leakage Models and Its Application to Side Channel Attacks and Responsibly Engineered Simulators[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 254-283.

[15] GAO S, OSWALD E, PAGE D. Towards Micro-architectural Leakage Simulators: Reverse Engineering Micro-architectural Leakage Features Is Practical[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 284-311.

[16] BELENKY Y, BUGAENKO V, AZRIEL L, et al. Redundancy AES Masking Basis for Attack Mitigation (RAMBAM)[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 69-91.

[17] WANG W, GUO C, YU Y, et al. Side-Channel Masking with Common Shares[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 290-329.

[18] AZOUAOUI M, BRONCHAIN O, GROSSO V, et al. Bitslice Masking and Improved Shuffling: How and When to Mix Them in Software?[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 140-165.

[19] FELDTKELLER J, KNICHEL D, SASDRICH P, et al. Randomness Optimization for Gadget Compositions in Higher-Order Masking[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 188-227.

[20] WU Q, CHENG W, GUILLEY S, et al. On Efficient and Secure Code-based Masking: A Pragmatic Evaluation[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 192-222.

[21] NAITO Y, SASAKI Y, SUGAWARA T. Secret Can Be Public: Low-Memory AEAD Mode for High-Order Masking[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 315-345.

[22] KNICHEL D, MORADI A, MULLER N, et al. Automated Generation of Masked Hardware[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 589-629.

[23] DUBEY A, AHMAD A, PASHA MA, et al. Modulonet: neural networks meet modular arithmetic for efficient hardware masking[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 506-556.

4 拜占庭协议和拜占庭广播/安全消息传输/其他真实协议

拜占庭协议是分布式计算中的重要组成部分。该协议有多个参与方，其中可能有一些参与方是恶意的。协议开始时每个参与方有各自的输入，协议运行结束后保证所有诚实的参与方输出同样的值。

作为运行效率的一个重要衡量指标，如何减少拜占庭协议的轮数是一个重要的问题。之前的研究已经证明了该类协议的轮数下界，如果恶意参与方的数量 t 相对于总参与方数量 n 是线性的 ($t=\Theta(n)$)，那么任意的 r 轮拜占庭协议的运行失败概率至少是 $(cr)^r$ (其中 c 是常数)。然而，目前的协议只能达到 2^{-r} 的失败概率，如何构建一个达到失败概率下界的拜占庭协议是一个长期未解问题。对此，Ghinea 等人^[1]给出了首个达到上述下界(只相差常数因子)的拜占庭协议。该协议基于公钥基础设施和可信设置，可以容许适应性敌手腐蚀最多 $t=(1-\varepsilon)n/2$ 个参与方。

此外，基于拜占庭协议等共识算法在区块链上的重要性，如何在处理长信息(如区块信息)时降低通信复杂度是一个值得考虑的问题。对长度为 1 比特的信息，之前已知的通信复杂度最低是 $O(nl \cdot \text{poly}(\kappa))$ 或 $O(nl + n^2 \cdot \text{poly}(\kappa))$ 。Bhangale 等人在文献[2]中对此做了优化，构建了通信复杂度为 $O(nl + n \cdot \text{poly}(\kappa))$ 的拜占庭协议。该协议基于可信设置，可抵御适应性敌手的攻击。

与拜占庭协议不同,拜占庭广播有一个指定的信息发送方,虽然协议结束也要求诚实的参与方输出相同的信息,但是这个信息是发送方发送的。该类广播协议广泛应用于各种密码协议,如秘密分享、安全多方计算和区块链共识等。类似地,还有一个满足大多应用场景的并行拜占庭广播问题,该问题与上述朴素的拜占庭广播的区别是可以有多个信息发送方。对此,Tsimos 等人在文献[3]中提出了基于八卦(Gossiping)信息传播技术的高效广播协议。其中,不需要可信设置的并行广播协议达到了 $\tilde{O}(n^3\kappa^2)$ 的通信复杂度,而基于可信设置的协议的通信复杂度 $\tilde{O}(n^2\kappa^4)$ 几乎达到该类协议可能达到的最优界。

安全通信协议是被广泛应用的一类密码协议,其中安全连接协议更是加密了超过 90% 的万维网数据。针对 2018 年部署的重要安全连接协议传输层安全协议 TLS 的最新版本 TLS 1.3,之前已经有大量的可证明安全性分析。这些分析中给出的证明大多是渐进安全但不是紧致安全的,即在安全参数很大时可以保证协议安全性,而当实际部署的协议选取具体的标准安全参数时也许无法给出很好的具体安全性。对此,基于前人对 TLS 1.3 完整握手模式的紧致安全性研究,Davis 等人在文献[4]中对 TLS 1.3 较为简单的预共享秘钥 PSK 握手模式给出了首个紧致安全证明。此外,该证明的安全性假设也相对较弱,其所用方法也有望对 TLS 1.3 完整握手模式进行更好的证明。

近年来,随着人们对大规模监控的不满,端到端的安全消息传输协议成为一个重要的密码学研究方向。这类协议也属于安全通信协议,但其主要用于用户之间长期(如数月数年)的信息传输(如即时通信),而不是如安全连接协议一样用于客户端和服务端之间短期(如几分钟甚至几秒)的信息传输(如网页浏览)。由于安全消息传输协议的会话较长,用户被攻击的风险更大,因此该类协议的安全性保障比安全连接协议也更强。目前应用最广的协议是由 Signal 公司提出的同名协议。对该协议的安全性分析目前已经有大量已有工作,而之前的工作基本都没有对 Signal 协议的通用可组合(UC)安全性进行研究。Canetti 等人^[5]和 Bienstock 等人^[6]同时给出了针对 Signal 协议的 UC 安全性分析,其中文献[5]的 UC 框架对该类协议的建模更加全面,而文献[6]对 Signal 协议本身的安全性进行了更加细致的解剖并提出了相应的改进方案。此外,针对 Signal 协议的密码核心双棘轮算法,Pijnenburg 和 Poettering 在文献[7]中对其即时解密性质进行了安全性分析。该工作指出了目前的双棘轮算法的即时解密特性会损害前向安全性,并提出了更强的安全模型和满足该安全性的协议。该协议在经过预设时间阈值后会自动使密文过期,从而可以在基本不影响即时解密性质的前提下增强前向安全性。

以上讨论的都是针对两个用户之间的安全消息传输,而多用户之间的群组消息传输可以由两两之间的传输协议实现,如在每两个用户之间建立 Signal 协议通道。然而,这种朴素的方法会导致较大的通信复杂性,一个很重要的研究问题是如何降低群组消息传输协议的通信复杂度。对此,类似 TLS 的消息层安全协议 MLS 目前正在构建当中,而其安全性也已经有

大量的分析工作。Alwen 等人在文献[8]中首次对 MLS 协议在有恶意用户的情况下的安全性进行了分析,指出了一系列针对该类安全性漏洞的实际攻击和修补方案,而这些补丁已经被 MLS 协议采纳。

除了上述协议,一些其他重要的真实协议最近也有相应研究。针对分布式计算中重要的领导选举问题, Komargodski 等人^[9]将之前的最优协议轮数由 $O(\log_2(\log_2 n))$ 轮优化到 $O(\log_2 n)$ 轮,并保持了同样的博弈论公平性。关于大规模分布式场景(如区块链)中适应性安全协议的构建,一个重要问题是如何高效地选举随机匿名委员会; Matteo Campanelli 等人在文献[10]将其中的关键挑战——隐私状态在委员会之间的传递问题单独抽象为一个名为“面向未来加密”的密码学原语,并给出了安全实例。在之前对选举系统的研究中,如果其基于类似选择密文攻击(CCA)安全的加密算法,就无法很好地满足无收据性质(如只能加密很小的选票数据),为此 Devillez 等人^[11]提出了名为“可追踪无收据加密”的密码学原语用来构建无收据选举协议,并给出了该原语的安全实例。最后,在与拜占庭协议相关但更困难的状态机复制问题上,已有的协议所运行的网络通常是确定的,即同步或异步其一;最近基于不可知网络的安全状态机复制协议也被构造了出来, Alexandru 等人^[12]给出了更安全和更低通信复杂度的该类协议,其协议甚至可以在经常变动的网络中运行。

本节作者:陈杉(南方科技大学)

参考文献

- [1] GHINEA D, GOYAL V, LIU-ZHANG C D. Round-Optimal Byzantine Agreement[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 96-119.
- [2] BHANGALE A, LIU-ZHANG C D, LOSS J, et al. Efficient adaptively-secure byzantine agreement for long messages[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 504-525.
- [3] TSIMOS G, LOSS J, PAPAMANTHOU C. Gossiping for communication-efficient broadcast[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 439-469.
- [4] DAVIS H, DIEMERT D, GUNTHER F, et al. On the Concrete Security of TLS 1.3 PSK Mode[C]. Advances in Cryptology-EUROCRYPT 2022: EUROCRYPT 2022, Trondheim, Norway, May 30-June 3, 2022. Cham: Springer International Publishing, 2022: 876-906.

[5] CANETTI R, JAIN P, SWANBERG M, et al. Universally composable end-to-end secure messaging[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 3-33.

[6] BIENSTOCK A, FAIROZE J, GARG S, et al. A more complete analysis of the signal double ratchet algorithm[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 784-813.

[7] PIJNENBURG J, POETTERING B. On Secure Ratcheting with Immediate Decryption[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 89-118.

[8] ALWEN J, JOST D, MULARCZYK M. On the insider security of MLS[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 34-68.

[9] KOMARGODSKI I, MATSUO S I, SHI E, et al. $\log^*$ -Round Game-Theoretically-Fair Leader Election[C]. Advances in Cryptology-CRYPTO 2022: CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022. Cham: Springer Nature Switzerland, 2022: 409-438.

[10] CAMPANELLI M, DAVID B, KHOSHAKHLAGH H, et al. Encryption to the future: a paradigm for sending secret messages to future (anonymous) committees[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 151-180.

[11] DEVILLEZ H, PEREIRA O, PETERS T. Traceable receipt-free encryption[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 273-303.

[12] ALEXANDRU A B, BLUM E, KATZ J, et al. State machine replication under changing network conditions[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 681-710.

密码实现

量子计算机的发展对传统公钥密码算法（如 RSA 和椭圆曲线密码）产生了威胁，规模足够大的量子计算机能在多项式时间内破解传统公钥密码算法，因此能抵抗量子计算威胁的抗量子密码备受学界关注。除此之外，对称密码的性能优化能提升系统响应速度、降低系统运行的成本，也是较为重要的研究课题。本章主要针对格密码实现和对称密码实现进行介绍。

1 格密码实现

在格密码性能优化研究的相关工作中，最受研究者关注的是多项式乘法优化，大量工作表明数论变换（Number Theoretic Transform, NTT）算法能够加速格密码的多项式乘法运算。NTT 算法的核心操作之一是模乘运算，目前最常用的模乘算法是 1985 年提出的 Montgomery 算法和 1986 年提出的 Barrett 算法，大多数密码方案均采用这两类模乘算法。2021 年，Plantard 提出了一个优于 Montgomery 算法和 Barrett 算法的模乘算法，称为 Plantard 算法。原始的 Plantard 算法无法直接用于格密码中，为了解决这个问题，黄军浩等人^[1]对 Plantard 算法进行了改进，并提出了改进的 Plantard 算法。将其应用到格密码后，性能测试结果表明其优于目前广泛使用的 Montgomery 算法和 Barrett 算法，刷新了格密码 CRYSTALS-Kyber 方案在 ARM Cortex-M4 平台上的性能记录。同样是针对模乘运算，Becker 等人在文献[2]中使用 ARMv8 Neon 指令将 Montgomery 算法和 Barrett 约减算法结合起来，提出了更快的模乘算法用于加速 NTT。在 ARM Cortex-A72 CPU 上的测试表明，相比于之前最快的实现，Kyber 与 Saber 的矩阵多项式乘法运算分别加速 1.7 倍和 2 倍。与此同时，在 Apple M1 CPU 上的测试表明，Kyber 与 Saber 的矩阵多项式乘法运算分别加速了 2.1 倍和 1.9 倍。

NTT 参数的选择决定着不同的实现方式，Alkim 等人在文献[3]中针对 NTRU 以及 NTRU Prime 中使用的参数，设计了 NTT 的实现方案，该文献通过结合 Vector-Radix FFT 以及 Good-Thomas FFT，设计了一种混合基的蝴蝶变换。相较于现有的技术，该文献所提出的实现方案性能提高 2.8%~10.3%，且不依赖于特定的多项式环结构，能够适用于不同参数的 NTT。Chen 等人^[4]提出了一种高效且无内存映射冲突的实现方案，该方案支持多个并行的蝴蝶变换单元和多种 NTT 基数的选择，与其他方法相比具有更广泛的适配性。基于该方案，通过算法-硬件协同设计，Chen 等人提出了一种可扩展的基为 2 和基为 4 的 NTT 乘法架构。该

架构使用特有的调度方法减少基为 4 的蝴蝶变换单元中的模加/模减和模乘的数量。实验结果和理论分析结果表明,在交叉存储系统中,相同并行度下基为 4 的 NTT 的 ATP 指标优于基为 2 的 NTT,当并行度增加时,优势将更大。

为了降低格密码的内存占用,Becker 等人在文献[5]中通过利用 ARM Cortex-M4 CPU 中的 M-profile Vector Extension (MVE) 指令集,实现了一种内存占用低的 Toom-Cook 算法。区别于经典的 Toom-Cook 算法,其在多项式评估 (Evaluation) 与插值 (Interpolation) 阶段优化了算法的访存模式,降低了内存访问次数,通过加载多项式系数至寄存器,并利用相关并行指令进行打包读取从而降低读写内存的复杂度。与现有的实现方案相比,该算法具有 3~5 倍的性能提升,且该文献提出的多项式乘法技术已经被后量子算法 Saber 所采用。

此外,流水线架构设计也是常用的优化手段,Zhao 等人在文献[6]中介绍了一种针对 CRYSTALS-Dilithium 数字签名方案的优化方法,实现了一种紧凑且高效的硬件架构。在该文献中提出了分段流水线处理方法,旨在减少存储需求和处理时间,并设计了几个优化模块,包括流水线 NTT 模块、采样模块、Decompose 模块和 3 个约减模块,以提高所提出架构的效率。相比于类似平台上现有的 Dilithium 设计,其实现所需的 LUT、FF、BRAM、DSP 资源分别减少了 5/7、5/7、1/3、2/9,对于按照 NIST 安全级别为 5 级选取参数的 Dilithium,密钥生成、签名生成和签名验证所需时间分别减少 5/22、10/17、5/7。

2 对称密码实现

关于对称密码实现的研究也取得了很多进展。Liu 等人^[7]提出了一种搜索低延迟线性层实现的启发式算法。与之前大多数算法所采用的前向式搜索不同,该算法采用了新的后向式的搜索架构。这种搜索方法通过限制表达式的分裂次数可以找到接近最低深度的电路实现。Liu 等人通过该搜索方法搜索到了 AES 算法列混淆矩阵目前最优的实现结果 (103 个异或门,电路深度为 3)。同时,他们搜索到了 $M_4(GL(8, \mathbb{F}_2))$ 上具有更小实现代价的对合 MDS 矩阵,该矩阵在深度为 3 的实现电路上仅需要 86 个异或门,相比之前的最优结果减少了两个异或门。

在电池驱动的嵌入式设备中低能耗是密码算法设计的目标之一。2008 年,Canier 和 Preneel 提出了 Trivium 同步流密码,该密码结构是专门为资源受限的硬件设备量身定制的,其硬件实现电路面积占用小。Caforio 等人^[8]提出了流密码领域第一个启发式的能量模型——完美树能量模型,将状态更新函数的底层代数拓扑结构和能量消耗建立联系,借此可以找到多个能耗最优版本的 Trivium 替代算法。此外,Caforio 等人还提出了 Trivium-LE(F)和 Trivium-LE(S)两种算法,其相比于原始的 Trivium 分别减少了 15%和 25%的能量消耗,是至今能耗最低的加密原语,且与原始的 Trivium 算法具有同样的安全等级。

本节作者：刘哲（南京航空航天大学/之江实验室），何德彪（武汉大学），周璐（南京航空航天大学）

参考文献

- [1] HUANG J, ZHANG J, ZHAO H, et al. Improved Plantard Arithmetic for Lattice-based Cryptography[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 614-636.
- [2] BECKER H, HWANG V, KANNWISCHER M J, et al. Neon NTT: Faster Dilithium, Kyber, and Saber on Cortex-A72 and Apple M1[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 221-244.
- [3] ALKIM E, HWANG V, YANG B Y. Multi-Parameter Support with NTTs for NTRU and NTRU Prime on Cortex-M4[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 349-371.
- [4] CHEN X, YANG B, YIN S, et al. CFNTT: Scalable radix-2/4 NTT multiplication architecture with an efficient conflict-free memory mapping scheme[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 94-126.
- [5] BECKER H, MERA J M B, KARMAKAR A, et al. Polynomial multiplication on embedded vector architectures[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 482-505.
- [6] ZHAO C, ZHANG N, WANG H, et al. A Compact and High-Performance Hardware Architecture for CRYSTALS-Dilithium[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 270-295.
- [7] LIU Q, WANG W, FAN Y, et al. Towards Low-Latency Implementation of Linear Layers[J]. IACR Transactions on Symmetric Cryptology, 2022(1): 158-182.
- [8] CAFORIO A, BANIK S, TODO Y, et al. Perfect Trees: Designing Energy-Optimal Symmetric Encryption Primitives[J]. IACR Transactions on Symmetric Cryptology, 2021(4), 36-73.

设备安全/物理攻击/故障注入攻击/ 掩码保护

1 设备安全

现代计算机体系结构中的信息安全问题日益重要, 尽管计算机系统的处理器和内存等组件受到物理访问控制, 但是这些组件的功耗仍然可能暴露出信息。这些泄露可以被利用, 从而破坏计算机系统的安全性。Marshall^[1]等人提出了一个新的工具 Miracle, 用来评估不同计算机设备中的不同指令集架构的功率泄露, 包括 32 比特微控制器以及各类指令集和微架构。并且可以帮助研究人员了解泄露的程度和性质, 包括提出精确的泄露模型、选择合适的评估设备以及针对包括掩码的各类具体情况。而 Casagrande^[2]等人探讨了小米健身追踪生态系统的安全漏洞, 以及如何利用这些漏洞来实现攻击。此外, Schneider^[3]等人探讨了将可信执行分离和整合到更多计算架构中的机制, 以便可以满足新一代应用程序的安全性需求, 并描述了一种新的系统架构, 称为 Composite Enclave。Composite Enclave 改变了传统的可信执行方法, 它可以将可信执行分解成多个安全起点, 每个安全起点都可以在不同的环境中安全运行, 并在运行过程中提供强大的安全性保护。它还可以支持多种安全协议, 如基于密钥的访问控制、基于角色的访问控制、可信任的计算等。这些特性使得 Composite Enclaves 能够支持更复杂的应用程序, 并且抵御更多的攻击。

Carrier^[4]等人关注的是一类特定的加密算法, 称为线性码。线性码通常用于数据传输和纠错编码。解码线性码的问题可以被形式化为一个线性规划问题, 它的解可以使用标准算法如单纯形法来获得。然而, 在实践中, 线性码通常会通过添加噪声的方式来保护它们免受攻击。这些噪声可以使解码问题变得非常困难, 这是由于现有的解码算法通常不能有效地处理噪声。因此, 需要一种新的方法来解决这个问题。Carrier 等人将解码问题转化为解决 LPN 问题。LPN (Learning Parity with Noise) 问题是一个经典的密码学问题, 其目标是从带有噪声的线性方程组中恢复向量。Carrier 等人基于线性规划的技术来解决 LPN 问题, 从而解码线性码。此外, Beere^[5]等人探讨了逻辑锁定技术。逻辑锁定的目的是保护电路的知识产权不受生产商的影响, 方法是将电路的原始逻辑修改成一个新的“锁定”电路, 若不具备密钥则无法获取原始电路的任何信息。虽然逻辑锁定为芯片的外包制造提供了一个有前景的解决方案, 但是

当前提出的几个逻辑锁定系统均已经被攻破。缺乏既定的安全技术，一部分原因是缺乏对逻辑锁安全概念的严格处理，以及理论和实践之间的脱节。Beere1 等人通过引入形式化的定义来捕捉逻辑锁定方案所需的安全性，进而解决这一问题。

Kathrin^[6]基于物理不可克隆函数（Physical Unclonable Functions, PUF）提出了一种无电池电容式的外壳，PUF 用于解密密钥链的密钥加密密钥（Key Encryption Key, KEK）的生成，以保护设备免受物理攻击。为了补偿噪声和环境影响，PUF 密钥中使用了纠错码机制。然而，一些破坏外壳的钻孔攻击可能会改变 PUF 响应。Kathrin 通过分析 PUF 响应的行为，从 q 元极化码中构建一条窃听信道，进而提出了一种新的解决方案。该方案具有 100 位的物理安全性，并让 PUF 密钥具有 306 位熵。这种方法使得基于 PUF 的安全外壳具有商用性的可能。Deric^[7]等人探讨的问题是零信任芯片系统的完整性检查技术。当前行业趋势正朝着使用小芯片替代现代芯片的单片制造方向发展。每个芯片是一个单独生产的硅片，通过将芯片封装在一个硅片夹层或桥上，就形成了一个系统芯片（SoC）。芯片实现了 IP 重用、异构集成，并能更好地利用适合成本的工艺节点。然而，用单独生产的组件创建系统也带来了安全风险，如可能存在芯片交换、易受插板探测或篡改的影响。在零信任的安全态势下，用户不应盲目地认为芯片是在一个友好的环境中运行的。在文献中提出了一种新的零信任晶片系统的完整性检查技术——Between-Die Delay PUF。该技术允许单个小芯片与它的邻居启动一个协议，测量传入信号传播延迟的独特变化，以此实现在晶片间的完整性检查，从而使系统能够更有效地检测到攻击行为。该技术使用了延迟 PUF 技术，可以提高系统的可靠性，并且可以检测到跨芯片的攻击行为。该技术还提供了一种不需要外部存储器的认证方法，从而可以更有效地降低系统的成本。此外，在文献中还提出了一种新的基于 PUF 的认证方法，可以有效地检测到攻击行为，从而确保系统的安全性。

Udugama^[8]等人提出了一种名为 VITI 的轻量级传感器，可用于测量可编程逻辑器件（FPGA）中的功耗变化。这种传感器可以在 FPGA 芯片中嵌入，能够自校准并提供准确的功耗测量结果，可用于电源管理和故障检测等应用，如故障检测、电源管理和性能优化等。该电路自校准功能可调整可调延迟元件，无须反复试验即可实现功率变化测量。可以通过对不同的 FPGA 芯片进行测试来验证 VITI 的性能，并可以准确地测量功耗变化。通过使用 VITI 传感器，能够识别 FPGA 中不同元件之间的功耗差异，并可以更精确地测量功耗变化。将该模块用于片上远程旁路信息采集与攻击，所消耗资源将小于基于 TDC 和基于 RO 而设计的传感器。实验证明，该模块所采集的功耗变化可以针对硬件实现的 AES 进行完整的密钥恢复攻击。

2 物理攻击

故障注入攻击对加密实现构成严重威胁，其中激光故障注入（Laser Fault Injection, LFI）被认为是对安全设备实施攻击的最强大的半侵入性故障注入方法。Selmke^[9]等人首次讨论了非线性双光子吸收（TPA）效应在 LFI 中的应用。结果表明，TPA 相比常规的线性 LFI 方法有几个优越性。TPA 的效应允许在不增加诱导能量的情况下对非薄型器件进行处理并因此增加器件上的应力。Bartewitz 等人采用轻量级分组密码 SKINNY，并配备各种形式的基于 EDC（Error-Detection Codes）的 CED（Concurrent Error Detection）方案，在 40nm 的 AISC 上以密码协作处理器的形式进行硬件实现，然后评估它们对激光故障输入攻击的抵抗能力。结果表明，当代码的长度 n 小于其等级 k 的两倍时（没有完全冗余），很容易攻克防御对策的保护。这并不是由基础设计方法或概念中的任何缺陷造成的，而只是证明了定义的对手模型是很容易被克服的。文献中使用普通的单次 LFI 设置对目标进行标准的黑盒扫描，就足以在 $n < 2k$ 时偶尔注入比底层对手模型所约束的更多的单比特故障。事实证明，这种事件的概率大到足以在几小时内通过差分故障分析（DFA）成功进行密钥恢复攻击。因此，可以认为不要把基于代码的 FI 对策的冗余度限制在小于每个字的比特数。

Barua^[11]等人提出了一种名为 PreMSat 的防御措施，以防止对无源霍尔传感器的饱和攻击。PreMSat 的核心思想是，它可以产生一个内部磁场，其强度与攻击者注入的外部磁场相同但极性相反。因此，PreMSat 产生的内部磁场可以使注入的外部磁场失效，同时防止在传感器的线性区域的故意欺骗以及饱和区域内的饱和攻击。此外，Götte^[12]等人介绍了一种针对物理攻击的新对策——惯性硬件安全模块（Inertial Hardware Security Modules, IHSM）。传统系统有着共同点：它们的安全需要设计精细的传感器结构，从而对被监控的安全边界或体积的微小操作做出响应。该方法的新颖之处是因为降低了安全网和其他传感器的灵敏度要求，并通过高速旋转安全网或传感器来增加任何操作的复杂性，从而为攻击者提供一个移动目标，并且建立了一个概念验证的硬件原型来展示对该概念主要工程挑战的解决方案。作为这个概念验证的一部分，可以发现一个使用由商业印制电路板和汽车高重力加速度计制成的粗略安全网的系统就可以提供一个有用的安全水平。

3 故障注入攻击

故障攻击利用密码设备中出现错误时暴露出来的信息来对其进行攻击。作为一种主动型攻击，故障攻击通常可以分为两个阶段。在第一阶段中，攻击者会对密码设备进行干扰使其

出错,这一阶段通常被称为故障注入阶段;在第二阶段中,攻击者会利用设备出错后泄露的信息对其进行攻击,这一阶段通常被称为故障分析阶段。

在 2018 年 CHES 会议上, Zhang 等人提出了持久性故障分析 (Persistent Fault Analysis, PFA), 研究表明恢复出密钥信息大约需要 2200 条故障密文。持久性故障分析因为其更为宽松的要求, 近年来受到了很多关注。这种攻击方法一般会攻击密码算法的常量, 如分组密码的 S 盒, 因为这能一直影响整个加密过程。在 2018 年提出的方案中仅使用了分组密码最后一轮故障泄露, 为了完善这一缺陷, 张帆^[13]等人提出了代数持久性故障分析 (Algebraic Persistent Fault Analysis, APFA) 通过引入代数分析, 将分组密码以及多轮中的故障泄露转化为 SAT 问题下的代数表示, 并交由 SAT 求解器来计算密钥。APFA 对分组密码的轮函数操作进行了抽象, 并提出了持久性故障的通用代数表示, 其能够让攻击者按照特定规则为不同分组密码建立对应的方程并构建约束。通过 APFA, 张帆等人攻破了原始 PFA 难以解决的 PRESENT-128, 并将 SKINNY-64-64 所需的密文数量从 1550 条降至 10 条, 而且他们还将 APFA 应用至 Feistel 结构下的 LBlock-80 算法以及重量级分组密码 AES-128。

此外, Soleimant^[14]等人在经典持久性故障分析的基础上, 向 S 盒注入多个持续性故障的攻击, 高效地恢复出了 AES 算法的密钥。PFA 在以往仅能针对单个已知的故障注入, 且注入单个故障需要攻击者有较强的能力 (使用激光或 Rowhammer 等方法)。而现实是廉价的电磁故障攻击等方法往往会向 S 盒中注入多个未知故障, 因此 Soleimant 等人提出了实用的持久性多故障分析 (Practical Multiple Persistent Faults Analysis, PMPFA)。PMPFA 巧妙地利用了多故障下 S 盒不可能输出集合在每个字节中有着相同分布的性质, 通过收集每个密文字节中的未输出值集合并根据算法恢复密钥。在 8 个故障下, PMPFA 仅用 671 条密文就能够将密钥的搜索空间降至 $2^{14.56}$ 。在 16 个故障下, 利用 477 条密文将密钥搜索空间降至 $2^{24.52}$ 。

4 掩码保护

掩码是一种广泛使用的硬件保护技术, 旨在防止旁路攻击。然而, 掩码电路仍然存在泄露问题, 其中包括电平转换泄露 (Transitional Leakage), 这是由于连续调用 (如时钟周期) 导致存储器内容的不符预期的重新组合。这些重新组合可能导致不安全的信息泄露, 因此需要对其进行建模和分析。Müller^[15]等人在探讨了掩码电路中的电平转换泄露问题中扩展了泄露验证框架 SILVER, 通过同时考虑毛刺问题和存储器电平转换问题来评估带掩码的硬件电路的泄露情况。为此 Müller 等人扩展了 SILVER 处理和分析迭代数字电路的能力。通过该框架, 可以准确建模毛刺问题和电平转换问题, 分析了由 Boss 等人提出的 S 盒结构, 并报告了由电平转换所造成的信息泄露。此外, Müller 等人集成的硬件隐私电路 (Hardware Private Circuit), 可以用来评估由“输出探针隔离非干扰” (O-PINI) 构造的电路。

Knichel^[16]等人提出了一种新颖的通用框架,该框架可以为任意向量布尔函数轻松构建可组合的计算组件,可以将较大较复杂的电路组合转换为掩码计算组件。基于毛刺扩展探测模型和 PINI 的安全可组合概念,该框架可以生成一阶安全的掩码计算组件,可以很好地集成当前仅通过布尔函数便可以自动生成硬件掩码方案的 EDA 工具,为将任意布尔函数转换为一阶安全的可组合掩码组件提供了可能性。之后,Knichel^[17]等人再次提出了一种称为 COMAR 的方法,它仅用 6 个新的随机比特便可以实现任意一阶安全计算小组件的安全组合,并保证其硬件实现在毛刺情况下的探测模型的安全性。为此,Knichel 等人构建了可以实现安全组合的两个输入的或门电路和与门电路,并且展示了如何利用这两个门电路实现任意数量输入的安全电路组合。虽然实现任意数量的安全电路会消耗更多的新的随机掩码,但是这些随机掩码可以被其他电路组件再次利用。这证明了该方法可以在保证安全性和组合性的情况下,大大减少随机数资源的消耗,并且利用 SILVER 和 TVLA 检测手段进行了验证。此外,Beyne^[18]等人结合二阶掩码方法和针对高阶阈值实现的评估框架,提出了两个掩码技术。这两个掩码技术降低了二阶掩码实现所需要的分摊数和随机数资源。其中,实现二阶安全性的分摊数降到了最小值 3。在噪声探测模型的安全保障下,结合上述两种掩码技术,可以提供 LED、MIDORI、SKINNY 和 PRINCE 的二阶硬件掩码实现(基于 FPGA),这些实现不需要在每个时钟周期都刷新随机掩码,加上其分摊数降低到了最小,因此大大降低了成本和计算延迟。

线性码掩码是当前最先进最通用的掩码方案,通过指定不同的生成矩阵,它可以实例化为当前常用的编码方案,如布尔掩码、内积掩码等。最重要的是,它可以减少掩码实现的信息泄露,并且有望抵御故障注入攻击。2020 年出现了基于线性码掩码的通用编码器和私有计算方案,并且提出了一种称为“成本分摊”的方法,从而减少掩码实现的成本。Wu^[19]等人在这个基础上进行了一些改进和研究。首先,设计了一种新的编码器。该编码器的生成矩阵因其稀疏特征可以大大降低计算消耗,这意味着新设计的编码器可以在降低大量计算成本的情况下保留原线性编码器的通用性、安全性和抵御故障攻击的潜力,并且兼容“成本分摊”方案。在优化编码器的基础上改善了掩码的私有计算方案,去除了一些原有方案中的冗余操作,并且将成本较高的矩阵乘法优化为等价的矩阵联结操作,再次在保证安全性的情况下大大降低了计算成本。此外,还提供了实际的掩码实现结果,以 AES 为例,给出了完整算法所需要的实际时钟周期数,为基于线性码掩码的高效实现提供了一条基础线。

Antonio^[20]等人提出了一个名为“ABE Squared”的框架,用于准确地评估和比较基于属性加密(Attribute-Based Encryption, ABE)方案的效率。在传统评估方法的基础上,ABE Squared 考虑了多个优化层次对方案实现的影响,包括使用现代密码库提供的优化算法、选择椭圆曲线、计算顺序以及在选定曲线上实例化方案等。同时,Antonio 等人指出不同的设计目标会影响优化方法,如解密算法的优化。为了使比较更加公平,ABE Squared 还引入了手动启发

式类型转换技术，并通过实现多种方案进行了验证。实验结果表明，不同的设计目标会导致不同的优化方法和方案效率，与传统评估方法的结果是有所不同的。

本节作者：张帆、吴茜眉、杨博麟、徐国瑞（浙江大学）

参考文献

[1] MARSHALL B, PAGE D, WEBB J. Miracle: Micro-architectural leakage evaluation[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 175-220.

[2] CASAGRANDE M, LOSIOUK E, CONTI M, et al. BreakMi: Reversing, Exploiting and Fixing Xiaomi Fitness Tracking Ecosystem[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022: 330-366.

[3] SCHNEIDER M, DHAR A, PUDDU I, et al. Composite Enclaves: Towards Disaggregated Trusted Execution[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 630-656.

[4] CARRIER K, DEBRIS-ALAZARD T, MEYER-HILFIGER C, et al. Statistical Decoding 2.0: Reducing Decoding to LPN[C]. Advances in Cryptology-ASIACRYPT 2022: ASIACRYPT 2022, Taipei, Taiwan, December 5-9, 2022. Cham: Springer Nature Switzerland, 2023: 477-507.

[5] BEEREL P, GEORGIOU M, HAMLIN B, et al. Towards a Formal Treatment of Logic Locking [J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(2): 92-114.

[6] GARB K, XHEMRISHI M, KÜRZINGER L, et al. The Wiretap Channel for Capacitive PUF-Based Security Enclosures [J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(3): 165-191.

[7] DERIC A, HOLCOMB D. Know Time to Die-Integrity Checking for Zero Trust Chiplet-based Systems Using Between-Die Delay PUFs[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(3): 391-412.

[8] UDUGAMA B, JAYASINGHE D, SAADAT H, et al. VITI: A tiny self-calibrating sensor for power-variation measurement in FPGAs[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 657-678.

[9] SELMKE B, POLLANKA M, DUENSING A, et al. On the application of Two-Photon Absorption for Laser Fault Injection Attacks: Pushing the physical boundaries for Laser-based Fault Injection [J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 862-885.

[10] BARTKEWITZ T, BETTENDORF S, MOOS T, et al. Beware of Insufficient Redundancy: An Experimental Evaluation of Code-based FI Countermeasures[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(3): 438-462.

[11] BARUA A, AI FARUQUE M A. PreMSat: Preventing Magnetic Saturation Attack on Hall Sensors[J]. IACR Trans. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(4): 438-462.

[12] GÖTTE J S, SCHEUERMANN B. Can't Touch This: Inertial HSMs Thwart Advanced Physical Attacks[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 69-93.

[13] ZHANG F, FENG T, LI Z, et al. Free Fault Leakages for Deep Exploitation: Algebraic Persistent Fault Analysis on Lightweight Block Ciphers[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(2): 289-311.

[14] SOLEIMANT H, BAGHERI N, HADIPOUR H, et al. Practical multiple persistent faults analysis[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 367-390.

[15] MÜLLER N, KNICHEL D, SASDRICH P, et al. Transitional Leakage in Theory and Practice: Unveiling Security Flaws in Masked Circuits[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(2): 266-288.

[16] KNICHEL D, SASDRICH P, MORADI A. Generic Hardware Private Circuits Towards Automated Generation of Composable Secure Gadgets[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 323-344.

[17] KNICHEL D, MORADI A. Composable Gadgets with Reused Fresh Masks: First-Order Probing-Secure Hardware Circuits with only 6 Fresh Masks[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(3): 114-140.

[18] BEYNE T, DHOOGHE S, MORADI A, et al. Cryptanalysis of Efficient Masked Ciphers: Applications to Low Latency[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(1): 679-721.

[19] WU Q, CHENG W, GUILLEY S, et al. On Efficient and Secure Code-based Masking: A Pragmatic Evaluation[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(3): 192-222.

[20] DE LA PIEDRA A, VENEMA M, ALPÁR G. ABE Squared: Accurately Benchmarking Efficiency of Attribute-Based Encryption[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022(2): 192-239.

第二部分

全同态加密

全同态加密应用发展报告*

洪澄¹, 汤殿华^{2,3}, 郑中翔⁴

1. 蚂蚁集团, 杭州, 311121

2. 保密通信全国重点实验室, 成都, 610041

3. 中国电子科技集团公司第三十研究所, 成都, 610041

4. 中国传媒大学, 北京, 100024

通讯作者: 汤殿华, E-mail: tangdianhua86@163.com

摘要: 自 2009 年首次提出以来, 全同态加密算法已取得了长足的发展。随着一系列在安全性归约、算法结构以及编码技术等方面的改进相继提出, 全同态加密逐渐具备了在特定场景中解决数据安全问题的能力。本文针对全同态加密应用中的典型场景, 将其归纳为以下几个方面: 全同态加密在机器学习中的应用、全同态加密在隐私信息检索中的应用及全同态加密在隐私集合求交中的应用, 并对近年来全同态加密在上述应用的解决方案进行了总结与讨论。

关键词: 全同态加密; 机器学习; 隐私信息检索; 隐私集合求交

Report on the Development of Fully Homomorphic Encryption Applications

HONG Cheng¹, TANG Dianhua^{2,3}, ZHENG Zhongxiang⁴

1. Ant Group, Hangzhou, 311121, China

2. National Key Laboratory of Security Communication, Chengdu, 610041, China

3. The 30th Research Institute of China Electronic Technology Group Corporation, Chengdu, 610041, China

* 基金项目: 国家重点研发计划(2023YFB3106200); 国家自然科学基金(62202437, 12271306, 62272040, 62172005)。

4. Communication University of China, Beijing, 100024, China

Corresponding author: TANG Dianhua, E-mail: tangdianhua86@163.com

Abstract: Since it was first proposed in 2009, fully homomorphic encryption algorithm has made great progress. With a series of improvements in security reduction, algorithm structure and encoding technique, full homomorphic encryption now has the ability to solve data security problems in specific scenarios. In this paper, we focus on the typical scenarios in the application of full homomorphic encryption in recent years, which are summarized into the following classical problems, i.e., the application of full homomorphic encryption in machine learning, the application of full homomorphic encryption in the retrieval of private information and the application of full homomorphic encryption in the intersection of private sets, and make summarization and discussion about these solutions.

Keywords: Full Homomorphic Encryption; Machine Learning; Private Information Retrieval; Private Set Intersection.

1 全同态加密应用发展概述

随着人工智能、区块链、隐私计算等新技术的发展和新应用的出现,数据从传统的“静态”保护模式逐步向“动态”保护模式过渡,如在多个参与方动态计算数据的场景下保护各自数据的隐私性。数据安全需求也呈现出跨学科融合趋势,例如,人工智能模型训练和预测过程中数据的隐私性和完整性需求、大规模分布式架构中的数据隐私保护以及可扩展性需求。这些需求一方面对全同态加密的实际应用提出了新的挑战,另一方面也牵引着全同态加密应用的研究向着实用化不断前进和发展。

2009年,第一代全同态加密算法的提出标志着全同态加密从概念走向实际^[1],而使用该算法执行密文计算的效率较明文计算的效率慢约10亿(10^9)分之一,在这样巨大的额外开销下,所有应用的讨论有如天方夜谭。而此后的十余年间,全同态加密迎来了高速发展时期,针对全同态加密算法的设计与优化取得了显著的进展。

其中较为典型的优化技术包括:Brakerski等人提出的模数转换技术(Modulus Switching Technique)能够显著降低全同态加密的密文计算过程中错误增长的速度^[2-3],从而令全同态加密算法在参数选取方面更加灵活;Gentry等人提出的非对称的乘法技术有效地控制了密文乘法运算中的错误增长速度,为第三代全同态加密算法的高效自举计算奠定了基础^[4];Cheon等人将LWE问题中的错误作为计算误差进行处理,通过计算近似结果大幅提高算法的效率和参

数选取的灵活度等^[5]，这些优化技术在全同态加密算法的实用化过程中均发挥了重要的作用，为全同态加密的应用提供支撑作用。

1. 外包存储及外包计算

同态加密最直接的应用场景是外包计算。该应用中，云端提供大规模存储和高性能计算服务，客户端（如小型企业）将私有数据以同态加密形式保存在云端。云端可利用自身的计算能力和同态加密性质直接对这些密文数据执行操作并将密文结果返回给客户端，客户端对密文解密即可获得需要的计算结果。这里，同态加密算法提供了一种简洁的云计算安全解决方案，既保护了云上数据的安全性，又使云平台具备了对云上数据操作的能力。

2. 机器学习

机器学习方法要求输入大量数据以训练高质量模型，当数据涉及个人隐私或网络安全问题时，可以利用全同态加密通过密文计算的方式实施模型训练任务，在保证数据安全的前提下获取高质量模型。在该场景中，所需计算量较大，因此对全同态加密的密文计算效率有较高的需求。

3. 物联网和边缘计算

在物联网场景中，照明设备、安防监控、智能空调等诸多智能家居通过各自独立的传感器和芯片收集、处理用户数据，并可通过边缘计算部署的云平台实现协同。在这一过程中，智能设备收集的数据涵盖个人隐私，可通过全同态加密的方法实现对数据安全的保护。在该场景中，受限于 IoT 设备的计算能力和带宽限制，同态加密的计算效率和密文大小都是应用部署过程中需要解决的挑战。

4. 云计算

在云计算场景中，数据集中在云存储服务器上，用户可以通过指令使用云平台提供的计算和存储资源完成各项任务。对于单用户访问情况而言，云计算场景与外包计算类似，可以直接使用全同态加密技术实现对用户数据的隐私保护。而当涉及多用户场景时，例如，需要使用不同的用户数据共同完成某个计算任务时，需要完成不同密钥加密的数据之间的密文计算任务，此时不能直接使用基本同态加密方案，而需要采用代理重加密同态算法或多密钥同态算法。

通过以上几个典型场景中的案例可知，全同态加密的应用需要结合场景的具体需求，采取相应的优化技术和方法，如此才能够实现全同态加密的应用落地，完成实用化的最后一公里。在这些典型场景中，全同态加密算法的应用可以归纳为以下几个方面：全同态加密在机器学习中的应用、全同态加密在隐私信息检索中的应用以及全同态加密在隐私集合求交中的应用。

2 全同态加密的隐私保护机器学习应用

2.1 基于 BFV/BGV 类方案

微软研究院提出的 CryptoNets^[6]可以说是基于全同态加密来做隐私保护神经网络的开山之作。他们使用 YASHE 同态加密算法对图片数据进行加密，并用一个 5 层明文卷积神经网络（CNN）模型对加密图片执行了推理。为了降低同态计算的乘法深度，他们使用平方函数代替 CNN 中的 ReLU 激活函数（这要求明文模型是使用平方激活函数训练的）；类似地，他们也要求模型的池化层是基于 average pooling 而不是 max pooling 训练的。CryptoNets 的两个创新点包括：①基于中国剩余定理（CRT）使用 5 个密文来编码一份明文，从而可以支持 5 倍的明文空间长度；②使用基于 CRT 的打包（Packing）机制来支持同时对多张加密图片进行推理；CryptoNets 最终可以在 200 多秒内完成 4096 张 MNIST 加密图片的推理并且可以达到 99% 的准确率。

相比神经网络的推理，在同态加密的数据上进行神经网络训练要困难得多。文献[7]最早尝试实现这一工作：他们使用 BGV 算法逐个比特的加密 MNIST 数据集，然后在其上训练一个由 3 个全连接层、10 万多个参数组成的简单模型；模型使用的激活函数是 Sigmoid，最后分类的 loss 函数使用二次函数近似。值得一提的是，他们使用查找表而不是多项式近似来实现非线性运算，这是因为查找表更适合逐比特加密的输入。本文的创新点之一是将多张图片的同一个维度的数据打包在同一个密文内，以并行化的方式同时支持多张图片的训练。最终单线程训练一个 minibatch 60 张图片耗时为 1.5 天。该工作还尝试了一个更小的 2700 个参数的三层神经网络，并使用 30 个线程，这时训练一个 minibatch 的耗时可以压缩到 40min。

2.2 基于 TFHE 的方案

早期关于隐私模型预测的方案（如 CryptoNets 等）利用的都是 BFV/BGV 类的层次同态加密算法（leveled-FHE）。但由于层次同态加密的特性，这些方案能够进行的同态运算深度非常有限，且参数需要选得较大。与之对比，TFHE 算法的特点是能用较小的参数实现快速的自举。文献[19]是最早使用 TFHE 算法实现隐私保护的机器学习应用的论文之一，其在 TFHE 算法以及神经网络模型上分别提出了一些创新和改进。

文献[19]首先基于 TFHE 适合加密整数或用整数表示的定点数的特点，提出了离散化的深度神经网络（Discretized Neural Network, DiNN）。在这样的网络中，输入以及参数都用固定大小范围的整数来表示。为了契合 TFHE 算法的特点，输入值的范围选择了 $\{-1, 1\}$ 。而参数值则

通过对现成的、实数值的网络的参数围绕固定值的倍数取整得来。在激活函数方面，此方案使用了符号函数 Sgn。作者在 MNIST 数据集上分别训练了包含 30/100 个神经元的单隐层网络进行了验证，证明了离散化后的网络的训练结果与原始网络相比，准确度只有轻微的下降，其中在包含 30 个神经元的网络上仅损失了 1% 左右的准确度，在包含 100 个神经元的网络上准确度约下降了 0.3%。

在 TFHE 算法方面，文献[19]的作者提出了一些实用的改进，包括：①使用 LWE 加密形式来进行初始的数据传输步骤以减小通信量；②提出了 BlindRotate 算法的一个新的实现方式，在密钥数量增加 50% 的情况下将计算复杂度降低了约一半。利用这些改进，文献[19]对 MNIST 套用了上述两个网络进行同态推理。实验结果显示同态推理的准确度与明文上的推理准确度极为接近，且耗时基本上都在 1s 左右。

除了文献[19]，TAPAS 同样也是使用 TFHE 全同态加密算法进行模型预测的方案^[20]，其在文献[19]之后不久提出。TAPAS 方案针对的是二元神经网络（Binary Neural Network, BNN），其与文献[19]中的 DiNN 类似，其中输入值和参数值都为整数，只不过在 BNN 中不管输入值还是权重值（Weight）都选自 $\{-1, 1\}$ ，只有偏置值（Bias）是任意整数。在该方案中，作者提出了一系列技巧用于提升同态模型推理的效率，包括两种针对二元向量的内积的电路：归约树和排序网络，以及一种用于将同态运算稀疏化的“+1 技巧”。前者是两种亚线性深度的逻辑电路，后者通过将权重向量 w 与 1 向量相加，可以消去其中约占一半数量的 -1 项，使需要同态运算量能大致减半。文献[20]对包括 MNIST 在内的一些常见数据集进行了实验。由于该方案中的同态加密运算是对于二进制值的精确运算，因此准确度与明文下的 BNN 模型相同。而在明文下对于同样的数据集，BNN 模型与使用浮点数的模型的准确度差别微小。然而在效率方面，TAPAS 方案的运行耗时较长，部分数据集即使在多线程下进行完整的同态预测，也仍需要小时级别的耗时。

2.3 基于 CKKS 的方案

CKKS 横空出世之后，全同态应用的性能得到非常大的改进。让 CKKS 声名鹊起的是 2017 年的 iDASH 竞赛：首尔大学团队基于 CKKS 的方案^[13]仅用了 6min 就完成了在 1579 行 18 列基因数据上的逻辑回归（Logistic Regression）训练，并达到了预期的准确率，取得了同态加密赛道的冠军；相比之下，基于 BFV/BGV 的方案在同等情况下的耗时则达数十分钟乃至数百分钟。文献[13]基于逻辑回归的特性，巧妙地将多行数据表的特征和标签一起编码到 CKKS 密文内，并使用 3~7 阶多项式来近似 Sigmoid 函数。它们使用的参数是 $n=32768$ ， $q=1284\text{bit}$ ，这样可以支持 40 多个乘法深度，可以完成 7~9 次梯度下降迭代，足以将模型训练到足够的精度。

第二年 iDASH2018 的同态赛题是求多个 SNP (Single Nucleotide Poly-morphism) 的 p -value: 这实际上是要求同时训练多个模型, 如果还是简单地逐个模型训练, 其耗时必然是不可接受的。UCSD 团队 (主力依然是从首尔大学过去的) 基于 CKKS 设计了 SIMD 并行训练多个模型的方案^[14], 蝉联赛道冠军。他们在普通的四核笔记本上仅花费数十秒就完成了 245 行 10643 列 SNP 数据的建模。值得一提的是, 这一年包括 IBM、Duality 等在内的知名参赛队伍都选择了 CKKS 算法而非 BFV/BGV 算法。这也显示了 CKKS 在机器学习这类可容忍误差的应用中的巨大性能优势。

2.4 基于 PBS 的方案

Zama 是一家专注于 TFHE 算法相关技术的公司。文献[8-9]都是 Zama 团队的工作, 分别利用可编程的自举 (Programmable Boot Strapping, PBS) 的优势来实现对神经网络模型和基于树结构的机器学习模型的同态预测。众所周知, 现有的全同态加密方案全部依赖自举技术来实现不限深度的同态运算。而在此基础上, TFHE 算法能够支持可编程的自举 PBS, 即其自举结果所加密的是原来消息在某个函数下的取值, 且这个函数是可以任意选定的。因此, 使用 PBS 可以很好地处理机器学习的一些复杂操作。

文献[9]针对的是神经网络模型, 其与文献[19]的主要区别在于该工作没有针对同态加密来训练特殊的模型, 其只对训练好的模型进行必要的离散化。文中将神经网络推理需要的操作分为无须 PBS 的层 (如卷积层、全局平均池层等) 及需要进行 PBS 的层 [包括激活函数层 (如 ReLU)、最大池层等]。后面这类操作对于别的同态加密算法来说往往需要使用烦琐的多项式拟合来实现, 且容易受到层数的限制, 而 TFHE 算法的 PBS 则恰好能避免这些麻烦。文献[9]在 MNIST 数据集上分别训练了 20 层、50 层、100 层的神经网络 NN-20、NN-50、NN-100, 这些网络每个都包含卷积层、激活函数层和稠密层, 且每个隐藏层至少包含 92 个神经元。随后其使用 TFHE 方案对它们进行同态推理。实验结果显示, NN-20 的神经网络在 AWS2 云服务器上最快仅需不到 1s 即可完成同态推理 (其中安全级别为 80bit)。而在使用 128bit 安全级别的参数时, 这三个网络分别仅需约 8s、19s 和 38s。在前两个网络上, 同态推理的准确度都能达 90%以上且与明文下的推理准确度接近。而 NN-100 的同态推理准确度比较低, 不同参数下得到的准确度为 83%~91%。

与前面几项关于神经网络的工作不同, 文献[8]针对基于树结构的机器学习模型, 如决策树、随机森林和 XGBoost。其研究如何将基于树的模型转化为同态加密所兼容的运算, 进而实现这些模型在密文状态下的同态推理。在使用同态加密对树模型进行预测时, 容易遇到一些限制, 如密文包含的消息精度有限、难以实现条件操作等。针对这些限制, 作者对模型进行了与之前工作类似的离散化, 同时将条件操作转化为输出 0/1 值的函数从而能用 PBS 对其求值。

根据作者的实验,全同态加密下的同态模型预测的耗时基本在 10s 以内,有些实例甚至在 1s 以内。而在准确度方面,同态模型预测所达到的准确度与使用 32 位浮点数的明文准确度极为接近。

2.5 基于 CKKS+Bootstrapping 的方案

随着 RNS-CKKS 等 CKKS 算法方面的不断优化,以及 CKKS Bootstrapping 技术的不断改进,首尔大学首次实现了在全同态加密的数据上运行 ResNet110 级别的大规模 CNN 推理^[11]。他们为图片数据设计了更紧凑更适合 Bootstrapping 的多路 Packing 编码,使用非常高(15~27)阶的多项式来近似 ReLU,并提出了新的误差更低的 Bootstrapping 技巧,最终他们在 Cifar-10 上进行一次 ResNet110 推理的耗时约为 13000s(其中 75%的时间花在 Bootstrapping 上),且准确率与明文相差无几。

HyPHEN 则剑走偏锋,使用了另一套解决方案^[15]:他们尽可能地将模型中的 ReLU 使用低(2-3)阶多项式来替换,并使用 GPU 对整体的 CKKS 计算进行加速,最终仅花费 2s 左右即可在 Cifar-10 上进行一次 ResNet32 推理(其中约 50%的时间花在 Bootstrapping 上),已经接近实用了。这类方案的主要缺点是可扩展性不高:可能在某些数据集和某些网络上确实可以付出很大代价搜索到这种低阶多项式替换,但是在其他数据集和其他网络上并不一定能通用。

2.6 混合同态加密方案

由上述几节可知,目前基本上只有 CKKS 算法能够支持大规模的机器学习应用,而 CKKS 的主要问题是只能支持加法和乘法,也就是多项式操作,难以实现 ReLU 这类非多项式操作。反观 TFHE、FHEW 等方案能够通过查找表、布尔电路等方法实现任意非多项式操作,但是其加法和乘法又无法像 CKKS 等方案那样天然地支持基于 CRT 的 Packing 编码,导致性能较低。有没有办法同时取二者的长处? Chimera 是首个提出“混合同态加密”思想的方案^[18]:他们设计了 TRLWE、TGSW 等新的数据格式来支持密文在 TFHE、CKKS、BFV 之间互相切换,在需要计算除法这类复杂操作时使用 TFHE,而在需要计算矩阵乘这种 SIMD 乘法时切换到 CKKS。Chimera 方案的主要问题是性能较低,密钥体积较大。他们使用这一方案参加了 iDASH18 的竞赛,但是其性能并不理想,未能在与纯 CKKS 的方案比赛中胜出。

在 Chimera^[18]的基础上,阿里巴巴进一步提出了 Pegasus(飞马)方案^[10],大大提升了混合同态加密方案的性能。Pegasus 并没有使用 Chimera 的 TRLWE 这类中间格式,而是直接在 RLWE 和 LWE 形态之间来回转换,并使用近似的 FHEW 查找表来在 LWE 密文上计算非多项式函数,这种方案在支持较大的明文范围方面比 Chimera 快得多(代价是计算结果是阶梯型近

似的)。以机器学习中常用的操作 Sigmoid 为例, Pegasus 计算单个 Sigmoid 只需要秒级, 而同等情况下 Chimera 需要分钟级。Pegasus 还设计了新的基于线性变换的 RLWE、RLWE 转换方案, 其转换密钥只需要 GB 级, 而同等参数下 Chimera 需要百 GB 级。Pegasus 还实现了 Kmeans、决策树等算法, 并在多个数据集上进行了测试, 展示了近似查找表对机器学习/统计类应用的准确率确实存在, 但是影响不大(不超过 3%)。

HEDA 是混合同态加密在数据库上的一个典型应用^[12]。此前的加密数据库是很难支持形如 `SELECT SUM(salary) FROM emp WHERE salary BETWEEN 5000 AND 6000` 这种 SQL 操作的, 而有了混合同态加密之后就可以相对高效地解决这一问题: HEDA 首先将 salary 数据使用 BFV 编码方式进行 RLWE 加密, 在需要判断区间(如是否大于 5000)时减去 5000 然后转换到 LWE 密文, 再使用与 Pegasus 类似的查找表方法来提取 LWE 密文的正负号, 进而转换回 RLWE 密文与 salary 密文进行密文内积, 得到最终的 sum 结果。与 Pegasus 不同的是, HEDA 为了回到 BFV 编码, 其 LWE 与 RLWE 转换使用的是陈昊等人的方案^[21], 而不是 CKKS Bootstrapping。

2.7 同态与安全多方计算结合的方案

还有很多工作并不是仅使用全同态加密技术, 而是将全同态加密技术与秘密共享、不经意传输等安全多方计算(MPC)协议结合, 构造交互式的隐私保护解决方案。这类方案一般只需要消耗一次同态乘法, 不需要考虑自举(Bootstrapping)操作, 因此往往比纯 FHE 的解决方案更为高效。HE-MPC 交互式方案付出的代价是各个参与方之间需要多轮通信, 且通信量远大于纯 FHE 方案, 因此要求参与方具有较好的网络条件, 在整个计算过程中能够一直在线参与。典型的工作有 Gazelle^[22]、CryptFlow2^[23]、Cheetah^[16]等, 它们的创新点很多都与具体的 MPC 协议相关, 因此本文不一一介绍。

3 全同态加密的隐私信息检索应用

3.1 隐私信息检索概述

隐私信息检索(Private Information Retrieval, PIR)是一类特定用途的密码协议, 它允许客户端在服务器 Server 的数据库 DB 中查询索引为 index 的某条记录 DB_{index} , 但不向数据库服务器泄露哪条记录被查询。隐私信息检索是隐私保护中重要的关键技术, 可应用于数据库隐匿查询、匿名信息传递、联系人发现、联系人追踪、隐私保护导航、安全浏览等场景。

1995 年, Chor 等人^[24]在首次提出隐私信息检索概念时, 将其分为两类: 信息理论安全隐私信息检索 (Information Theoretic PIR, IT-PIR); 计算安全隐私信息检索 (Computational PIR, CPIR)。信息理论安全 PIR 协议一般将数据库复制存放在多个不共谋的服务器上, 也称为多服务器 PIR 协议; 计算安全 PIR 协议数据库存放在单一服务器, 并以数学困难问题为基础来构造, 也称为单服务器 PIR 协议。基于全同态加密构造的隐私信息检索协议属于计算安全隐私信息检索协议。

衡量隐私信息检索协议的性能主要有三个指标: 请求尺寸、响应尺寸和服务器计算开销。一个 PIR 方案很难同时将这三个指标做到很小, 不同设计方法往往通过不同方法实现了三个指标的权衡。一般来说, PIR 综合开销是较大的, 其本质原因在于: 对于来自客户端的每个询问, 数据库中的所有记录都需要参与计算, 以生成相应的响应。假设在生成响应结果时, 数据库中某些记录不参与运算, 那么说明客户端查询的目标不在这些记录中, 将泄露了一定的信息。

全同态加密具有密文计算能力, 能够在查询、数据库记录上执行密文计算, 并且计算结果也是加密的, 这非常适合用于设计计算安全隐私信息检索协议, 并且能够有效降低交互轮数、通信量。2009 年, Gentry 在他的博士论文^[1]中首次基于全同态加密方案构造了 PIR 协议, 实现了亚线性的通信复杂度, 但计算复杂度较高, 需要的同态乘法深度也较大。之后, 基于全同态加密方案的 PIR 协议方案得到较快的发展。

3.2 国内外发展现状

2009 年, Gentry 提出了首个基于全同态加密的 PIR 协议^[1]。该协议将查询索引逐比特同态加密, 然后把这些同态密文发送给服务器, 服务器通过同态计算一个布尔电路来获取查询与数据库索引的加密匹配值 (若相等匹配, 则为 1 的加密; 否则为 0 的加密), 再对每条数据记录同态乘以相应的加密匹配值, 最后进行同态加法求和, 得到响应结果发送给客户端, 客户端解密出响应结果。该协议相比之前的方案更为简洁, 并实现了亚线性通信复杂度, 但因逐比特同态计算导致执行效率非常低。2011 年, Brakerski 等人^[2]设计基于 LWE 全同态加密方案时, 融合对称密码和全同态加密, 提出了基于 LWE 的 PIR 协议, 对于每个单比特询问, 通信复杂度为 $k \cdot \text{polylog}(k) + \log_2 |\text{DB}|$, 其中 k 为安全参数。该协议使用对称密码算法来加密查询索引, 并使用同态加密方案加密对称密钥, 并发送给服务器, 然后服务器同态计算对称密码的解密算法和 Gentry 的 PIR 方案中计算函数^[1], 得到同态加密的查询结果, 客户端用同态加密方案的私钥解密获得查询结果。

2012—2016 年, 全同态加密的效率得到了较大的提升, 尤其基于 RLWE 构造的 BGV^[3]、BFV^[25]等方案性能表现出色。这也为 PIR 协议实用化设计带来了新希望。为了保证计算安全隐

私信息检索协议的实用性,2016年,Melchor等人^[26]采用基于RLWE的同态加密方案,设计一个高效可用的隐私信息检索协议XPIR,并基于HElib库实现了该协议。该方案不使用指标索引作为询问,而是用维数与数据库规模相同的一个汉明重量为1的 $\{0,1\}$ 向量作为询问,并使用同态加密按分量加密该询问向量,服务器执行密文询问向量与数据库记录的内积,得到查询结果。

XPIR方案询问为一个密文向量,导致询问通信开销较大。2018年,Angel等人^[27]基于BFV同态加密方案设计了询问尺寸更小的隐私信息检索协议SealPIR。相比XPIR方案,询问尺寸被压缩到原有的 $1/274$,并降低了加密询问生成时间、响应结果的解密时间。SealPIR协议将询问的指标索引index编码为单项式 x^{index} ,然后对其执行同态加密得到密文询问。数据库服务器在执行询问查询时,需要额外引入精巧的扩展算法,将密文询问扩展成类似于XPIR协议的密文询问向量。SealPIR在含有1048576条记录(每条记录288字节)的数据库中,执行1次查询,总耗时为6.505s,通信开销为320KB。另外,SealPIR协议采用概率性批量编码技术,实现了多询问PIR方案,对查询计算实现了40倍加速。

表1所示为XPIR协议与SealPIR协议的性能比较。

表1 XPIR协议与SealPIR协议的性能比较

协议方案		XPIR			SealPIR		
数据库规模		65536	262144	1048576	65536	262144	1048576
客户端 计算开销	密文询问生成	13.83ms	27.57ms	55.14ms	3.37ms	3.37ms	3.37ms
	响应结果提取	0.34ms	0.29ms	0.30ms	1.37ms	1.39ms	1.69ms
服务器 计算开销	数据库编码	0.15s	0.57s	2.27s	0.23s	1.04s	4.26s
	扩展算法	—	—	—	0.05s	0.11s	0.23s
	密态内积	0.21s	0.63s	2.12s	0.13s	0.50s	2.01s
通信开销	询问尺寸	4384KB	8768KB	17536KB	64KB	64KB	64KB
	响应尺寸	256KB	256KB	256KB	256KB	256KB	256KB

注:数据库每条记录大小为288B;服务器配置:处理器为16核3.6GHz Intel Xeon E5-2667,内存为112GB;客户端配置:12核2.4GHz Intel Xeon E5-2673,内存为32GB。

2019年,Gentry等人^[28]提出了可压缩的全同态加密方案,并将其用于构造隐私信息检索协议,从理论上获得了密文响应与记录之间的比率接近于1的PIR协议。2022年,Park等人^[29]基于TFHE全同态加密和查找表电路,设计了SHECS-PIR方案,将通信开销降低到 $O(\log_2 n)$,其中 n 为数据库记录的数量。

SealPIR 方案使得询问尺寸得到大幅压缩,但其响应尺寸较大、密态查询计算开销也较高。为了缩短响应尺寸,2021 年 Mughees 等人^[30]在 SealPIR 基本框架下设计了一个响应高效的单服务器隐私信息查询协议 OnionPIR。该协议充分利用 RLWE 密文与 RGSW 密文的外积,具备了噪声增长率小的优势,使得方案的参数设置会更小。在计算开销差不多的情况下 OnionPIR 协议将响应的明密文扩展比由 100 降低到 4.2,这使得通信开销得到大幅降低。例如,在规模为 1048576 的数据库上查询 30KB 的数据库记录,SealPIR 需要返回 3200KB 的密文响应,服务器计算耗时为 403s,而 OnionPIR 仅需要返回 128KB,服务器计算耗时为 400.9s。Mughees 等人通过开发同态计算复制网络的新技术,设计了带状态的 OnionPIR,在规模为 $2^{16}\sim 2^{24}$ 的数据库上,计算效率提升了 1.8~22 倍,响应尺寸降低了 1/3900~1/27。

表 2 所示为 SealPIR 协议与 OnionPIR 协议的性能比较。

表 2 SealPIR 协议与 OnionPIR 协议的性能比较

协议方案		SealPIR			OnionPIR		
数据库规模		65536	262144	1048576	65536	262144	1048576
服务器 计算开销	扩展算法	5.4s	10.7s	21.5s	3.6s	4.1s	4.6s
	密态内积	20.7s	91.2s	381.6s	21.3s	97.0s	396.3s
	总计算耗时	26.1s	101.9s	403.1s	24.9s	101.1s	400.9s
通信开销	询问尺寸	32KB	32KB	32KB	64KB	64KB	64KB
	响应尺寸	3200KB	3200KB	3200KB	128KB	128KB	128KB

注:数据库每条记录的大小为 30KB,配置环境为亚马逊 EC2 云服务器实例(具有 32GB 和 8 个 CPU 核的 t2.2xlarge 实例)。

2021 年, Ahmad 等人^[31]在设计元数据隐私保护语音通信系统时,基于 BFV 同态加密方案的密文旋转算法提出了一个隐私信息检索协议 FastPIR。相比 XPIR 和 SealPIR 协议, FastPIR 协议具有更短的服务器计算耗时和更小的响应尺寸。

基于全同态加密方案构造 PIR 协议,往往在不同指标之间权衡,很难同时在多个指标上做得很好。SealPIR 方案压缩了询问尺寸, OnionPIR 方案在计算开销差不多的情况下压缩了响应尺寸。其实由于不同数据库的记录比特长度不同,因此通过响应尺寸、规模尺寸、计算耗时而来判断方案效率,就存在一定的不合理性。应该将记录比特长度、数据规模作为考虑因素,例如,将数据库数据总量与计算耗时的比值称为服务器吞吐率,将响应尺寸与实际记录的尺寸的比值称为响应占比。2022 年, Menon 等人^[32]融合应用两种格基同态加密方案,设计了快速且高占比的单服务器隐私信息检索协议族,主要包括 SPIRAL、SPIRALSTREM 和 SPIRALPACK 方案。SPIRAL 协议族使用了矩阵型 Regev 方案和 GSW 方案,并引入了新型的密文转换方法,

同时实现了服务器吞吐率、响应占比这两个指标的提升。OnionPIR 方案的服务器吞吐率为 149MB/s, 响应占比为 0.24; SPIRALSTREM 方案的服务器吞吐率为 1.5GB/s, 响应占比为 0.49。SPIRALPACK 方案在大参数情况下服务器吞吐率可达到 1.9GB/s, 响应占比可达到 0.81。文献[32]给出了 SealPIR、FastPIR、OnionPIR、SPIRAL、SPIRALSTREM 等性能比较, 如表 3 所示。

表 3 基于全同态加密的 PIR 协议方案性能比较

数据库规模	性能项	SealPIR	FastPIR	OnionPIR	SPIRAL	SPIRALSTREM
$2^{20} \times 256\text{KB}$	询问尺寸	66KB	33MB	63KB	14KB	8MB
	响应尺寸	328KB	66KB	127KB	21KB	20KB
	计算耗时	3.19s	1.44s	3.31s	1.69s	0.85s
	响应占比	0.0008	0.0039	0.0020	0.0122	0.0125
	吞吐率	85MB/s	186MB/s	81MB/s	159MB/s	314MB/s
$2^{18} \times 30\text{KB}$	询问尺寸	66KB	8MB	63KB	14KB	15MB
	响应尺寸	3MB	262KB	127KB	84KB	62KB
	计算耗时	74.91s	50.52s	52.73s	24.46s	8.99s
	响应占比	0.0092	0.1144	0.2363	0.3575	0.4803
	吞吐率	105MB/s	156MB/s	149MB/s	322MB/s	875MB/s
$2^{14} \times 100\text{KB}$	询问尺寸	66KB	524KB	63KB	14KB	8MB
	响应尺寸	11MB	721KB	508KB	242KB	208KB
	计算耗时	19.03s	23.27s	14.38s	4.92s	2.38s
	响应占比	0.0092	0.1387	0.1969	0.4129	0.4811
	吞吐率	86MB/s	70MB/s	114MB/s	333MB/s	688MB/s

3.3 XPIR 隐私信息检索方案

目前大部分基于全同态加密构造的高效率 PIR 协议都以 XPIR 协议框架为蓝本。XPIR 协议的基本原理为:将原始询问的指标索引 index 编码为一个维度与数据库规模相同的零么向量, 其中仅有第 index 个元素为 1, 其余为 0; 客户端使用全同态加密方案加密向量的每个元素, 得到一个加密向量, 并将其作为询问发送给服务器; 服务器将该向量与数据库所有记录执行同

态内积运算, 得到加密的响应, 并发送给客户端; 客户端解密便可以得到查询结果。具体协议如下:

输入: 用户输入一个索引 $\text{idx} \in \{0, 1, \dots, n-1\}$, 数据库服务器输入一个规模为 N 的数据库 DB , 设数据库中的每个记录为 DB_i , $i = 0, 1, \dots, n-1$ 。

(1) 预处理数据库: 将数据库 DB 按照一定格式进行编码。

(2) 询问生成: 对于 $i = 0, 1, \dots, n-1$, 用户计算一组密文 $\mathbf{c} = (c_0, c_1, \dots, c_{n-1})$, 其中若 $i = \text{idx}$, 则 $c_i = \text{HE.Enc}(pk, 1)$, 若 $i \neq \text{idx}$, 则 $c_i = \text{HE.Enc}(pk, 0)$ 。用户将密文向量 $\mathbf{c}$ 作为询问发送给数据库服务器。

(3) 响应生成: 服务器使用同态加密方案中的明密文乘法算法和同态加法算法, 计算 $r = \sum_{i=0}^{n-1} c_i \cdot \text{DB}_i$, 服务器将密文 r 返回给用户。注意, 此处的求和符号 Σ 表示同态加法, 乘积符号 “ $\cdot$ ” 表示同态明密文乘法。

(4) 结果输出: 用户使用私钥 sk 解密密文 r , 便可得到数据库中 idx 所对应的记录 $\text{DB}_{\text{idx}} = \text{HE.Dec}(sk, r)$ 。

很容易验证协议正确性。可以看出密文 $\mathbf{c} = (c_0, c_1, \dots, c_{n-1})$ 是向量 $(0, 0, \dots, 0, 1, 0, \dots, 0)$ 的加密, 其中 “1” 处于向量的第 idx 个位置。因此

$$\begin{aligned} \text{HE.Dec}(sk, r) &= \text{HE.Dec}\left(sk, \sum_{i=0}^{n-1} c_i \cdot \text{DB}_i\right) = \text{HE.Dec}\left(sk, \text{HE.Enc}\left(pk, \sum_{i=0}^{n-1} b_i \cdot \text{DB}_i\right)\right) \\ &= \text{HE.Dec}(sk, \text{HE.Enc}(pk, \text{DB}_{\text{idx}})) = \text{DB}_{\text{idx}} \end{aligned}$$

3.4 SealPIR 隐私信息检索方案

XPIR 方案的询问需要 n 个同态密文, 导致询问尺寸过大。虽然将数据库折叠为一个 d 维超立方体, 询问可以降低为 $d\sqrt[n]{n}$ 个密文, 但是这导致出现 “密文乘以密文” 的同态乘法操作, 密文噪声过大。2018 年, Angel 等人^[27]提出了压缩询问的新方法, 将询问由 n 个同态密文压缩到一个同态密文, 进一步设计了更为高效的隐私信息检索方案 SealPIR。

假设客户端希望查询数据库中的第 idx 个元素, SealPIR 方案将询问加密为 $\text{HE.Enc}(x^{\text{idx}})$, 相当于将询问表示成一个次数为 idx 的单项式并加密。然后专门设计一个扩展算法 Expand, 在服务器上将 $\text{HE.Enc}(x^{\text{idx}})$ 扩展成 $\mathbf{c} = (c_0, c_1, \dots, c_{n-1})$, 使其满足: 若 $i = \text{idx}$, 则 $c_i = \text{HE.Enc}(1)$, 否则 $c_i = \text{HE.Enc}(0)$ 。扩展算法 Expand 的应用使得基于同态加密的隐私信息检索方案的效率得到大幅提升, 后续知名的 OnionPIR、SPIRAL 方案都采用这种设计方法。

SealPIR 方案以 BFV 方案为基础, 简要介绍一下该方案。令 BFV 的代数结构为 $R_q = \mathbb{Z}_q[x]/(x^N + 1)$ 。BFV 方案的公钥 $\mathbf{pk} = (pk_0, pk_1) = (b = -as + e, a)$, 私钥 $\mathbf{sk} = (1, s)$ 。令 $\Delta = \lfloor \frac{q}{t} \rfloor$, BFV 密文形式一般为 $\mathbf{ct} = (ct_0, ct_1) = (pk_2r + e_2 + \Delta m, pk_1r + e_1)$ 。解密算法表达为 $m = \lfloor \frac{t}{q} ct^T sk \rfloor \bmod t = \lfloor \frac{t}{q} (ct_0 + ct_1s) \rfloor \bmod t$ 。可以看出明文到密文的扩展率为 $\frac{2 \log_2 q}{t}$ 。

扩展算法 Expand 需要一个较为特殊的同态操作 $\text{Sub}(\mathbf{ct}, k)$: 首先将 $\mathbf{ct} = (ct_0(x), ct_1(x))$ 中的 x 替换为 x^k , 则变换为 $\mathbf{ct}' = (ct_0(x^k), ct_1(x^k))$, 此时所对应的解密密钥变化为 $s(x^k)$; 为了一致性, 再用密钥切换算法处理密文 $\mathbf{ct}'$, 得到 $\mathbf{ct}''$, 使其对应的解密密钥为 $s(x)$, 并且 $\mathbf{ct}''$ 变成对消息 $m(x^k)$ 的加密。

下面重点介绍和分析扩展算法 Expand。

扩展算法 $(c_0, c_1, \dots, c_{n-1}) \leftarrow \text{Expand}(\text{query}, n)$

输入: 询问 $\text{query} = \text{HE.Enc}(x^{\text{idx}})$, 数据库的尺寸 n 。

输出: 一组密文 $\mathbf{c} = (c_0, c_1, \dots, c_{n-1})$ 。

- (1) 找一个最小的 $m = 2^l$, 使得 $m \geq n$ 。
- (2) 用询问 query 初始化一个密文向量 $\mathbf{ciphertexts} = [\text{query}]$ 。随着算法的运行, $\mathbf{ciphertexts}$ 向量的元素个数不断增加。
- (3) 对于 $j = 0, 1, \dots, l-1$ 迭代循环运行。
 - ① 对于 $k = 0, 1, \dots, 2^j - 1$ 迭代循环运行:

$$c_0 \leftarrow \mathbf{ciphertexts}[k]$$

$$c_1 \leftarrow c_0 \cdot x^{-2^j}$$

$$c'_k \leftarrow c_0 + \text{Sub}(c_0, N / 2^j + 1)$$

$$c'_{k+2^j} \leftarrow c_1 + \text{Sub}(c_1, N / 2^j + 1)$$
 - ② 内层迭代循环运行完毕后, 设置 $\mathbf{ciphertexts} \leftarrow [c'_0, c'_1, \dots, c'_{2^{j+1}-1}]$, 继续执行下一次外层迭代循环。
- 外层迭代循环运行完毕后, 将得到 $\mathbf{ciphertexts} \leftarrow [c'_0, c'_1, \dots, c'_{2^l-1}]$ 。
- (4) $\text{inverse} \leftarrow m^{-1} \bmod t$ 。
- (5) 对于 $j = 0, 1, \dots, n-1$, 设置 $c_j = \text{inverse} \cdot \mathbf{ciphertexts}[j]$ 。

接下来我们分析 Expand 的正确性。首先给出一个定理 1, 并给出与文献[27]不同且更容易理解的证明方法。

定理 1 令 N 是一个 2 的幂, 满足 $N > n$, 询问为 $\text{query} = \text{HE.Enc}(x^{\text{id}_x})$, 则扩展算法 Expand 输出的 n 个密文 $c_0, c_1, \dots, c_{n-1}$, 对于所有的 $0 \leq k \leq n-1$, 满足以下条件。

$$c_k = \begin{cases} \text{HE.Enc}(1), & \text{如果 } k = \text{id}_x \\ \text{HE.Enc}(0), & \text{如果 } k \neq \text{id}_x \end{cases}$$

证明: 假设一个小于 n 的正整数的二进制表示为 $a = \sum_{i=0}^{l-1} 2^i a_i$ 。令 $\bar{a}_j = \sum_{i=j+1}^{l-1} 2^i a_i$, 即存在一个正整数 b , 使得 $\bar{a}_j = 2^j b$ 。令一个 2^l 维向量 $\mathbf{pt} = (pt_0, pt_1, \dots, pt_{2^l-1})$ 作为状态, $\mathbf{pt}$ 在 j 时刻的状态表示为 $\mathbf{pt}^{(j)}$, 元素表示为 $pt_i^{(j)}$, $i = 0, 1, \dots, 2^l - 1$, 并且规定 $\mathbf{pt}^{(j)}$ 第 0 到 $2^{j+1} - 1$ 个连续的元素为有效。该向量较为特殊, 其中仅有 1 个元素为非零, 即 $pt_{k_j}^{(j)} = h_j x^{\bar{a}_j}$, 其余均为 0。这样就可以用 $st_j = (k_j, h_j, \bar{a}_j)$ 来刻画向量 $\mathbf{pt}$ 在 j 时刻的状态。令初始时刻为 $j = -1$; 状态表示为 $st_{-1} = (k_{-1}, h_{-1}, \bar{a}_{-1}) = (0, 1, a)$ 。

首先我们考察 Expand 中在 j 时刻迭代运行 1 次内循环, 将状态 st_{j-1} 更新到 st_j 。根据状态的定义, 我们知道在 $j-1$ 时刻, 仅有 1 个元素不为零, $pt_{k_{j-1}}^{(j-1)} = h_{j-1} x^{\bar{a}_{j-1}}$, 其余均为 0。那么在 j 时刻, 根据 Expand 算法的内循环表达, 则有

更新 $\mathbf{pt}$ 第 0 到 $2^j - 1$ 个位置的值:

$$pt_k^{(j)}(x) = pt_k^{(j-1)}(x) + pt_k^{(j-1)}(x^{N/2^j+1}) \bmod x^N + 1$$

定义为 $pt_k^{(j)} \leftarrow \text{LowTrans}(pt_k^{(j-1)}), k = 0, 1, \dots, 2^j - 1$ 。

激活 $\mathbf{pt}$ 第 2^j 到 $2^{j+1} - 1$ 个位置的值:

$$pt_{k+2^j}^{(j)} = pt_k^{(j-1)}(x) \cdot x^{-2^j} + pt_k^{(j-1)}(x^{N/2^j+1}) (x^{-2^j})^{N/2^j+1} \bmod x^N + 1$$

定义为 $pt_{k+2^j}^{(j)} \leftarrow \text{HighTrans}(pt_k^{(j-1)}), k = 0, 1, \dots, 2^j - 1$ 。

图 1 所示为扩展算法的迭代过程。

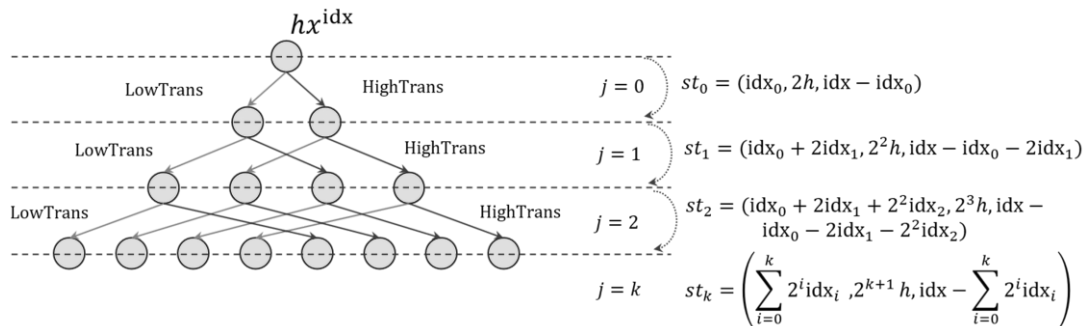


图 1 扩展算法的迭代过程

很容易看出对于输入为 0 的情况，状态更新函数 LowTrans、HighTrans 的输出均为 0；我们只需要考察输入不为 0 的情况，由状态 $st_{j-1} = (k_{j-1}, h_{j-1}, \bar{a}_{j-1})$ 可知， $pt_{k_{j-1}}^{(j-1)} = h_{j-1} x^{\bar{a}_{j-1}}$ ，我们只需要关注 $pt_{k_{j-1}}^{(j)}$ 、 $pt_{k_{j-1}+2^j}^{(j)}$ 的值。结合 $x^N = -1 \bmod x^N + 1$ ， $\bar{a}_{j-1} = \sum_{i=0}^{l-1} 2^i a_i = 2^j \sum_{i=0}^{l-j-1} 2^i a_{j+i}$ ，则有

$$pt_{k_{j-1}}^{(j)} = \text{LowTrans}(pt_{k_{j-1}}^{(j-1)}) = h_{j-1} \cdot x^{\bar{a}_{j-1}} + h_{j-1} \cdot (-1)^{a_j} x^{\bar{a}_{j-1}} \bmod x^N + 1$$

$$pt_{k_{j-1}+2^j}^{(j)} = \text{HighTrans}(pt_{k_{j-1}}^{(j-1)}) = h_{j-1} \cdot x^{\bar{a}_{j-1}-2^j} + h_{j-1} \cdot (-1)^{a_j-1} x^{\bar{a}_{j-1}-2^j} \bmod x^N + 1$$

当 $a_j = 0$ 时， $pt_{k_{j-1}}^{(j)} = 2h_{j-1} \cdot x^{\bar{a}_{j-1}}$ ， $pt_{k_{j-1}+2^j}^{(j)} = 0$ ；即 $k_j = k_{j-1}$ ， $h_j = 2h_{j-1}$ ， $\bar{a}_j = \bar{a}_{j-1}$ 。

当 $a_j = 1$ 时， $pt_{k_{j-1}}^{(j)} = 0$ ， $pt_{k_{j-1}+2^j}^{(j)} = 2h_{j-1} \cdot x^{\bar{a}_{j-1}-2^j}$ ；即 $k_j = k_{j-1} + 2^j$ ， $h_j = 2h_{j-1}$ ， $\bar{a}_j = \bar{a}_{j-1} - 2^j$ 。

因此，可以得到 $k_j = k_{j-1} + a_j 2^j$ ， $h_j = 2h_{j-1}$ ， $\bar{a}_j = \bar{a}_{j-1} - a_j 2^j$ ，且仅有 $pt_{k_j}^{(j)}$ 不为 0。根据初始值 $(k_{-1}, h_{-1}, \bar{a}_{-1}) = (0, 1, a)$ ，我们可以确定第 j 时刻的状态为

$$k_j = \sum_{i=0}^j a_i 2^i, h_j = 2^{j+1}, \bar{a}_j = a - \sum_{i=0}^j a_i 2^i$$

因此经过 l 次内循环迭代，我们得到 $l-1$ 时刻的状态为 $k_{l-1} = a$ ， $h_{l-1} = 2^l$ ， $\bar{a}_j = 0$ ；即 $pt_a^{(l-1)} = 2^{l-1}$ ， $pt^{(l-1)}$ 中其余的元素均为 0。然后计算 $(2^l)^{-1} pt^{(l-1)} \bmod n$ ，再取第 0 到 $n-1$ 个元素 $(pt_0^{(l-1)}, pt_1^{(l-1)}, \dots, pt_{n-1}^{(l-1)})$ ，即满足 $pt_a^{(l-1)} = 1$ ，其余元素为 0。

我们将以上整个明文过程投射到密文计算过程,就是完整 Expand 算法过程。只不过初始状态变化为密文状态(0, 1, HE.Enc(x^{idx}))。Expand 算法输出的密文向量 $\mathbf{c} = (c_0, c_1, \dots, c_{n-1})$, 对于所有的 $0 \leq k \leq n-1$, 满足以下条件。

$$c_k = \begin{cases} \text{HE.Enc}(1), & \text{如果 } k = \text{idx} \\ \text{HE.Enc}(0), & \text{如果 } k \neq \text{idx} \end{cases}$$

证毕。

可以看出 Expand 算法涉及同态加法、同态替换、单项式乘法等操作,这些同态操作的计算耗时都不多。经过分析可知,Expand 算法一共包含 $2^{t+1} - 2$ 次同态加法、同态替换、单项式乘法,都属于 $O(n)$ 级别。

为了进一步降低 Expand 算法的复杂度,SealPIR 采用了 Stern 等人所提出的数据库矩阵化编排方法,将 n 个元素的数据库编排成一个 $\sqrt{n} \times \sqrt{n}$ 矩阵 $\mathbf{M}$ 。

$$\mathbf{M} = \begin{bmatrix} DB_{0,0} & DB_{0,1} & \cdots & DB_{0,\sqrt{n}-1} \\ DB_{1,0} & DB_{1,1} & \cdots & DB_{1,\sqrt{n}-1} \\ \vdots & \vdots & & \vdots \\ DB_{\sqrt{n}-1,0} & DB_{\sqrt{n}-1,1} & \cdots & DB_{\sqrt{n}-1,\sqrt{n}-1} \end{bmatrix}$$

我们可以通过行和列的坐标来定位元素的位置。客户端需要发送行、列共两个询问来查询元素。假设期望查询的元素处于矩阵 $\mathbf{M}$ 中第 r 行、第 c 列,其中 $0 \leq r, c \leq \sqrt{n}-1$, 那么客户端需要向数据库服务器发送两个询问 $\text{query}_{\text{row}} = \text{HE.Enc}(x^r)$ 、 $\text{query}_{\text{col}} = \text{HE.Enc}(x^c)$ 。服务器采用 Expand 算法将询问扩展为 2 个密文向量 $\mathbf{c}_{\text{row}} = (c_{\text{row},0}, c_{\text{row},1}, \dots, c_{\text{row},\sqrt{n}-1})$ 、 $\mathbf{c}_{\text{col}} = (c_{\text{col},0}, c_{\text{col},1}, \dots, c_{\text{col},\sqrt{n}-1})$ 。此时,询问的尺寸大小为 $4n \log_2 q$ 。Expand 算法开销降低为 $O(\sqrt{n})$ 次同态加法、同态替换、单项式乘法。

首先我们处理矩阵的行,得到一个密文行向量 $\mathbf{v} = \mathbf{c}_{\text{row}}^T \mathbf{M} = (v_0, v_1, \dots, v_{\sqrt{n}-1})^T$, 其中 $v_k = \sum_{i=0}^{\sqrt{n}-1} c_{\text{row},i} DB_{i,k}, k = 0, 1, \dots, \sqrt{n}-1$ 。然后处理列 $\mathbf{vc}_{\text{col}} = \sum_{k=0}^{\sqrt{n}-1} v_k c_{\text{col},k}$ 。

两个同态密文的乘法会导致密文噪声增长过大。为了保证整个协议的正确性,需要将同态密码方案的参数设置较大,这带来了计算效率降低。为了避免此情况发生,SealPIR 采用了基于密文分块的同态密文乘法 $\mathbf{c}_{\text{cMult}} \leftarrow \text{chunksHEMult}(c_0, c_1, F)$ 。该方法需要依次执行密文分块、明密文乘法两个过程。为了解密 $\mathbf{c}_{\text{cMult}}$ 得到 $m_0 m_1$, 需要对其按顺序执行解密、重构、解密三个过程(图 2),我们称为基于密文分块的同态密文解密算法: $m \leftarrow \text{chunksDec}(sk, \mathbf{c}_{\text{cMult}})$ 。

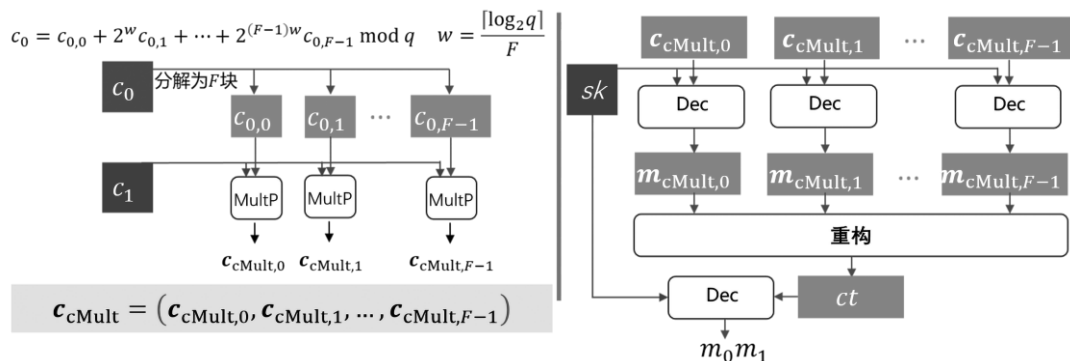


图2 基于密文分块的同态密文乘法与解密

(注意, 图中的每个 $c_{cMult,i}, i = 0, 1, \dots, F-1$ 含有两个同态密文)

首先描述 chunksHEMult 算法:

基于密文分块的同态密文乘法 $c_{cMult} \leftarrow \text{chunksHEMult}(c_0, c_1, F)$

输入: 两个同态加密的密文 c_0, c_1 , 分块尺寸 F 。

输出: $2F$ 个密文 $c_{cMult} = (c_{cMult,0}, c_{cMult,1}, \dots, c_{cMult,2F-1})$ 。

- (1) 密文分块: 令 $w = \lceil \log_2 q \rceil / F$, 将密文 c_0 做 2^w 进制分解为 F 个块 $c_{0,0}, c_{0,1}, \dots, c_{0,F-1}$ 。满足 $c_0 = c_{0,0} + 2^w c_{0,1} + \dots + 2^{(F-1)w} c_{0,F-1} \bmod q$ 。注意: 根据 BFV 密文形式, 可知每小块含有两个环元素 $c_{0,i} = (c_{0,i,0}, c_{0,i,1})$, $i = 0, 1, \dots, F-1$ 。
- (2) 明密文乘法: 将 $c_{0,i} = (c_{0,i,0}, c_{0,i,1})$, $i = 0, 1, \dots, F-1$ 当作明文与密文 c_1 相乘, 得到一个密文向量 $c_{cMult} = (c_{cMult,0}, c_{cMult,1}, \dots, c_{cMult,2F-1})$, 其中 $c_{cMult,2i} = c_{0,i,0} c_1$, $c_{cMult,2i+1} = c_{0,i,1} c_1$, $i = 0, 1, \dots, F-1$ 。

chunksHEMult 算法将系数较大的密文切成系数较小的密文块, 并将密文块看作明文, 从而执行明密文乘法操作。那么为了解密密文 c_{cMult} , 需要两次解密, 下面描述 chunksDec 算法:

基于密文分块的同态密文解密算法 $m_{cMult} \leftarrow \text{chunksDec}(sk, c_{cMult})$

输入: chunksHEMult 算法生成的一组密文 $c_{cMult} = (c_{cMult,0}, c_{cMult,1}, \dots, c_{cMult,2F-1})$ 。

输出: 解密得到消息 $m \in R_q$ 。

- (1) 分块解密: 对 c_{cMult} 的每个密文进行解密 $m_{cMult,i} = \text{HE.Dec}(sk, c_{cMult,i})$, $i = 0, 1, \dots, 2F-1$ 。
- (2) 重构: 计算 $ct_0 = \sum_{i=0}^{F-1} 2^{iw} m_{cMult,2i} \bmod q$, $ct_1 = \sum_{i=0}^{F-1} 2^{iw} m_{cMult,2i+1} \bmod q$ 。然后形成新的密文 $ct = (ct_0, ct_1)$ 。
- (3) 解密: $m_{cMult} = \text{HE.Dec}(sk, ct)$ 。

有了数据库矩阵编码、扩展算法、基于密文分块的同态密文乘法，就可以得到完整的 SealPIR 协议，具体如下：

输入：

- (1) 规模为 n 的数据库 DB，设数据库中的每个记录为 DB_i ， $i=0,1,\dots,n-1$ 。
- (2) 索引 $\text{idx} \in 0,1,\dots,n-1$ ，即客户端期望查询数据库中第 idx 个记录。

输出：客户端获得 DB_{idx} 。

- (1) 服务器将数据库 DB 编排成一个 $\sqrt{n} \times \sqrt{n}$ 矩阵 M 。
- (2) 客户端将索引 idx 转换为 $(\text{idx}_0, \text{idx}_1)$ ，其中 $\text{idx}_0 = \lceil \text{idx} / \sqrt{n} \rceil - 1$ ， $\text{idx}_1 = \text{idx} - \text{idx}_0 \sqrt{n}$ 。
- (3) 客户端生成加密询问 $\text{Query} = (\text{query}_{\text{row}}, \text{query}_{\text{col}})$ ，并发送给服务器，其中， $\text{query}_{\text{row}} = \text{HE.Enc}(pk, x^{\text{idx}_0})$ ， $\text{query}_{\text{col}} = \text{HE.Enc}(pk, x^{\text{idx}_1})$ 。
- (4) 服务器接收到来自客户端的加密询问 Query ，运行扩展算法：

$$c_{\text{row}} = (c_{\text{row},0}, c_{\text{row},1}, \dots, c_{\text{row},\sqrt{n}}) \leftarrow \text{Expand}(\text{query}_{\text{row}}, \sqrt{n}) ;$$

$$c_{\text{col}} = (c_{\text{col},0}, c_{\text{col},1}, \dots, c_{\text{col},\sqrt{n}}) \leftarrow \text{Expand}(\text{query}_{\text{col}}, \sqrt{n}) .$$

- (5) 服务器采用明密文同态乘法、同态加法计算 $v = c_{\text{row}}^T M$ ，即 $v_k = \sum_{i=0}^{\sqrt{n}-1} c_{\text{row},i} DB_{i,k}$ ， $k = 0,1,\dots,\sqrt{n}-1$ 。

- (6) 服务器采用基于密文分块的同态密文乘法，计算 $c_{\text{cMult},k} \leftarrow \text{chunksHEMult}(v_k, c_{\text{col},k}, F)$ ， $k = 0,1,\dots,\sqrt{n}-1$ 。然后

采用同态加法进行求和 $c_{\text{cMult}} = \sum_{k=0}^{\sqrt{n}-1} c_{\text{cMult},k}$ 。注意， c_{cMult} 含有 $2F$ 个密文 $(c_{\text{cMult},0}, c_{\text{cMult},1}, \dots, c_{\text{cMult},2F-1})$ 。

- (7) 服务器将 c_{cMult} 发送给客户端。

- (8) 客户端采用基于密文分块的同态密文解密算法，计算 $DB_{\text{idx}} \leftarrow \text{chunksDec}(sk, c_{\text{cMult}})$ 得到查询结果。

3.5 OnionPIR 隐私信息检索方案

SealPIR 方案的询问尺寸较小，但其响应尺寸较大，计算开销也较大。SealPIR 方案的响应尺寸相比数据库原始记录增大了 100 倍，如客户端为了查询获取 30KB 的数据库记录，必须下载约 3MB 的数据；从 100 万条记录规模的数据中获取 30KB 的数据记录，服务器的计算耗时约为 400s。

为了解决 SealPIR 存在的缺点，2021 年 Mughees 等人^[30]在 SealPIR 的基础上设计了 OnionPIR 方案。该方案采用了 BFV、RGSW 两种同态加密方案，利用 BFV 密文与 RGSW 密

文外积，来降低密文乘法的噪声增长，从而将响应尺寸的扩展率由 100 降低到 4.2，计算开销与 SealPIR 相当，但询问尺寸较大。

我们用 BFV.Keygen 、 BFV.Enc 、 BFV.Dec 、 BFV.Add 、 BFV.Mul 分别表示 BFV 同态加密方案的密钥生成、加密、解密、同态加法、同态乘法。用 RGSW.Keygen 、 RGSW.Enc 、 RGSW.Dec 、 RGSW.Add 、 RGSW.Mul 分别表示 RGSW 同态加密方案的密钥生成、加密、解密、同态加法、同态乘法。

1. RGSW 同态加密方案

给定一个基 B ，令 $l = \lceil \log_B q \rceil$ 。首先引入工具向量 $\mathbf{g}$ 和工具矩阵 $\mathbf{G}$ ：

$$\mathbf{g} = (1, B, B^2, \dots, B^{l-1})$$

$$\mathbf{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \mathbf{g} = \begin{bmatrix} \mathbf{g} & \mathbf{0} \\ \mathbf{0} & \mathbf{g} \end{bmatrix} \in R_q^{(2l \times 2)}$$

对于 $b \in R_q$ ，将其进行以 B 为底的分解， $b = \sum_{i=0}^{l-1} B^i \cdot b_i$ ，其中 $\|b_i\|_\infty \leq \frac{B}{2}$ 。可以定义 $\mathbf{g}^{-1}(b) = (b_0, b_1, \dots, b_{l-1})$ 。

那么对于 $\mathbf{v} = (a, b) \in R_q^2$ ，可以定义 $\mathbf{G}^{-1}(\mathbf{v}) = (\mathbf{g}^{-1}(a), \mathbf{g}^{-1}(b))$ 。可以看出 $\mathbf{g}^T \mathbf{g}^{-1}(a) = a$ ，

$\mathbf{G}^T \mathbf{G}^{-1}(\mathbf{v}) = \mathbf{v}$ 。

$s \leftarrow \text{RGSW.Keygen}(1^\lambda)$ ：输入安全参数 λ ，生成满足安全强度的参数，在密钥分布上随机抽样，得到密钥 s ，令解密密钥为 $\mathbf{sk} = (1, s)$ 。

$\mathbf{C} \leftarrow \text{RGSW.Enc}(s, m)$ ：输入消息 $m \in R_l$ 和密钥 s ，在 R_q 上随机均匀选取 $2l$ 维向量 $\mathbf{a} = (a_0, a_1, \dots, a_{2l-1})$ ，在离散高斯分布 χ_σ 上选取 $2l$ 维错误向量 $\mathbf{e} = (e_0, e_1, \dots, e_{2l-1})$ ，计算 m 的加密，得到 R_q 上 $2l \times 2$ 维矩阵，即密文：

$$\mathbf{C} = \begin{bmatrix} -a_0 s + e_0 & a_0 \\ -a_1 s + e_1 & a_1 \\ \vdots & \vdots \\ -a_{2l-1} s + e_{2l-1} & a_{2l-1} \end{bmatrix} + \Delta m \begin{bmatrix} \mathbf{g} & \mathbf{0} \\ \mathbf{0} & \mathbf{g} \end{bmatrix} \triangleq \mathbf{Z} + \Delta m \mathbf{G}$$

可以看出 $\mathbf{Z} \cdot \mathbf{sk} = \mathbf{e}$ ， $\mathbf{C} \cdot \mathbf{sk} = \mathbf{e} + \Delta m \mathbf{G} \cdot \mathbf{sk}$ 。

2. BFV 密文与 RGSW 密文的外积

输入一个 BFV 密文 $\mathbf{ct} = (ct_0, ct_1) \in R_q^2$ ，是消息 m_0 的加密；一个 RGSW 密文 $\mathbf{C} \in R_q^{2l \times 2}$ ，是消息 m_1 的加密。这两个密文对应的解密密钥都是 s 。则我们可以定义密文外积：

$$\mathbf{C} \boxtimes \mathbf{ct} = \lfloor \frac{t}{q} \mathbf{C}^T \mathbf{G}^{-1}(\mathbf{ct}) \rfloor \in R_q^2$$

$$\begin{aligned} \text{粗略地分析: } \text{BFV.Dec}(s, ct \boxdot C) &= \left\lfloor \left(\frac{t}{q} \right)^2 (ct \boxdot C)^T sk \right\rfloor = \left\lfloor \left(\frac{t}{q} \right)^2 (G^{-1}(ct))^T C \cdot sk \right\rfloor = \\ &= \left\lfloor \left(\frac{t}{q} \right)^2 (G^{-1}(ct))^T Z \cdot sk + m_1 \frac{t}{q} ct^T sk \right\rfloor \bmod t. \end{aligned}$$

由于 ct 是消息 m_0 的加密, 因此 $\frac{t}{q} ct^T sk = m_0 + e' \bmod t$ 。

$$\text{则有 } \text{BFV.Dec}(s, ct \boxdot C) = \left\lfloor m_0 m_1 + m_1 e' + \left(\frac{t}{q} \right)^2 (G^{-1}(ct))^T e \right\rfloor \bmod t$$

当 $\left\| m_1 e' + \left(\frac{t}{q} \right)^2 (G^{-1}(ct))^T e \right\| < \frac{1}{2}$ 时, 则有 $\text{BFV.Dec}(s, ct \boxdot C) = m_0 m_1$ 。可以看出外积得到的

密文噪声并不是两个密文的乘积, 是一个线性关系。

利用外积, 我们将第 1 维的查询采用 BFV 密文, 第 2 到 d 维的查询采用 RGSW 密文 (图 3), 这样可以将查询表示为一系列同态密文外积计算, 降低噪声增长率, 使得参数设置更小。

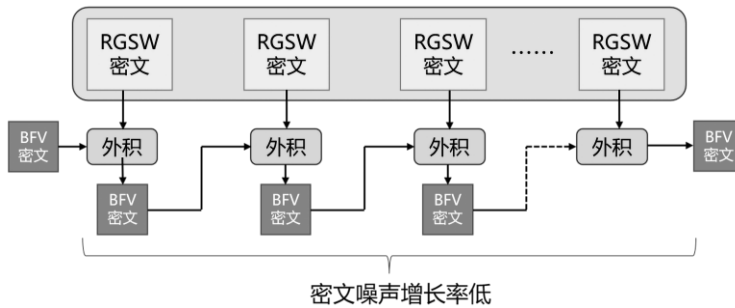


图 3 OnionPIR 方案中基于外积的查询计算

3. 分块明密文乘法

明文空间为 R_t , 明文多项式的每个系数最多可以表示成 $l_p = \lceil \log_2 t \rceil$ 比特, 则明文可分解成两个 $\lceil l_p / 2 \rceil$ 比特的多项式, 即存在 $pt_0, pt_1 \in R_{2^{\lceil l_p / 2 \rceil}}$, 满足 $pt = pt_0 + 2^{\lceil l_p / 2 \rceil} \cdot pt_1$, 我们将明文分块定义为 $\mathbf{pt} = (pt_0, pt_1) \leftarrow \text{DecompPlain}(pt)$ 。

分块加密算法 $ct' = (ct_0, ct_1) \leftarrow \text{DecompEncrypt}(b)$: 输入一个比特 $b \in \{0, 1\}$, 执行加密 $ct_0 = \text{BFV.Enc}(pt, b)$, $ct_1 = \text{BFV.Enc}(pk, 2^{\lceil l_p / 2 \rceil} b)$ 。

分块明密文乘法 $ct \leftarrow \text{DecompMult}(pt', ct')$: 输入明文分块 $pt' = (pt_0, pt_1)$, 以及分块加密算法产生的密文 $ct' = (ct_0, ct_1)$ 。计算 $ct = pt_0 ct_0 + pt_1 ct_1 \bmod q$ 。

很容易验证 $\text{BFV.Dec}(ct) = b \cdot pt$ 。

分块明密文乘法的密文噪声增长更小, 在最坏情况下, 将噪声由 $\frac{t}{2} \cdot e$ 降低为 $\frac{\sqrt{t}}{2} \cdot e$ 。

4. 询问压缩

OnionPIR 方案将数据库 DB 编排为 d 维空间的超立方体, 其中第 1 维的尺寸为 $N_1=128$, 后续第 2 到 d 维的尺寸都为 $N_2 = N_3 = \dots = N_d = 4$ 。用 d 个向量 $b_0 \in \{0,1\}^{128}, b_1 \in \{0,1\}^4, \dots, b_{d-1} \in \{0,1\}^4$ 表示超立方体每个维度空间的明文询问向量, 这些向量仅有一个的分量值为 1, 其余为 0。

OnionPIR 方案如果继续采用 SealPIR 中的询问生成方法, 将导致询问为 d 个密文。为了保持较小的询问尺寸, OnionPIR 方案提出了新的询问压缩算法 QueryPack, 将这 d 个向量压缩为 1 个明文, 然后对其采用 BFV 方案加密, 得到一个 BFV 密文作为询问。

SealPIR 方案采用基于密文分块的同态密文乘法, 来降低同态密文乘法噪声增长, 但是这种方法存在一定限制, 只能密态执行一次, 无法持续做同态乘法。因此 SealPIR 方案只能支持数据库矩阵编排。OnionPIR 方案引入同态外积, 可以处理维度更高的数据库编排。

OnionPIR 方案为了进一步降低计算过程中密文噪声增长, 对第 1 维度的明密文计算采用了分块明密文乘法。

令明文多项式为 $pt = pt_0 + pt_1 x + \dots + pt_{N-1} x^{N-1}$, 首先我们将 pt 初始化为 0。

第 1 维度: $pt_{2i} = b_{0,i}, \quad pt_{2i+1} = 2^{\lceil l_p/2 \rceil} b_{0,i}, \quad i = 0, 1, 2, \dots, 127$

第 2 维度: $pt_{256+i+l+j} = B^j b_{1,i}, \quad i = 0, 1, 2, 3, \quad j = 0, 1, \dots, l-1$

...

第 k 维度: $pt_{256+(k-2) \cdot 4l+i+l+j} = B^j b_{k-1,i}, \quad i = 0, 1, 2, 3, \quad j = 0, 1, \dots, l-1$

...

第 d 维度: $pt_{256+(d-2) \cdot 4l+i+l+j} = B^j b_{d-1,i}, \quad i = 0, 1, 2, 3, \quad j = 0, 1, \dots, l-1$

接下来, 使用 BFV 同态加密方案, 加密明文 pt , 得到 $\text{query} = \text{BFV.Enc}(pt)$ 。那么询问压缩算法描述如下。

询问压缩算法 $\text{query} \leftarrow \text{QueryPack}(b_0, b_1, \dots, b_{d-1})$

输入: d 个询问向量 $b_0 \in \{0,1\}^{N_1}, b_1 \in \{0,1\}^{N_2}, \dots, b_{d-1} \in \{0,1\}^{N_d}$ 。

输出：一个 BFV 密文 **query**。

- (1) 设置明文多项式 $pt = 0$ 。
- (2) $pt_{2i} = b_{0,i}$, $pt_{2i+1} = 2^{\lceil l/2 \rceil} b_{0,i}$, $i = 0, 1, 2, \dots, N_1 - 1$ 。
- (3) 设置 $ptr = 2N_1$ 。
- (4) 对于 $i = 1, 2, \dots, d-1$ 执行：
 - ① 对于 $j = 0, 1, \dots, N_i - 1$ 计算： $pt_{ptr+j} = B^j b_{i,j}$, $j = 0, 1, \dots, l-1$ 。
 - ② 计算 $ptr = ptr + l \cdot N_i$ 。
- (5) 计算加密 **query** = BFV.Enc(pt)。

5. 询问解压缩

由 OnionPIR 方案设计思想可知，我们需要将 **query** 扩展成 d 个密文向量，分别对应 $b_0, b_1, \dots, b_{d-1}$ 的加密。为了持续处理数据库超立方体的每一维，并保证同态乘法操作带来的噪声增长较小，OnionPIR 方案需要使得 $b_1, \dots, b_{d-1}$ 被 RGSW 方案加密。

$$\mathbf{C}_{\text{BFV}} = (\mathbf{c}_{\text{BFV},0}, \mathbf{c}_{\text{BFV},1}, \dots, \mathbf{c}_{\text{BFV},255})$$

$$\mathbf{C}_{\text{RGSW}}^1 = (\mathbf{C}_{\text{RGSW},0}^1, \mathbf{C}_{\text{RGSW},1}^1, \mathbf{C}_{\text{RGSW},2}^1, \mathbf{C}_{\text{RGSW},3}^1)$$

$$\mathbf{C}_{\text{RGSW}}^2 = (\mathbf{C}_{\text{RGSW},0}^2, \mathbf{C}_{\text{RGSW},1}^2, \mathbf{C}_{\text{RGSW},2}^2, \mathbf{C}_{\text{RGSW},3}^2)$$

...

$$\mathbf{C}_{\text{RGSW}}^{d-1} = (\mathbf{C}_{\text{RGSW},0}^{d-1}, \mathbf{C}_{\text{RGSW},1}^{d-1}, \mathbf{C}_{\text{RGSW},2}^{d-1}, \mathbf{C}_{\text{RGSW},3}^{d-1})$$

满足 $\mathbf{c}_{\text{BFV},i} = \text{BFV.Enc}(b_{0,i})$, $i = 0, 1, 2, \dots, 127$ 。

$$\mathbf{C}_{\text{RGSW},i}^j = \text{RGSW.Enc}(b_{j,i}), j = 1, 2, \dots, d-1, i = 0, 1, 2, 3。$$

借鉴 SealPIR 的思路，采用增强型扩展算法 AugmentedExpend^[30]来提取系数，获得 $\text{BFV.Enc}(pt_i)$, $i = 0, 1, \dots, 255 + (d-1) \cdot 4l$ 。

第 1 维度，密文向量为： $\{\text{BFV.Enc}(pt_i)\}_{i=0}^{255}$

第 2 维度，密文向量为： $\{\text{BFV.Enc}(pt_i)\}_{i=256}^{255+4l}$

...

第 k 维度，密文向量为： $\{\text{BFV.Enc}(pt_i)\}_{i=256+(k-2)4l}^{255+(k-1)4l}$

...

第 d 维度，密文向量为： $\{\text{BFV.Enc}(pt_i)\}_{i=256+(d-2)4l}^{255+(d-1)4l}$

明显, 这不符合我们的要求, 因此需要解决 $\{\text{BFV.Enc}(pt_i)\}_{i=256+(k-2)4l}^{255+(k-1)4l}$ 到 $(\mathbf{C}_{\text{RGSW},0}^{k-1}, \mathbf{C}_{\text{RGSW},1}^{k-1}, \mathbf{C}_{\text{RGSW},2}^{k-1}, \mathbf{C}_{\text{RGSW},3}^{k-1})$ 的转换, $k=2,3,\dots,d$, 来获得 $d-1$ 组 RGSW 密文。

$\{\text{BFV.Enc}(pt_i)\}_{i=256+(k-2)4l}^{255+(k-1)4l}$ 对应 $\{\text{BFV.Enc}(B^j b_{k-1,i})\}_{i=0,1,2,3,j=0,1,\dots,l-1}$ 。由此可以构建一个密文矩阵, 令 $\text{BFV.Enc}(B^j b_{k-1,i})$ 可以表示为

$$(-a_j s + e_j + \Delta B^j b_{k-1,i}, a_j) = (-a_j s + e_j, a_j) + \Delta b_{k-1,i} (B^j, 0)$$

由于

$$\begin{aligned} \text{BFV.Enc}(s B^j b_{k-1,i}) &= (-a'_j s + e'_j + \Delta B^j b_{k-1,i} s, a'_j) \\ &= (- (a'_j - \Delta B^j b_{k-1,i}) s + e'_j, a'_j - \Delta B^j b_{k-1,i} s) + (0, \Delta B^j b_{k-1,i} s) \end{aligned}$$

令 $a''_j = a'_j - \Delta B^j b_{k-1,i}$, 则有

$$\text{BFV.Enc}(s B^j b_{k-1,i}) = (-a''_j s + e'_j, a''_j) + \Delta b_{k-1,i} (0, B^j)$$

由外积的性质可知, 用 $\text{RGSW.Enc}(s) \square \text{BFV.Enc}(B^j b_{k-1,i})$ 来计算得到 $\text{BFV.Enc}(s B^j b_{k-1,i})$ 。

因此可以构建一个密文向量

$$\begin{aligned} \begin{bmatrix} \text{BFV.Enc}(B^0 b_{k-1,i}) \\ \text{BFV.Enc}(B^1 b_{k-1,i}) \\ \vdots \\ \text{BFV.Enc}(B^{l-1} b_{k-1,i}) \\ \text{RGSW}(s) \square \text{BFV.Enc}(B^0 b_{k-1,i}) \\ \text{RGSW}(s) \square \text{BFV.Enc}(B^1 b_{k-1,i}) \\ \vdots \\ \text{RGSW}(s) \square \text{BFV.Enc}(B^{l-1} b_{k-1,i}) \end{bmatrix} &= \begin{bmatrix} (-a_0 s + e_0, a_0) + \Delta b_{k-1,i} (B^0, 0) \\ (-a_1 s + e_1, a_1) + \Delta b_{k-1,i} (B^1, 0) \\ \vdots \\ (-a_{l-1} s + e_{l-1}, a_{l-1}) + \Delta b_{k-1,i} (B^{l-1}, 0) \\ (-a''_0 s + e'_0, a''_0) + \Delta b_{k-1,i} (0, B^0) \\ (-a''_1 s + e'_1, a''_1) + \Delta b_{k-1,i} (0, B^1) \\ \vdots \\ (-a''_{l-1} s + e'_{l-1}, a''_{l-1}) + \Delta b_{k-1,i} (0, B^{l-1}) \end{bmatrix} \\ &= \mathbf{Z}_{k-1,i} + \Delta b_{k-1,i} \mathbf{G} \end{aligned}$$

该密文向量形式上可以看作 $\text{RGSW}(b_{k-1,i})$, 充分运用此方法, 我们就可以将 $\{\text{BFV.Enc}(B^j b_{k-1,i})\}_{i=0,1,2,3,j=0,1,\dots,l-1}$ 转换为 $(\text{RGSW.Enc}(b_{k-1,0}), \text{RGSW.Enc}(b_{k-1,1}), \text{RGSW.Enc}(b_{k-1,2}), \text{RGSW.Enc}(b_{k-1,3}))$ 。

询问解压缩算法 $(\mathbf{C}_{\text{BFV}}, \{\mathbf{C}_{\text{RGSW}}^i\}_{i=1}^{d-1}) \leftarrow \text{QueryUnpack}(\text{query}, \mathbf{C}_s)$

输入：一个加密 **query**，该查询是询问压缩算法产生的密文，即 $\text{query} = \text{QueryPack}(\mathbf{pk}, \mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_{d-1})$ ；一个密文 $C_s = \text{RGSW}.\text{Enc}(s)$ 。

输出： $(C_{\text{BFV}}, \{C_{\text{RGSW}}^i\}_{i=1}^{d-1})$ 。

(1) 首先调用增强型扩展算法，得到 N 个密文： $\mathbf{ct}_{\text{BFV}} = \text{AugmentedExpend}(\text{query})$ ，其中 $\mathbf{ct}_{\text{BFV}} = (\mathbf{ct}_{\text{BFV},0}, \mathbf{ct}_{\text{BFV},1}, \dots, \mathbf{ct}_{\text{BFV},N-1})$ 。

(2) $\mathbf{c}_{\text{BFV},i} = \mathbf{ct}_{\text{BFV},i}$ ， $i = 0, \dots, 2N_1 - 1$ 。

(3) 设置 $\text{ptr} = 2N_1$ 。

(4) 对于 $i = 1, 2, \dots, d-1$ 执行：

① 对于 $j = 0, 1, \dots, N_i - 1$ 计算：

$$C_{\text{RGSW},j,k}^i = \mathbf{ct}_{\text{BFV},\text{ptr}+j+k}, \quad k = 0, 1, \dots, l-1;$$

$$C_{\text{RGSW},j,k+l}^i = C_s \boxtimes \mathbf{ct}_{\text{BFV},\text{ptr}+j+k}, \quad k = 0, 1, \dots, l-1。$$

② 计算 $\text{ptr} = \text{ptr} + l \cdot N_i$ 。

输出：

$$C_{\text{BFV}} = (\mathbf{c}_{\text{BFV},0}, \mathbf{c}_{\text{BFV},1}, \dots, \mathbf{c}_{\text{BFV},2N_1-1});$$

$$C_{\text{RGSW}}^i = (C_{\text{RGSW},0}^i, C_{\text{RGSW},1}^i, \dots, C_{\text{RGSW},N_i-1}^i), \quad i = 1, 2, \dots, d-1。$$

其中 $C_{\text{RGSW},j}^i = (C_{\text{RGSW},j,0}^i, C_{\text{RGSW},j,1}^i, \dots, C_{\text{RGSW},j,2l-1}^i)$ 。

6. OnionPIR 协议

有了以上组件，我们就可以组装形成完整的 OnionPIR 协议。

输入：

(1) 规模为 n 的数据库 DB，设数据库中的每个记录为 DB_i ， $i = 0, 1, \dots, n-1$ 。

(2) 索引 $\text{idx} \in \{0, 1, \dots, n-1\}$ ，即客户端期望查询数据库中第 idx 个记录。

输出：客户端获得 DB_{idx} 。

(1) 服务器首先将数据库 DB 的每个元素做明文分块：

$$\mathbf{pt}_i \leftarrow \text{DecompPlain}(\text{DB}_i), i = 0, 1, \dots, n-1$$

(2) 客户端将索引 idx 转换为 d 个询问向量 $\mathbf{b}_0 \in \{0, 1\}^{N_1}, \mathbf{b}_1 \in \{0, 1\}^{N_2}, \dots, \mathbf{b}_{d-1} \in \{0, 1\}^{N_d}$ 。

(3) 客户端运行询问压缩算法生成询问 $\text{query} \leftarrow \text{QueryPack}(\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_{d-1})$ ，并且计算 $C_s = \text{RGSW}.\text{Enc}(s)$ 。然后将 query 和 C_s 发送给数据库服务器。

(4) 服务器接收到来自客户端的加密询问 **query** 和 C_s ，运行询问解压算法：

$$(C_{BFV}, \{C_{RGSW}^i\}_{i=1}^{d-1}) \leftarrow \text{QueryUnpack}(\text{query}, C_s)$$

(5) 服务器处理数据库第 1 维：对于 $j = 0, 1, \dots, n / N_1 - 1$ ，

$$\overline{DB}_j = \sum_{k=1}^{N_1} \text{DecompMul}((c_{BFV,2k}, c_{BFV,2k+1}) pt_{k+j \cdot N_1})$$

(6) 服务器处理数据库第 2 到 d 维：初始化 $len = n / N_1$ ，对于 $i = 2, 3, \dots, d$ ，循环迭代运行：

① 设置 $len = len / N_i$ ；

② 对于 $j = 0, 1, \dots, len - 1$ ，

$$\overline{DB}_j = \sum_{k=1}^{N_i} c_{RGSW,k}^{i-1} \boxtimes \overline{DB}_{k+j \cdot N_i}$$

③ 将 $\overline{DB}_0, \overline{DB}_1, \dots, \overline{DB}_{len-1}$ 作为计算结果，进入到下一轮循环迭代。

(7) 服务器迭代运行后，得到 $\overline{DB}_0$ ，并将其发送给客户端。

(8) 客户端解密 $DB_{idx} \leftarrow \text{BFV.Dec}(sk, \overline{DB}_0)$ ，得到查询结果。

3.6 未来发展方向

大部分高效率的计算安全隐私信息检索协议都是基于全同态加密来构造的，并且百万量级规模的数据库隐匿查询计算效率达到秒级。对于大型数据库仍然需要相当大的服务器计算开销。进一步增强 PIR 协议的大规模数据库处理能力，降低通信开销和计算开销，仍然是研究的重点。为此我们提出以下发展方向。

(1) 基于全同态加密的 PIR 协议高性能实现。在 OnionPIR 的实现中，80% 的服务器计算时间用于 NTT 计算。GPU 和 FPGA 可以大大加快 NTT 速度，进而加快基于 RLWE 的全同态加密的执行效率。因此将 GPU 和 FPGA 高性能实现整合到 PIR 协议中将是发展方向之一。

(2) 可更新的带状态 PIR 协议设计。目前带状态的 PIR 框架只适合静态数据库，但是实际应用场景中数据库往往处于不断更新中。因此可以提出可更新的带状态的 PIR 协议。

(3) 基于全同态加密的关键词 PIR 协议设计。在实际应用场景中，客户端一般不知道指标索引与记录内容的对应关系，而是习惯用关键词去查询。因此，设计基于全同态加密的关键词 PIR 协议也是一个符合实际需要的研究点。

(4) 恶意模型下的基于全同态加密的 PIR 协议设计。用户向数据库发起隐私查询时, 数据库服务器可以不遵守协议执行, 导致返回错误的查询结果, 而且用户无法判断该结果是否为期望结果。因此如何抵御数据库服务器的恶意行为, 设计恶意模型下的基于全同态加密的 PIR 协议是一个有趣的方向。

4 全同态加密的隐私集合求交应用

4.1 隐私集合求交协议概述

隐私集合求交 (Private Set Intersection, PSI) 允许一组互不信任参与方联合计算参与方私有集合的交集, 但不泄露交集之外的任何额外信息。在协议完成后, 某些参与方或全部参与方将得到正确的交集。隐私集合求交协议是一种重要的密码学原语, 对数据隐私保护具有重要的应用价值, 可以应用于联系人发现、行为匹配、远程诊断、新冠接触者追踪、在线广告实际效果计算、基因序列匹配检测等场景。

根据参与方的数量、规模, 可以将其分为两方隐私集合求交、多方隐私集合求交。根据参与方腐化情况, 可分为半诚实安全、恶意模型安全。根据参与方集合规模的对等与否, 可以将两方隐私集合求交协议分为平衡 PSI、非平衡 PSI。Pinkas 等人^[33]根据技术路线的不同将隐私集合求交协议分为基于公钥的 PSI、基于电路的 PSI、基于 OT 的 PSI、基于第三方的 PSI。基于全同态加密构造的非平衡 PSI 协议, 效率非常高, 已经达到实用化水平。

4.2 国内外发展现状

1986 年, Meadows 等人^[34]提出了第一个安全的隐私集合求交协议。之后。1999 年, Huberman 等人^[35]对该协议进行了详细描述, 这类协议以 Diffie-Hellman 密钥交换为基础, 具有通信开销低的优势, 但由于需要执行较多的模幂运算, 计算开销较大。2004 年, Freedman 等人^[36]采用部分同态加密算法, 提出了一个基于不经意多项式计算的隐私集合求交协议。目前大多数高效隐私集合求交协议一般采用 Diffie-Hellman 交换协议、RSA 公钥加密、不经意传输来构造, 但这些方案的通信复杂度与参与方中最大集合的规模呈线性关系。

2017 年, Chen 等人^[37]设计了第一个基于全同态加密的安全实用的两方 PSI 协议, 通信复杂度与参与方中较小集合的规模呈线性关系。该方案将发送方的所有元素作为多项式根来构造多项式, 并把交集计算表示为“判断某个元素是否为该多项式的根”。同态执行该交集计算以计算出交集。为了避免同态乘法深度过大带来低效性, Chen 等人^[37]在协议中融合 SMID

同态计算、置换哈希、布谷鸟哈希、分窗机制等多个优化技术,得到效率较高的方案,在两个集合规模分别为 2^{24} 、11041 上执行隐私集合求交协议,通信开销为 37.4MB、耗时为 18.9s。Chen 等人^[37]将基于全同态加密方案的 PSI 协议与基于 OT 的 PSI 协议进行了性能对比,如表 4 所示。可见在非平衡场景下,基于全同态加密方案的 PSI 协议在通信效率、计算效率方面均优于基于 OT 的协议方案。

表 4 基于全同态加密的安全实用的两方 PSI 协议与其他 PSI 协议的性能比较

参与方集合规模		协议方案	通信量	总耗时			
N_X	N_Y			10Gbps 网络	100Mbps 网络	10Mbps 网络	1Mbps 网络
2^{24}	11041	文献[37]	23.2MB	115.4s	117.8s	134.4s	290.8s
		PSZ16 方案 ^[38]	480.9MB	40.5s	88.0s	448.5s	4084.8s
		KKRT16 方案 ^[39]	975.0MB	70.8s	188.7s	1269.1s	12156.7s
	5535	文献[37]	20.1MB	105.2s	107.2s	120.3s	211.1s
		PSZ16 方案 ^[38]	480.4MB	40.1s	87.9s	449.2s	4080.6s
		KKRT16 方案 ^[39]	962.1MB	70.4s	188.3s	1263.5s	12153.2s
2^{20}	11041	文献[37]	11.5MB	12.8s	14.0s	22.2s	105.4s
		PSZ16 方案 ^[38]	30.9MB	3.3s	7.0s	29.8s	263.7s
		KKRT16 方案 ^[39]	58.5MB	4.5s	11.6s	79.4s	688.1s
	5535	文献[37]	5.6MB	8.6s	9.2s	13.3s	53.6s
		PSZ16 方案 ^[38]	30.4MB	3.1s	6.8s	29.0s	260.0s
		KKRT16 方案 ^[39]	57.3MB	4.4s	11.5s	79.3s	686.0s

注:配置环境:2 台服务器,配置为 18 核 Intel Xeon CPU E5-2699v3 @2.3GHz, 256GB 内存,单线程下运行。

2018 年,Chen 等人^[40]在文献[37]的基础上优化设计了具有恶意安全的基于全同态加密的标签 PSI 协议,采用了 OPRF 预处理来改善性能和提升安全性,引入了更普适的 SIMD 编码算法,将集合元素比特长度由 32 比特提升到可支持任意比特长度,将协议拓展为带标签的 PSI 协议,具有抵御恶意发送者行为,安全性更高。

2021 年,Cong 等人^[41]基于同态加密设计了具有较小计算通信开销的带标签 PSI 协议。该协议利用了 Paterson-Stockmeyer 算法来改进计算复杂度,并实现了计算与通信开销的有效平衡。相比 Chen 等人的方案,在两个集合规模分别为 2^{28} 、 20^{48} 的情况下,所设计的隐私集合求交协议在线计算开销降低了 71%,通信开销降低了 63%。

针对多个参与方的情况,2021 年 Badrinarayanan 等人^[42]在 PKC2021 会议上基于门限全同态加密方案设计多方门限隐私集合求交协议,具有亚线性通信复杂度。

4.3 基础协议

全同态加密能够在不解密的情况下直接对加密数据执行算术电路、布尔电路计算,是一种计算功能强大的密码原语。隐私集合求交协议本质上在两个或多个集合上计算出交集,因此交集计算可以表示成一个电路。将集合同态加密为加密集合,并采用全同态加密在相应加密集合上同态执行交集计算电路,理论上就可以得到加密的交集计算结果,私钥持有者解密获得交集。按照此思路,我们可以构造一个安全的隐私集合求交协议。

以两方 PSI 为例:持有集合规模较小的参与者 R ,使用同态加密来加密集合,然后发给另一个参与者 S 。最后 S 将接收到的加密集合、自己未加密的集合作为输入,同态计算交集计算电路,便可以得到加密的交集结果,并发送给参与者 R 。参与者 R 解密出交集结果。整个过程的通信量仅为: $2 \times \text{密文扩展} \times \text{规模较小集合的数据量}$ 。可以看出这种方法在非平衡 PSI 上具有明显的通信效率优势。下面描述基于全同态密码的隐私集合求交的基础协议。

令两个参与方分别为发送方 Sender、接收方 Receiver,持有的私有集合分别为 X 、 Y ,并且其集合元素的比特长度为 σ ,接收方的集合大小为 N_X ,发送方的集合大小为 N_Y 。

该基础协议的功能:令来自接收方的输入为(RECEIVE, Y),令来自发送方的输入为(SEND, X),其中 $X, Y \subseteq \{0,1\}^\sigma$, $|X|=N_X$, $|Y|=N_Y$,并将(OUTPUT, $X \cap Y$)作为输出发送给接收方,不发送任何信息给发送方。

输入:发送方 Sender 持有私有集合 X ,集合大小为 N_X ;接收方 Receiver 持有私有集合 Y ,集合大小为 N_Y 。这两个集合中元素的比特长度都为 σ 。

输出:接收方 Receiver 输出 $X \cap Y$;发送方输出 $\perp$ 。

(1) 系统建立:发送方 Sender 和接收方 Receiver 联合协商使用相同的全同态加密方案,其方案参数也相同。接收方 Receiver 生成一个公私钥对(pk, sk),并仅自己持有该私钥 sk 。

(2) 集合加密:接收方采用全同态加密方案,加密集合 Y 中的每个元素 y_i ,即 $c_i = \text{FHE.Enc}(pk, y_i)$, $i = 1, 2, \dots, N_Y$ 。

然后将 N_Y 个密文 $(c_1, c_2, \dots, c_{N_Y})$ 发送给发送方。

(3) 交集计算：对于每个密文 c_i ，发送方执行以下计算。

① 随机选取一个非零明文元素 r_i 。

② 同态计算

$$d_i = r_i \prod_{x \in X} (c_i - x)$$

发送方 Sender 将一组密文 $(d_1, d_2, \dots, d_{N_Y})$ 发送给接收方 Receiver。

(4) 结果提取：接收方 Receiver 解密所收到的密文 $(d_1, d_2, \dots, d_{N_Y})$ ，并输出 $X \cap Y = \{y_i : \text{FHE.Decrypt}(d_i) = 0\}$ 。

由于全同态加密方案是 IND-CPA 安全，发送方 Sender 获得来自接收方 Receiver 的 N_Y 个密文与随机数据是不可区分的，因此发送方 Sender 无法获得关于集合 X 中元素的信息。

接收方 Receiver 对 N_Y 个密文解密，也只能获得交集 $X \cap Y$ 的元素，得不到除交集之外元素的任何信息。具体原因为：如果接收方 Receiver 的某个元素 $y_{i_0} \in X$ ，则在集合 X 中存在某个元素 $x_{j_0} = y_{i_0}$ ，那么接收方 Receiver 将获得密文 d_{i_0} 解密， $\text{FHE.Dec}(sk, d_{i_0}) = r_{i_0} \prod_{x \in X} (y_{i_0} - x) = 0$ ；如果某个元素 $y_{i_1} \notin X$ ，同理，有 $\text{FHE.Dec}(sk, d_{i_1}) = r_{i_1} \prod_{x \in X} (y_{i_1} - x)$ ，由于 r_{i_1} 是随机的，因此 $\text{FHE.Dec}(sk, d_{i_1})$ 是 $\mathbb{Z}_t$ 上的随机非零元素。对于 Y 集合中非交集元素，接收方 Receiver 解密出的结果是一个随机数据。那么可以通过解密结果是否为 0，来定位集合中哪些元素为交集的元素。

这个协议的通信开销为 $2N_Y$ 个密文，但是其计算开销较大，需要 $O(N_X N_Y)$ 次同态乘法和加法，同态乘法深度为 $O(\log_2 N_X)$ 。在全同态加密方案中同态乘法数量多、同态乘法深度深都是导致效率不高的主要因素。

该协议框架是当前基于全同态加密设计隐私集合求交协议的主要思路。不同学者在此基础上融入一些优化方法，以降低通信和计算开销。

4.4 基于全同态加密的隐私集合求交协议

2017 年，Chen 等人^[37]在基础协议上融合 SMID 同态计算、置换哈希、布谷鸟哈希、分窗机制等多个优化技术，设计出安全实用的 PSI 协议方案。下面给出整个方案的设计过程。

1. 集合元素尺寸压缩

在实际场景中，发送方和接收方的集合元素比特长度较长，不利于协议执行和同态加密的使用。为此，需要采用一个随机哈希函数 $H': \{0,1\}^\sigma \rightarrow \{0,1\}^{\sigma'}$ ，其中 $\sigma' < \sigma$ ，将集合元素进行压缩， $X' = \{H'(x) | x \in X\}$ ， $Y' = \{H'(y) | y \in Y\}$ 。那么就可以在比特长度更小的集合 X' 、 Y' 上

执行隐私集合求交。

2. 哈希表生成

基础协议框架的本质是通过判断集合 Y 中每个元素 y_i 是否为多项式 $r_i \prod_{x \in X} (y - x)$ 的根, 来判定 y_i 是否属于集合 X 。对于非平衡场景, 发送方持有的集合 X 规模较大, 计算 $r_i \prod_{x \in X} (y - x)$ 的开销较大。为了解决这一问题, 文献[37]采用哈希表, 将该问题转化为小规模的多项式计算问题, 以提升效率。

布谷鸟哈希: 令一个 2^k 长的哈希表 T , 一组 h 个哈希函数 $H_1, H_2, \dots, H_h$ 。为了插入一个元素 x 到哈希表 T 中, 需要随机选择 $i \in \{1, 2, \dots, h\}$, 计算 $H_i(x)$ 作为地址。然后查看哈希表中第 $H_i(x)$ 个位置是否已经被占用: 若没有被占用, 则将 (x, i) 存放在该位置, 即 $T[H_i(x)] = (x, i)$; 若该位置已经被 (y, j) 占用, 则将 (x, i) 存放在该位置, 并重新选择 $j' \in \{1, 2, \dots, h\} \setminus \{j\}$, 计算 $H_{j'}(y)$, 并查看 $T[H_{j'}(y)]$ 是否被占用, 未被占用就在此位置放置 (y, j) , 被占用就挤出该位置的数据, 并为被挤出的数据重新选择位置, 如此迭代下去, 直到无占用情况发生。

基于置换的哈希函数: 假设 $m = 2^k$, k 为某个正整数。为了将长度为 σ 的比特串 x 插入到哈希表中, 首先将 x 分成两个比特串的拼接, 即 $x = x_L \parallel x_R$, 其中 x_R 的比特长度为 $\log_2 m$, x_L 的比特长度为 $\sigma - \log_2 m$ 。我们基于一组哈希函数 $H_1, H_2, \dots, H_h: \{0, 1\}^{\sigma - \log_2 m} \rightarrow \{0, 1\}^{\log_2 m}$ 来构造基于置换的哈希函数 (称为定位函数):

$$\text{Loc}_i(x) = H_i(x_L) \oplus x_R, \quad i = 1, 2, \dots, h$$

然后在哈希表中地址为 $\text{Loc}_i(x)$ 的空间中存放 (x_L, i) 。可以看出, 在已知这 h 个函数情况下, 我们可以通过 $\text{Loc}_i(x)$ 、 x_L 、 i 恢复出 x , 即 $x = x_L \parallel (\text{Loc}_i(x) \oplus H_i(x_L))$ 。这节约了哈希表的存储空间。假设在两个哈希表中都采用了这种基于置换的哈希函数, 如果两个哈希表某个位置存放的内容相同 $(x_L, i) = (y_L, j)$, 那么必然推断出 $x = y$ 。

结合布谷鸟哈希和基于置换的哈希函数, 可以将问题“ y_i 是否属于集合 X ”转换成“判断某个与 y_i 相关的数据是否属于某个小集合”, 这较大地提升了方案效率。

下面给出哈希表生成算法 $(\mathfrak{B}_X, \mathfrak{B}_Y) \leftarrow \text{Hashing}(X, Y, H_1, H_2, \dots, H_h)$ 。

输入: 接收方 Receiver 输入元素个数为 N_Y 的集合 Y 。发送方 Sender 输入元素个数为 N_X 的集合 X 。这两个集合的元素比特长度都为 σ 。参数 N_X 、 N_Y 、 σ 是公开的。两个参与方输入正整数型参数 h 、 m 、 B , 以及一组哈希函数 $H_1, H_2, \dots, H_h: \{0, 1\}^{\sigma - \log_2 m} \rightarrow \{0, 1\}^{\log_2 m}$ 。定位函数为 $\text{Loc}_i(x) = H_i(x_L) \oplus x_R$, 其中 $x = x_L \parallel x_R$, $x_L \in \{0, 1\}^{\sigma - \log_2 m}$, $x_R \in \{0, 1\}^{\log_2 m}$ 。

输出:

①接收方输出一个基于置换的布谷鸟哈希表 $\mathfrak{B}_Y$, 使得集合 Y 中的元素或空值“ $\perp$ ”被插入到该表中。

②发送方输出一个基于置换的哈希表 $\mathfrak{B}_X$, 使得集合 X 中的元素或空值“ $\perp$ ”被插入到该表中。

(1) 发送方: 假设哈希表 $\mathfrak{B}_X$ 可以看成由 m 个箱子组成, 每个箱子的容量是 B , 初始化为 $\{(\perp, \perp)^B\}$ 。对于每个

集合元素 $x \in X$ 和每个 $i \in \{1, 2, \dots, h\}$, 发送方选取箱子的一个空格子, 即选择 $j \leftarrow [B]$, $\mathfrak{B}_x[\text{Loc}_i(x)][j] = \perp$, 然后设置 $\mathfrak{B}_x[\text{Loc}_i(x)][j] := (x_i, i)$ 。如果箱子是满的, 导致发送方无法选取到一个可用的空格子, 那么哈希表生成算法执行失败并输出 $\perp$, 否则输出 $\mathfrak{B}_x$ 。

(2) 接收方: 假设哈希表 $\mathfrak{B}_y$ 可以看成由 m 个箱子组成, 但每个箱子的容量只有 1, 初始化为 $(\perp, \perp)$ 。对于每个集合元素 $y \in Y$, 接收方执行以下过程:

① 设置 $\omega = y$, 随机选择 $i \leftarrow \{1, 2, \dots, h\}$ 。

② 定义并调用函数 $\text{Insert}(\omega, i)$: 交换 (ω, i) 和 $\mathfrak{B}_y[\text{Loc}_i(\omega)]$, 如果 $(\omega, i) \neq (\perp, \perp)$, 重复调用 $\text{Insert}(\omega, j)$, 其中 $j \leftarrow \{1, 2, \dots, h\} \setminus \{i\}$ 。

如果对于任何一个 $y \in Y$, 重复调用 Insert 函数的次数超过系统限制的最大调用次数, 接收方将终止哈希表生成算法并输出 $\perp$, 否则输出 $\mathfrak{B}_y$ 。

图 4 所示为哈希表的生成过程。发送方 Sender 针对集合 X 中的每个元素, 遍历所有的定位函数 $\text{Loc}_i(x)$, 多次得到存放位置的地址, 将元素左边部分数据装入相应箱子里的某一格中。接收方针对集合 Y 中的每个元素, 以布谷鸟哈希的方式, 以某个哈希函数来定位装入到相应的箱子中。然后以这两个哈希表作为输入来运行 4.3 节的基础协议。

如果 $\mathfrak{B}_x$ 中某个箱子的所有格子都被装满了, 并且又接到插入数据的操作, 那么这会导致哈希表生成失败。所以需要设置合理的 B, m 参数, 使得这种情况发生的概率非常小。当然, $\mathfrak{B}_x$ 和 $\mathfrak{B}_y$ 中某些位置没有插入任何元素数据, 可将其初始化为空值 “ $\perp$ ”。文献[37]推荐将 $\mathfrak{B}_x$ 中的空值设置为 $(m, 0)$, 将 $\mathfrak{B}_y$ 的空值设置为 $(m-1, 0)$ 。

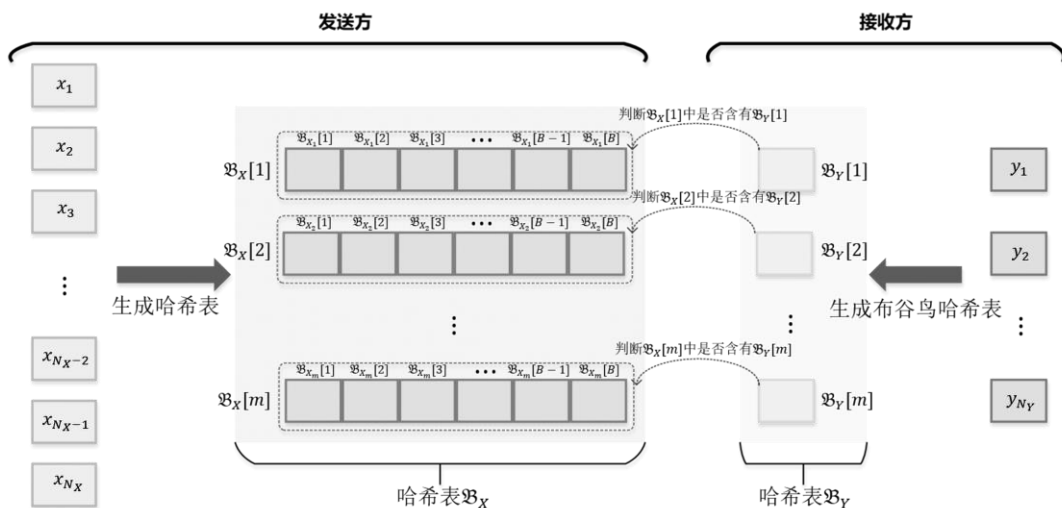


图 4 哈希表生成过程

3. 分割机制

生成哈希表后,面临的新任务是:判断 $\mathfrak{B}_Y[k]$ 是否为 $f_k(y) = \prod_{i=1}^B (y - \mathfrak{B}_X[k][i])$ 的根。按照 4.3 节的基础协议,需要计算一个 B 次多项式。令 $f_k(y) = r_k y^B + r_k a_{k,B-1} y^{B-1} + \cdots + r_k a_{k,1} y + r_k a_{k,0}$, 则可以为每个箱子 $\mathfrak{B}_X[k]$ 计算出相应的系数向量 $(r_k, r_k a_{k,B-1}, r_k a_{k,B-2}, \cdots, r_k a_{k,1}, r_k a_{k,0})$ 。为了计算 $f_k(\mathfrak{B}_Y[k])$, 可以将 $((\mathfrak{B}_Y[k])^B, (\mathfrak{B}_Y[k])^{B-1}, \cdots, (\mathfrak{B}_Y[k])^0, 1)$ 与系数向量做内积计算(图 5)。 $(\mathfrak{B}_Y[k])^B$ 的计算将消耗的同态乘法深度为 $\lceil \log_2(B+1) \rceil$ 。由于同态乘法操作非常耗时,噪声增长也较大,因此需要进一步优化,降低同态乘法深度,Chen 等人^[37]提出了分割机制、窗口机制来应对该问题。

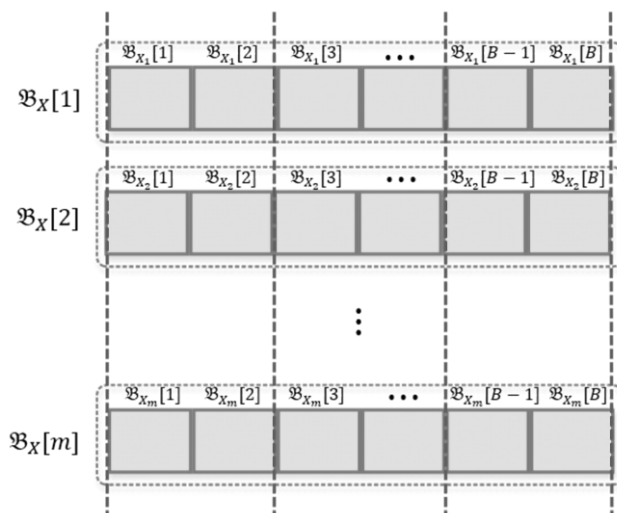


图 5 多项式分割

令 α 为分割机制的参数,将 $f_k(y)$ 用 α 个次数更小的多项式的乘积来表示, $f_k(y) = \prod_{j=1}^{\alpha} f_{k,j}(y)$, 令 $d' = \lceil B/\alpha \rceil$, 则可以粗略地表示为

$$f_{k,j}(y) = r_{k,j} \prod_{i=1}^{d'} (y - \mathfrak{B}_X[k][\alpha(j-1) + i]), \quad j = 1, 2, \dots, \alpha$$

这种分割机制虽然带来了同态乘法深度由 $\lceil \log_2(B+1) \rceil$ 降低到 $\lceil \log_2(B/\alpha + 1) \rceil$, 但这必须要求将 $f_{k,1}(\mathfrak{B}_Y[k]), f_{k,2}(\mathfrak{B}_Y[k]), \dots, f_{k,\alpha}(\mathfrak{B}_Y[k])$ 发送给接收方 Receiver, 这显然导致了发送方到接收方的通信量增加。由 4.3 节的基础协议可知, 对于每个元素 $\mathfrak{B}_Y[k]$, 在加密情况下, 发送

方需要计算 $d_{k,j} = f_{k,1}(\text{FHE.Enc}(\mathfrak{B}_Y[k]))$, $j=1,2,\dots,\alpha$ 。然后接收方将这 α 个密文进行解密, 查看解密出的 α 个明文是否含有 0, 若有则说明 $\mathfrak{B}_Y[k]$ 所对应的元素 y 属于交集 $X \cap Y$ 。

4. 窗口机制

采用分割机制后, 交集计算所需要的同态乘法深度为 $\lceil \log_2(B/\alpha+1) \rceil$ 。为了进一步降低同态乘法深度, Chen 等人^[37]引入了窗口机制, 基本原理为通过预先计算出一些低次单项式, 然后在交集计算时重构出高次单项式。

在分割机制中, 我们将交集计算的 B 次多项式拆解为 α 个 d' 次多项式的乘积, 令 $f_{k,j}(y) = r_{k,j}y^{d'} + r_{k,j}a_{k,d'-1}y^{d'-1} + \dots + r_{k,j}a_{k,1}y + r_{k,j}a_{k,0}$ 。每个箱子 $\mathfrak{B}_X[k]$ 预先计算出相应的系数向量 $(r_{k,j}, r_{k,j}a_{k,d'-1}, r_{k,j}a_{k,d'-2}, \dots, r_{k,j}a_{k,1}, r_{k,j}a_{k,0})$ 。同样, 为了计算 $f_{k,j}(\mathfrak{B}_Y[k])$, 需要 $((\mathfrak{B}_Y[k])^{d'}, (\mathfrak{B}_Y[k])^{d'-1}, \dots, \mathfrak{B}_Y[k], 1)$ 与系数向量做内积计算。

令窗口机制的参数为 ℓ , 我们可以将 $u \in \{1, 2, \dots, d'\}$ 进行 2^ℓ 进制分解, 则

$$u = \sum_{i=0}^{\lceil \log_2(d'/l) \rceil - 1} 2^{i\ell} \cdot u_i, \quad u_i \in \{0, 1, \dots, 2^\ell - 1\}$$

我们可以将计算 y^u 转换为计算 $\prod_{i=0}^{\lceil \log_2(d'/l) \rceil - 1} y^{2^{i\ell} \cdot u_i}$ 。由于 $y^0 = 1$, 因此对于 u_i 等于 0 的情况, 可以不用纳入此计算。那么我们可以预先计算:

$$y^{2^{i\ell} \cdot v}, 0 \leq i \leq \lceil \log_2(d'/l) \rceil - 1, 1 \leq v \leq 2^\ell - 1$$

然后重构出 $y^{d'}, y^{d'-1}, \dots, y$ 。将该窗口机制引入到隐私集合求交协议后, 需要提前计算:

$$\text{FHE.Enc}(pk, (\mathfrak{B}_Y[k])^{2^{i\ell} \cdot v}), 0 \leq i \leq \lceil \log_2(d'/l) \rceil - 1, 1 \leq v \leq 2^\ell - 1$$

显然, 在分割机制上使用窗口机制, 交集计算的同态乘法深度由 $\lceil \log_2(B/\alpha+1) \rceil$ 进一步降低到 $\log_2\left(\log_2\left(\left\lceil \frac{B}{\alpha\ell} \right\rceil + 1\right) + 1\right)$ 。窗口机制会导致增加接收方到发送方的通信开销。对于每个元素 $y \in Y$, 通信量由 1 个密文增加到 $\log_2\left(\left\lceil \frac{B}{\alpha\ell} \right\rceil + 1\right)(2^\ell - 1)$ 个密文。

5. 打包批量处理技术

分割机制、窗口机制都是在处理哈希表的每行, 以在密文下判断 $\mathfrak{B}_Y[k]$ 是否在箱子 $\mathfrak{B}_X[k]$ 中出现。由于每行的处理方法相同, 这非常适合使用批处理技术。基于 RLWE 的全同态加密方案能够在密文上执行 SIMD 操作, 该技术将一个向量加密成密文, 同态加法、乘法操作可实现向量相同位置的加法和乘法计算, 能够降低通信开销和计算开销。

在本协议中采用打包批量处理技术,可以对图 6 中的哈希表 $\mathfrak{B}_X$ 和 $\mathfrak{B}_Y$ 按列进行打包,然后进行批处理。令明文槽的总数为 n ,即一次可以加密 n 个明文。那么可以将通信开销和计算开销降低到原有的 $1/n$ 。

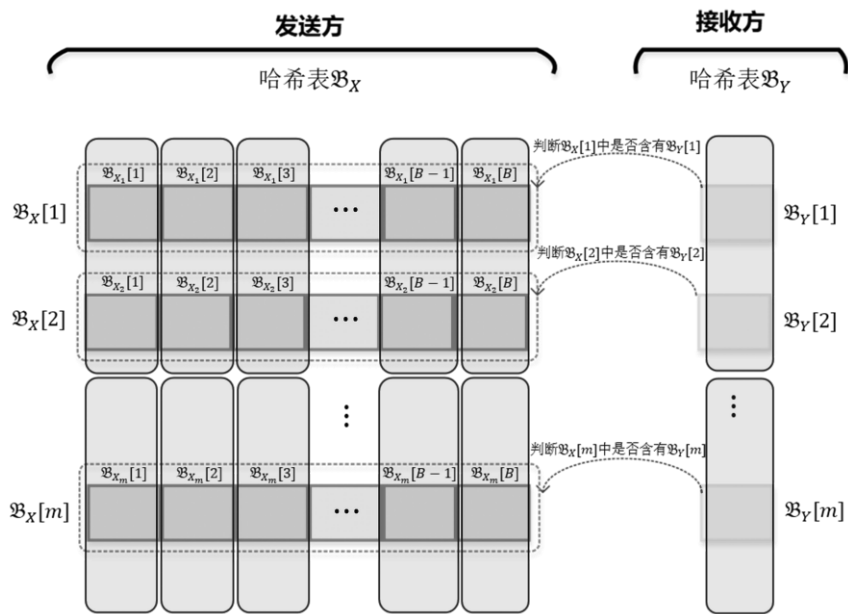


图 6 对哈希表采用打包批量处理技术

6. 模切换技术

模切换技术能够将密文的模值由较大的 q 降低为较小的 q' , 并且保持密文所对应的明文不变,最开始用于降低 BGV 全同态加密方案中密文噪声增长率。其实模切换还有一个好处: 将密文尺寸降低到原来的 $\frac{\log_2 q'}{\log_2 q}$ 。我们可以将模切换技术应用到密文计算结果返回阶段,压缩密文尺寸。Chen 等人在文献[37]中指出,用模切换技术,有效降低了 15%~20%的密文尺寸。

7. 完整协议

我们遵循了协议的实现过程将集合元素尺寸压缩、哈希表生成、分割机制、窗口机制、打包批处理、模切换等组件进行了详细分析。通过组合这些技术,就能够得到完整的隐私集合求交协议。

输入:接收方 Receiver 输入元素个数为 N_Y 的集合 $Y \subset \{0,1\}^\sigma$,发送方 Sender 输入元素个数为 N_X 的集合 $X \subset \{0,1\}^\sigma$ 。这两个集合中元素的比特长度都为 σ 。参数 N_X 、 N_Y 、 σ 是公开的。 κ 表示计算安全参数, λ 表示统计安全参数。

输出：接收方输出 $X \cap Y$ ；发送方输出 $\perp$ 。

(1) 执行哈希表生成：选择合适的哈希参数 h, m, B ，使得“哈希 hN_X 个球到最大容量为 B 的 m 个箱子，布谷鸟哈希 N_Y 个球到 m 个箱子”事件成功的概率 $\geq 1 - 2^{-\lambda}$ 。

① 哈希到更短的比特串：令 $\sigma' = 2\log_2(N_X + N_Y) + \lambda - 1$ 。如果 $\sigma > \sigma'$ ，那么发送方和接收方将自己的集合哈希到更短的比特表达。首先，选择一个随机哈希函数 $H': \{0, 1\}^\sigma \rightarrow \{0, 1\}^{\sigma'}$ ，令 $X' = \{H'(x) \mid x \in X\}$ ， $Y' = \{H'(y) \mid y \in Y\}$ ，在 (X', Y', σ') 执行后续协议，输出集合 X 和集合 Y 对应的元素作为交集。

② 哈希表生成：发送方和接收方将 h, m, B ，以及随机哈希函数为 $H_1, H_2, \dots, H_h: \{0, 1\}^{\sigma - \log_2 m} \rightarrow \{0, 1\}^{\log_2 m}$ 作为输入，运行哈希 Hashing 子程序。发送方执行哈希子程序的步骤 1，获得 $\mathfrak{B}_X$ 。接收方执行哈希子程序的步骤 2，获得 $\mathfrak{B}_Y$ 。

(2) 选择 FHE 参数：针对一个具有电路隐私的 IND-CPA 安全的全同态加密方案，发送方和接收方协商出参数 (n, q, t) 。选择足够大的 t ，满足 $\log_2 t > \sigma - \log_2 m + \lceil \log_2 h \rceil + 1$ 。

(3) 选择电路深度参数：发送方和接收方协商出窗口参数 $\ell \in [1, \log_2 B]$ 和分割参数 $\alpha \in [1, B]$ ，以尽可能地降低协议的总体开销。

(4) 预处理集合 X ：

① 分割：发送方将表 $\mathfrak{B}_X$ 纵向按照列分割成 α 个子表 $\mathfrak{B}_{X,1}, \mathfrak{B}_{X,2}, \dots, \mathfrak{B}_{X,\alpha}$ ，每个子表有 $B' = B/\alpha$ 列。

② 计算系数：针对每个子表的每一行 v ，发送者用多项式 $\prod_s (x - v_s)$ 的系数替换行 v ，即使用 $\text{Sym}(v) = \left((-1)^j \sum_{S \subseteq [B'] \mid |S|=j} \prod_{s \in S} v_s \right)_{0 \leq j \leq B'}$ 替换 v 。

③ 打包：对于每个子表，发送者将其每一列作为 m 维的向量，每个元素属于 $\mathbb{Z}_t$ 。那么发送者将每个向量打包为 m/n 个明文多项式。因此，每个子表都被转换为 $m/n \cdot B'$ 个多项式 $\mathbf{S}_{i,j}^{(r)}, 1 \leq i \leq m/n, 0 \leq j \leq B', 1 \leq r \leq \alpha$ 。

(5) 加密集合 Y ：

① 打包：接收者将表 $\mathfrak{B}_Y$ 理解为 $\mathbb{Z}_t$ 上的一个 m 维向量。将向量打包为 m/n 个明文多项式 $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_{m/n}$ 。

② 窗口化：对于每个打包的明文多项式 $\bar{Y}$ ，接收方逐分量计算 $\bar{Y}^{i \cdot 2^j}, 1 \leq i \leq 2^\ell - 1, 0 \leq j \leq \lfloor \log_2(B')/\ell \rfloor$ 。

③ 加密：接收方使用 FHE.Enc 加密每个 $\bar{Y}^{i \cdot 2^j}$ ，得到 m/n 个密文集合 $\{c_{i,j}\}$ 。接收方传输这些密文给发送方。

(6) 交集计算：

① 同态计算所有幂的加密：对每个密文集合 $\{c_{i,j}\}$ ，发送方同态计算一个向量 $\mathbf{c} = (c_0, c_1, \dots, c_{B'})$ ，以至 c_k 是 $\bar{Y}^k$ 的同态密文。最后，发送方得到 m/n 个向量 $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{m/n}$ 。

② 同态计算内积：发送方同态计算

$$r_{i,r} = \sum_{j=0}^{B'} c_i[B' - j] \cdot S_{i,j}^{(r)}, \quad 1 \leq i \leq m/n, \text{ 且 } 1 \leq r \leq \alpha$$

对密文 $r_{i,r}$ 执行模切换操作，以减小密文尺寸，并发送给接收方。

(7) 解密并获取结果：对于每个 $1 \leq r \leq \alpha$ ，接收方解密其收到的所有密文，并将 m/n 拼接成 1 个 m 维向量 $\mathfrak{R}_r$ 。最终，接收方输出 $Y \cap X = U_{1 \leq r \leq \alpha} \{y \in Y : \mathfrak{R}_r[\text{Loc}(y)] = 0\}$ 。

4.5 未来发展方向

在非平衡场景下，即一方的集合规模较多地大于另一方的集合时，基于全同态加密构造的隐私集合求交方案的计算效率和通信效率高于基于 OT 的 PSI 协议方案，这显示出较大的性能优势。在千万级或万级规模情况下，10Gbps 网络下单线程的交集计算耗时为 115.4s，在小规模数据场景已经达到实用化水平。目前基于全同态加密的 PSI 协议在面对大规模数据计算时仍然具有较大计算开销，并且对多个参与方的集合计算效率较低。

(1) 基于全同态加密的 PSI 协议高性能实现。现有方案一般以 SEAL、HElib 开源软件库为基础，这使得方案实现性能提升有限。由于当前全同态加密在 GPU、FPGA 上的实现技术发展迅速，加速比较快，因此研究基于全同态加密 PSI 协议在 GPU、FPGA 上的快速实现，将是一个未来发展方向，以提升这类协议的实现效率。

(2) 基于全同态加密的实用的多方 PSI 协议设计。2021 年，Badrinarayanan 等人在 PKC 2021 会议上对多方的情形进行了探索，但是实现效率较低，恶意安全考虑也不充分。因此，基于全同态加密的实用的多方 PSI 协议也是未来的发展方向之一，以不断提升其效率和安全性。

(3) 基于 NTRU/GSW 同态加密的 PSI 协议设计。现有方案主要以 BGV、BFV 方案为主来设计隐私集合求交方案，没有仔细考察不同方案对协议效率的影响。NTRU 方案的密文尺寸较短、GSW 密文外积的噪声增长小，可能是性能提升的努力方向。

5 总结与展望

时至今日，全同态加密算法已从理论研究阶段逐渐步入应用探索阶段，其在隐私保护机器学习、隐私信息检索、隐私集合求交等众多方面的应用已取得了较大的进展并持续吸引着研究者的关注。全同态加密在概念提出之际就有着“密码学中的瑞士军刀”的美称，这个称号体现了全同态加密算法能够在诸多安全场景下提供一种通用的、简洁的解决方案，即在某些具体场景下，基于全同态加密算法的方案或许并非最优，但其广泛的适用性和直观的设计

方式总能在各类应用场景中占据一席之地。随着全同态加密技术的改进及其与应用结合方式的进步,我们或将看到更多全同态加密技术在各类应用场景的落地部署,其将以“瑞士军刀”的姿态为我们在面对各类安全需求时提供一种普适的解决方案。

参考文献

- [1] GENTRY C. A fully homomorphic encryption scheme[D]. Stanford university, 2009.
- [2] BRAKERSKI Z, VAIKUNTANATHAN V. Efficient Fully Homomorphic Encryption from (Standard) LWE[C]//2011 IEEE 52nd Annual Symposium on Foundations of Computer Science. Palm Springs, CA, USA: IEEE, 2011: 97-106.
- [3] BRAKERSKI Z, GENTRY C, VAIKUNTANATHAN V. (Leveled) fully homomorphic encryption without bootstrapping[J]. ACM Transactions on Computation Theory (TOCT), 2014, 6(3): 1-36.
- [4] GENTRY C, SAHAI A, WATERS B. Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based[C]//CRYPTO 2013, Part I. Berlin: Springer, Heidelberg, 2013. volume 8042 of LNCS, 75-92.
- [5] CHEON J H, KIM A, KIM M, et al. Homomorphic encryption for arithmetic of approximate numbers[C]//ASIACRYPT 2017, Part I. Berlin: Springer, Heidelberg, 2017. volume 10624 of LNCS, 409-437.
- [6] GILAD-BACHRACH R, DOWLIN N, Laine K, et al. CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy[C]//International conference on machine learning. PMLR, 2016: 201-210.
- [7] NANDAKUMAR K, RATHA N, Pankanti S, et al. Towards deep neural network training on encrypted data[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.
- [8] FRERY J, STOIAN A, BREDEHOFT R, et al. Privacy-Preserving Tree-Based Inference with Fully Homomorphic Encryption[J]. arXiv preprint arXiv:2303.01254, 2023.
- [9] CHILLOTTI I, JOYE M, PAILLIER P. Programmable bootstrapping enables efficient homomorphic inference of deep neural networks[C]//Cyber Security Cryptography and Machine Learning: 5th International Symposium, CSCML 2021, Be'er Sheva, Israel, July 8-9, 2021, Proceedings 5. Springer International Publishing, 2021: 1-19.

- [10] LU W, HUANG Z, HONG C, et al. PEGASUS: bridging polynomial and non-polynomial evaluations in homomorphic encryption[C]//2021 IEEE Symposium on Security and Privacy (SP). IEEE, 2021: 1057-1073.
- [11] LEE E, LEE J W, LEE J, et al. Low-complexity deep convolutional neural networks on fully homomorphic encryption using multiplexed parallel convolutions[C]//International Conference on Machine Learning. PMLR, 2022: 12403-12422.
- [12] REN X, SU L, GU Z, et al. HEDA: Multi-Attribute Unbounded Aggregation over Homomorphically Encrypted Database[J]. Proceedings of the VLDB Endowment, 2022, 16(4): 601-614.
- [13] KIM A, SONG Y, KIM M, et al. Logistic regression model training based on the approximate homomorphic encryption[J]. BMC medical genomics, 2018, 11(4): 23-31.
- [14] KIM M, SONG Y, LI B, et al. Semi-parallel logistic regression for GWAS on encrypted data[J]. BMC Medical Genomics, 2020, 13: 1-13.
- [15] KIM D, PARK J, KIM J, et al. HyPHEN: A Hybrid Packing Method and Optimizations for Homomorphic Encryption-Based Neural Networks[J]. arXiv preprint arXiv:2302.02407, 2023.
- [16] HUANG Z, LU W, HONG C, et al. Cheetah: Lean and Fast Secure Two-Party Deep Neural Network Inference[C]//31st USENIX Security Symposium (USENIX Security 22). 2022: 809-826.
- [17] LU W, HUANG Z, ZHANG Q, et al. Squirrel: A Scalable Secure Two-Party Computation Framework for Training Gradient Boosting Decision Tree[C]//32nd USENIX Security Symposium (USENIX Security 23). 2023.
- [18] BOURA C, GAMA N, GEORGIEVA M, et al. Chimera: Combining ring-lwe-based fully homomorphic encryption schemes[J]. Journal of Mathematical Cryptology, 2020, 14(1): 316-338.
- [19] BOURSE F, MINELLI M, MINIHOLD M, et al. Fast homomorphic evaluation of deep discretized neural networks[C]//Advances in Cryptology-CRYPTO 2018: 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part III 38. Springer International Publishing, 2018: 483-512.
- [20] SANYAL A, KUSNER M, GASCON A, et al. TAPAS: Tricks to accelerate (encrypted) prediction as a service[C]//International conference on machine learning. PMLR, 2018: 4490-4499.
- [21] CHEN H, DAI W, KIM M, et al. Efficient homomorphic conversion between (ring) LWE ciphertexts[C]//Applied Cryptography and Network Security: 19th International Conference, ACNS 2021, Kamakura, Japan, June 21-24, 2021, Proceedings, Part I. Cham: Springer International Publishing, 2021: 460-479.

[22] JUVEKAR C, VAIKUNTANATHAN V, CHANDRAKASAN A. GAZELLE: A low latency framework for secure neural network inference[C]//27th USENIX Security Symposium (USENIX Security 18). 2018: 1651-1669.

[23] RATHEE D, RATHEE M, KUMAR N, et al. CryptFlow2: Practical 2-party secure inference[C]//Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. 2020: 325- 342.

[24] CHOR B, GOL DREICH O, KUSHILEVITZ E, et al. Sudan. Private Information Retrieval[C]//46th IEEE Symposium on Foundations of Computer Science (FOCS'95), Pittsburgh, PA, USA. 1995:41-50.

[25] FAN J, VERCAUTEREN F. Somewhat practical fully homomorphic encryption[J]. Cryptology ePrint Archive, 2012.

[26] AGUILAR MELCHOR C, BARRIER J, FOUSSE L, et al. XPIR : Private Information Retrieval for Everyone[J]. Proceedings on Privacy Enhancing Technologies, 2016, 2016(2): 155-174.

[27] ANGEL S, CHEN H, LAINE K, et al. PIR with Compressed Queries and Amortized Query Processing[C]//2018 IEEE Symposium on Security and Privacy (SP). San Francisco, CA: IEEE, 2018: 962-979.

[28] GENTRY C, HALEVI S. Compressible FHE with Applications to PIR[M]// HOFHEINZ D, ROSEN A. Theory of Cryptography: 11892. Cham: Springer International Publishing, 2019: 438-464.

[29] PARK J, TIBOUCHI M. SHECS-PIR: Somewhat Homomorphic Encryption-Based Compact and Scalable Private Information Retrieval[M]// Computer Security - ESORICS 2020: 12309. Cham: Springer International Publishing, 2020: 86-106.

[30] MUGHEES M H, CHEN H, REN L. OnionPIR: Response Efficient Single-Server PIR[C]//Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security. Virtual Event Republic of Korea: ACM, 2021: 2292-2306.

[31] AHMADI I, YANG Y, AGRAWAL D, et al. Addra: Metadata-private voice communication over fully untrusted infrastructure[C]//15th USENIX Symposium on Operating Systems Design and Implementation, OSDI2021, July 14-16, 2021.

[32] MENON S J, WU D J. Spiral: Fast, high-rate single-server PIR via FHE composition[C]//2022 IEEE Symposium on Security and Privacy (SP). IEEE, 2022: 930-947.

[33] PINKAS B, SCHNEIDER T, DARMSTADT T. Faster Private Set Intersection based on OT Extension[C]// Proceedings of the 23rd USENIX conference on Security Symposium. 2014:797-812.

[34] MEADOWS C. A more efficient cryptographic matchmaking protocol for use in the absence of a continuously available third party[C]//1986 IEEE Symposium on Security and Privacy. IEEE, 1986: 134-137.

[35] HUBERMAN B A, FRANKLIN M, HOGG T. Enhancing privacy and trust in electronic communities[C]//Proceedings of the 1st ACM conference on Electronic commerce. Denver Colorado USA: ACM, 1999: 78-86.

[36] FREEDMAN M J, NISSIM K, PINKAS B. Efficient Private Matching and Set Intersection[C]//Advances in Cryptology - EUROCRYPT 2004. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004: 1-19.

[37] CHEN H, LAINE K, RINDAL P. Fast Private Set Intersection from Homomorphic Encryption[C]//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. Dallas Texas USA: ACM, 2017: 1243-1255.

[38] PINKAS B, SCHNEIDER T, ZOHNER M. Scalable Private Set Intersection Based on OT Extension[J]. ACM Transactions on Privacy and Security, 2018, 21(2): 1-35.

[39] KOLESNIKOV V, KUMARESAN R, ROSULEK M, et al. Efficient Batched Oblivious PRF with Applications to Private Set Intersection[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. Vienna Austria: ACM, 2016: 818-829.

[40] CHEN H, HUANG Z, LAINE K, et al. Labeled PSI from Fully Homomorphic Encryption with Malicious Security[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. Toronto Canada: ACM, 2018: 1223-1237.

[41] CONG K, MORENO R C, DA GAMA M B, et al. Labeled PSI from Homomorphic Encryption with Reduced Computation and Communication[C]//Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security. Virtual Event Republic of Korea: ACM, 2021: 1135-1150.

[42] BADRINARAYANAN S, MIAO P, RAGHURAMAN S, et al. Multi-party Threshold Private Set Intersection with Sublinear Communication[C]//GARAY J A. Public-Key Cryptography-PKC 2021: Springer International Publishing, 2021: 349-379.

全同态加密的安全模型与安全性分析

陈隆¹, 周潭平², 郭晖³

1. 中国科学院软件研究所, 北京, 100878

2. 武警信息工程大学, 西安, 710086

3. 密码科学技术全国重点实验室, 北京, 100878

通讯作者: 陈隆, E-mail: chenlong@iscas.ac.cn

摘 要: 全同态加密 (Fully Homomorphic Encryption, FHE) 是一种重要的加密技术, 能够在不解密的情况下, 实现对加密数据的计算, 因此在隐私计算领域具有广泛的应用前景。当前主流的全同态加密方案均基于可证明安全理论。本文将介绍全同态加密的安全定义, 并介绍其安全假设的相关背景知识, 为相关领域研究者和工程技术人员提供参考。

关键字: 全同态加密; 安全模型; 格密码

The Security Definitions and Assumptions of Fully Homomorphic Encryption

CHEN Long¹, ZHOU Tanping^{1,2}, GUO hui³

1. Institute of Software Chinese Academy of Sciences, Beijing, 100878, China

2. College of Cryptography Engineering, Engineering University of People's Armed Police,
Xi'an 710086, China

3. The State Key Laboratory of Cryptology, Beijing, 100878, China

Corresponding author: CHEN Long, E-mail: chenlong@iscas.ac.cn

Abstract: Fully Homomorphic Encryption (FHE) is an important encryption technology, which can realize the computation of encrypted data without decryption, so it has a wide application prospect in the field of privacy computing. At present, the mainstream fully homomorphic encryption schemes are based on the provable security theory. In this paper, we will introduce the security

definition of fully homomorphic encryption and introduce the background knowledge of the underlying security assumptions, which will provide reference for researchers and engineers in related fields.

Keywords: Fully homomorphic encryption; Security models; Lattice based cryptography

1 引言

全同态加密 (Fully Homomorphic Encryption, FHE) 是一种重要的加密技术, 能够在不解密的情况下, 实现对加密数据的计算, 以达到相应明文计算的效果——“密文计算”, 即同态加密技术支持直接对密文数据进行有意义的计算。这种“可算不可见”的性质, 使同态加密在隐私保护的机器学习、智能电网、电子投票、基因数据的隐私处理等方面, 具有潜在的应用前景。

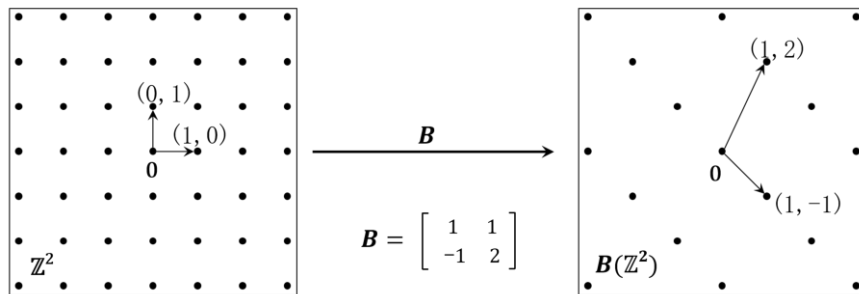
当前主流的全同态加密方案均基于可证明安全理论。具体来说, 常用的安全分析思路是首先建立安全模型, 通过构建归约, 将方案安全性建立在经典困难问题的复杂性基础之上; 然后通过评估给定参数的经典困难问题复杂性给出具体全同态加密方案的安全级别。由于主流全同态加密方案的安全性大多基于格上困难问题, 因此本文首先介绍格密码基础知识, 然后着重讨论全同态加密的安全模型和底层困难问题安全性, 从而揭示全同态加密的安全机理。

2 格密码基础

目前主流的同态加密通常基于格上困难问题假设构造, 属于基于格的密码方案的范畴, 也是抗量子密码体制的一个分支。基于格的密码体制凭借安全性强 (困难性基于最坏情况的格上问题)、计算效率高、可以抵抗已知量子算法攻击等优点, 受到密码学家的广泛关注。本节介绍格密码相关基础知识。

2.1 格的定义和性质

格是欧几里得空间中点的有规律的排列。最简单的 n 维空间格的例子是 $\mathbb{Z}^n$, 即所有 n 维整数向量的集合 (参阅图 1 左侧)。对整数格 $\mathbb{Z}^n$ 进行整系数线性变换后还是格, 如对整数格 $\mathbb{Z}^n$ 线性变换 $\mathbf{B}: \mathbb{R}^n \rightarrow \mathbb{R}^d$, 可以得到格 $\mathcal{L}(\mathbf{B}) = \{\mathbf{B}(\mathbb{Z}^n) = \mathbf{B}\mathbf{x} : \mathbf{x} \in \mathbb{Z}^n\}$ (参阅图 1 右侧)。等价地, 格 $\mathcal{L}(\mathbf{B})$ 可以描述为所有列向量的线性组合 $\sum_i \mathbf{b}_i x_i$, 其中 $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_n]$, $x_i \in \mathbb{Z}$ 。


 图 1 整数格 $\mathbb{Z}^2$ 和两维格 $B(\mathbb{Z}^2)$

定义 2-1 (格的定义) ^[1] 给定 k 个线性独立的向量 $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k \in \mathbb{R}^n$, 则由这组向量生成的格 $\mathcal{L}$ 记为

$$\mathcal{L}(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k) = \{z_1 \mathbf{b}_1 + z_2 \mathbf{b}_2 + \dots + z_k \mathbf{b}_k \mid z_i \in \mathbb{Z}\}$$

其中, 向量 $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k] \in \mathbb{R}^{n \times k}$ 被称为格 $\mathcal{L}$ 的基, 整数 k 和 n 分别被称为格的维数和秩。

格的定义与向量空间的定义非常类似, 但是格将线性组合的系数 z_i 限定为整数, 向量空间的系数是实数。同态加密方案中通常使用满秩的格, 满足条件 $k = n$ 。

“基本平行体” (Fundamental Parallelepiped) 是格理论中一个非常重要的概念。给定格 $\mathcal{L}$ 中的一个基 $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k \in \mathbb{R}^n$, 基本平行体定义为

$$P(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k) = \{z_1 \mathbf{b}_1 + z_2 \mathbf{b}_2 + \dots + z_k \mathbf{b}_k \mid -1/2 \leq z_i < 1/2\}$$

基本平行体中除了含有 0 这个格点, 不含有任何其他格点。通常一个格具有多个不同的基。格的多个不同基 $\mathbf{B}'$ 和 $\mathbf{B}$ 之间存在关系: 当且仅当存在一个幺模矩阵 $\mathbf{U}$ 使得 $\mathbf{B}' = \mathbf{B}\mathbf{U}$ 。格的多组基对应的基本平行体的体积是相同的。这个不变量称为格的行列式或格的体积, 格 $\mathcal{L}$ 的行列式定义为 $\det(\mathcal{L}) = |\det(\mathbf{B})|$ 。

在格密码方案的安全性分析过程中, 还会涉及对偶格 (Dual Lattice)、 q 元格等概念。

定义 2-2 (对偶格) 假设 $\mathcal{L}$ 是 $\mathbb{R}^n$ 上的一个格, 则 $\mathcal{L}$ 的对偶格 $\mathcal{L}^*$ 定义为

$$\mathcal{L}^* = \{\mathbf{x} \in \mathbb{R}^n \mid \forall \mathbf{b} \in \mathcal{L} : \langle \mathbf{x}, \mathbf{b} \rangle \in \mathbb{Z}\}$$

对偶格具有的性质: 对偶格的对偶格还是原来的格, 即 $(\mathcal{L}^*)^* = \mathcal{L}$; 如果 $\mathbf{B}$ 是格 $\mathcal{L}$ 的一组基, 那么有 $\mathcal{L}(\mathbf{B})^* = \mathcal{L}((\mathbf{B}^{-1})^T)$ 。

定义 2-3 (q 元格) 对于任意矩阵 $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, 其中 n, m, q 是整数。定义 q 元格为以下两个格:

$$\mathcal{L}_q(\mathbf{A}) = \{\mathbf{y} \in \mathbb{Z}^m \mid \mathbf{y} = \mathbf{A}^T \mathbf{s} \bmod q, \mathbf{s} \in \mathbb{Z}^n\}$$

$$\mathcal{L}_q^\perp(\mathbf{A}) = \{\mathbf{y} \in \mathbb{Z}^m \mid \mathbf{A}\mathbf{y} = 0 \bmod q\}$$

q 元格具有的性质: $\mathcal{L}_q^\perp(\mathbf{A})$ 是 $\mathcal{L}_q(\mathbf{A})$ 缩放过的对偶格, 满足 $\mathcal{L}_q^\perp(\mathbf{A}) = q \cdot \mathcal{L}_q(\mathbf{A})^*$ 。

2.2 理想格

介绍理想格之前, 我们需要补充相关数域的代数学基础知识。

定义 2-4 (H 空间^[2]) H 空间为 $\mathbb{R}^{s_1} \times \mathbb{C}^{2s_2}$ 的子空间, 定义为

$$H := \{(x_1, \dots, x_n) \in \mathbb{R}^{s_1} \times \mathbb{C}^{2s_2} : x_{s_1+s_2+j} = \overline{x_{s_1+j}}, \forall j \in 1, \dots, s_2\}$$

其中, s_1 和 s_2 均为正整数。

令 $s_1 + 2s_2 = n$, LRP10^[2]中证明 H 空间与 n 维实数空间 $\mathbb{R}^n$ 同构。

定义 2-5 (复数域上的分圆多项式) 定义复数 $\omega = \exp(2\pi\sqrt{-1}/m) \in \mathbb{C}$, m 次分圆多项式为 $\Phi_m(X) = \prod_{i \in \mathbb{Z}_m^*} (X - \omega^i) \in \mathbb{Z}[X]$ 。

注意: 复数域上的分圆多项式系数为整数, 且在复数域上不可约。特别地, 当 m 为 2 的幂次时, $\Phi_m(X) = X^{m/2} + 1$ 。

定义 2-6 (数域) 令 $f(x) \in \mathbb{Q}$ 为 n 次首一多项式且在有理数域 $\mathbb{Q}$ 上不可约。 ζ 为 $f(x)$ 的一个根, 即 $f(\zeta) = 0$ 。 n 阶数域 $K = \mathbb{Q}(\zeta)$ 表示将有理数域加入 ζ 得到的扩域, 而且存在自然的同构 $\zeta \mapsto x$ 使得 $K \cong \mathbb{Q}/(f(x))$ 成立。特别地, 当 $f(x)$ 为割圆多项式时, K 为割圆域。特别地, K 中的元素可以通过指数基 $\{1, \zeta, \dots, \zeta^{n-1}\}$ 的线性组合表示。

定义 2-7 (整数环) 数域 $K = \mathbb{Q}(\zeta)$ 上的整数环 R 定义为数域 K 中所有代数整数的集合, 即所有整系数多项式的集合 $R := \mathbb{Z}[\zeta] \cong \mathbb{Z}[X]/f(X)$ 。 R 可以视为以 $\{1, \zeta, \dots, \zeta^{n-1}\}$ 为基的 n 阶 $\mathbb{Z}$ 自由模 (free $\mathbb{Z}$ -module)。对于整数 q , 可以定义商环 $R_q = R/qR$ 。这里假设 R_q 中元素的系数都属于 $(-q/2, q/2]$ 。

由于数域 $K \cong \mathbb{Q}[x]/(f(x))$, 其中 n 次不可约多项式 f 存在 n 个复数域上的根。这些根彼此互为共轭根。固定一个根 ζ , 它的所有共轭根都可以表示成指数基 $\{1, \zeta, \dots, \zeta^{n-1}\}$ 的线性组合。于是可以定义 n 个 K 到复数域 $\mathbb{C}$ 的嵌入, 其中每个嵌入 $\sigma_i: K \rightarrow \mathbb{C}$ 将根 ζ 映射到它的某个共轭根。

定义 2-8 (正则嵌入) 正则映射 $\sigma: K \rightarrow \mathbb{C}^n$ 为上述 n 个嵌入的级联, 即

$$\sigma(a) = (\sigma_i(a))_{i \in n}, a \in K$$

同时定义正则嵌入范数为 $\|a\|_c^{\text{can}} = \|\sigma(a)\|_c, a \in R$ 。特别地, $\|a\|_c \leq c_m \cdot \|a\|_c^{\text{can}}$, 其中 c_m 为环常数。

定义 2-9 (理想) 理想 $\mathcal{I} \subseteq R$ 为 R 上的加法子群且关于乘以 R 上任意元素封闭, 即对任意 $r \in R$ 和 $x \in \mathcal{I}$ 有 $r \cdot x \in \mathcal{I}$ 。与整数环 R 一样, 任意理想可以视为秩为 n 的 $\mathbb{Z}$ 自由模, 因而也有 n 维整系数基。

将理想的定义更一般化, 我们可以定义分式理想的概念。

定义 2-10 (分式理想) 对于集合 $\mathcal{I} \subset K$, 如果对某个 $d \in R$ 有 $d\mathcal{I} \subset R$ 为理想, 那么 $\mathcal{I}$ 被称为一个分式理想。

为了区别, 我们称此前定义的理想为整理想。整理想可以视为特殊的分式理想。

定义 2-11 (理想格^[2]) 一个分式理想 $\mathcal{I} \subset K$ 通过正则映射可以得到 H 空间的一个格 $\mathcal{L} = \sigma(\mathcal{I}) \subset H$ 。这个格通常被称为理想格。自然地, 如果 $(b_1, \dots, b_n)$ 是理想 $\mathcal{I}$ 的整数基, 那么 $B := (\sigma(b_1), \dots, \sigma(b_n))$ 是格 $\mathcal{L}$ 的一组基。

2.3 格上困难问题

本节介绍格密码涉及的经典困难问题。

定义 2-12 (逐次最小和最短向量) 对于 n 维的格 $\mathcal{L}$, 任意的 $i \in [n]$, $\lambda_i(\mathcal{L})$ 表示能包含格中 i 个线性独立向量且中心在原点的最小的球体的半径。 $\lambda_i(\mathcal{L}) = \min\{r : \dim(\text{span}(\mathcal{L} \cap B(r))) \geq i\}$, 定义 $\lambda_1(\mathcal{L})$ 为格 $\mathcal{L}$ 中最短非零向量长度, 其中 span 是张成空间。

最短向量 $\lambda_1(\mathcal{L})$ 具有的性质: $\lambda_1(\mathcal{L}) \leq \sqrt{n}(\det(\mathcal{L}))^{1/n}$, $\lambda_1(\mathcal{L}) \geq \min_{i=1, \dots, n} \|\tilde{\mathbf{b}}_i\|_2$, 其中 $\tilde{\mathbf{b}}_i$ 是 Gram-Schmidt 正交化生成的基。

定义 2-13 (最短向量问题 SVP, Shortest Vector Problem) 给定格 $\mathcal{L}$ 的一个基 B , 求格 $\mathcal{L}$ 中的最短非零向量 $\mathbf{v} \in \mathcal{L}$ 。

在密码学中, 最短向量问题和最近向量问题并不适合直接应用于方案构造中, 方案中往往使用 SVP_γ 和 $\text{GapSVP}_{\gamma,r}$ 问题。

定义 2-14 (近似最短向量问题 SVP_γ , γ -approximate Shortest Vector Problem) 给定格 $\mathcal{L}$ 的一个基 B 和近似因子 γ , 求格 $\mathcal{L}$ 中的近似最短非零向量 $\mathbf{v} \in \mathcal{L}$, 满足 $\|\mathbf{v}\|_2 \leq \gamma \cdot \lambda_1(\mathcal{L})$ 。

定义 2-15 (判定性 SVP 问题 $\text{GAPSVP}_{\gamma,r}$, Decisional version of SVP) 给定格 $\mathcal{L}$ 的一个基 B , 间隙 $\gamma \geq 1$ 和参数 $r > 0$, 若 $\lambda_1(\mathcal{L}) \leq r$, 则输出 “YES”; 若 $\lambda_1(\mathcal{L}) > \gamma r$, 则输出 “NO”。

如果把输入的格限定为理想格, 就得到理想格上的 SVP 问题。

格上问题的困难性：上述格上问题经过数学家和密码学家广泛的研究，普遍被认为当近似因子足够小的情况下是计算困难的。目前已知的多项式时间算法，如 Arien Lenstra、Hendrik Lenstra 和 László Lovász 提出的 LLL 算法^[3]及其改进版^[4]只能解决亚指数近似因子 $\gamma = 2^{O(n \log \log n / \log n)}$ 的困难问题。而已知能解决多项式近似因子问题的算法^[5-8]，或者需要超多项式 $2^{O(n \log n)}$ 时间，或者需要 $2^{O(n)}$ 时间和空间。而对于计算复杂性而言，许多格问题甚至当近似因子取到亚多项式时都可以被证明是 NP 困难的，如 SVP 问题^[9-12]。然而，上述结论并不能直接应用于密码学。这是因为基于格的密码方案目前所基于的困难问题的近似因子至少为 $\gamma(n) \geq n$ 。有证据显示当近似因子大于 $\gamma(n) \geq \sqrt{n}$ 时，格上困难问题不是 NP 困难的，因为它们属于 $NP \cap coNP$ ^[13]。

2.4 格基约减算法

一个格可以用不同的格基来表示。在解决格上的困难问题时，同一个算法如果选择不同的格基作为输入，解决问题的时间和所得的结果差别十分明显。所以，研究如何找到有利于解决问题的那组基是有必要的。通常将寻找有利于解决格上问题的一组基的过程称为格基约减，将找到的基称为格上的一组约化基。本节介绍两个经典的格基约减算法：LLL（Arien Lenstra、Hendrik Lenstra 和 László Lovász）^[3]和 BKZ（Block Korkin-Zolotarev reduction）^[14]。BKZ 算法是一种块状迭代的格基约减算法，它是 LLL 算法的扩展。

1. LLL 算法

1982 年，LLL 算法由 Arjen Lenstra、Hendrik Lenstra 和 László Lovász 提出^[3]，它首先用来解决格中的最短向量问题（SVP）。其后的格基约减算法都由 LLL 算法衍生而来。LLL 算法的目标是找到格中一组近似正交的基。在介绍具体算法前，我们先介绍两个使用到的基本概念。

定义 2-16（施密特正交化） 给定 n 个线性独立的向量 $\{\mathbf{b}_1, \dots, \mathbf{b}_n\} \in \mathbb{R}^n$ ，向量的施密特正交化定义为 $\tilde{\mathbf{b}}_i = \mathbf{b}_i - \sum_{j=1}^{i-1} \mu_{i,j} \tilde{\mathbf{b}}_j$ ，其中 $\mu_{i,j} = \frac{\langle \mathbf{b}_i, \tilde{\mathbf{b}}_j \rangle}{\langle \tilde{\mathbf{b}}_j, \tilde{\mathbf{b}}_j \rangle}$ ， $1 \leq i \leq n$ ， $j < i$ 。定义 $\mu_{i,j}$ 组成的矩阵为施密特矩阵 μ 。

定义 2-17（ σ -LLL-约化基） 一组基 $\{\mathbf{b}_1, \dots, \mathbf{b}_n\} \in \mathbb{Z}^n$ 称为 σ -LLL-约化基，如果满足：

(1) Size 条件： $\forall 1 \leq i \leq n$ ， $j < i$ ，满足 $|\mu_{i,j}| \leq 1/2$ 。

(2) Lovász 条件： $\forall 1 \leq i \leq n$ ，满足 $\delta \|\tilde{\mathbf{b}}_i\|^2 \leq \|\mu_{i+1,i} \tilde{\mathbf{b}}_i + \tilde{\mathbf{b}}_{i+1}\|^2$ 。

在定义 2-7 中 Lovász 条件的含义是 $\tilde{\mathbf{b}}_{i+1}$ 不会比 $\tilde{\mathbf{b}}_i$ 短太多。LLL 算法如下：

算法 1: LLL 算法

 输入: 格 $\mathcal{L}$ 的一组基 $\{\mathbf{b}_1, \dots, \mathbf{b}_n\} \in \mathbb{Z}^n$

 输出: σ -LLL-约化基 $\mathcal{L}(B)$

 Start: compute $\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_n$

Reduction Step:

 for $i = 2$ to n do

 for $j = i - 1$ to 1 do

$$\mathbf{b}_i \leftarrow \mathbf{b}_i - c_{i,j} \mathbf{b}_j, \text{ 其中 } c_{i,j} = \left\lfloor \frac{(\mathbf{b}_i, \tilde{\mathbf{b}}_j)}{(\tilde{\mathbf{b}}_j, \tilde{\mathbf{b}}_j)} \right\rfloor$$

Swap Step

 if $\exists i$ s.t. $\|\tilde{\mathbf{b}}_i\|^2 > \|\mu_{i+1,i} \tilde{\mathbf{b}}_i + \tilde{\mathbf{b}}_{i+1}\|^2$ then

$$\mathbf{b}_i \leftrightarrow \mathbf{b}_{i+1}$$

goto start

 Output $\mathbf{b}_1, \dots, \mathbf{b}_n$

LLL 算法的约化效果明显、效率高。LLL 算法的循环次数为 $\text{poly}(M)$, 其中 $M := \max\{n, \log(\max_i \|\mathbf{b}_i\|)\}$ 。该算法可以输出一组较短的 σ -LLL-约化基, 满足 $\|\mathbf{b}_i\| \leq \left(\frac{4}{4\delta - 1}\right)^{(n-1)/2} \lambda_1(L)$ 。

2. BKZ 算法

1994 年, Schnorr 和 Euchner 提出一种 LLL 算法的衍生算法——BKZ 算法^[14]。BKZ 算法是目前最常用、最高效的格基归约算法, 在衡量全同态加密方案的安全性方面具有重要的作用。BKZ 算法将大维度的格基分成小区块, 并在小区块中调用 LLL 算法进行高效的约减。无论是 BKZ 算法还是其衍生和改进算法, 小区块中分组向量的长度 β 都是该类算法的一个重要参数。当 β 逐渐增大时, 算法的精确度会随之增加, 但是运行时间也会呈指数级增长。算法过程如下:

 算法 2: BKZ 算法^[15]

 输入: 格 $\mathcal{L}$ 的一组基 $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$, 分块长度 $\beta \in \{2, \dots, n\}$, $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ 生成的斯密特矩阵 μ 和 $\|\mathbf{b}_1^*\|^2, \dots, \|\mathbf{b}_n^*\|^2$ 。

 输出: BKZ 约化基 $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ 。

```

1.  $z \leftarrow 0, j \leftarrow 0$ ,  $\text{LLL}(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n, \mu)$  //LLL 约化基, 并更新矩阵  $\mu$ 

2. while ( $z < n-1$ ) do

3.    $j \leftarrow (j \bmod (n-1)) + 1; k \leftarrow \min(j + \beta - 1, n); h \leftarrow \min(k + 1, n)$  //定义区块大小

4.    $\mathbf{v} \leftarrow \text{Enum}\left(\mathbf{u}_{[j,k]}, \|\mathbf{b}_1^*\|^2, \dots, \|\mathbf{b}_n^*\|^2\right)$  //Enum 算法可以找到  $\mathbf{v} = (\mathbf{v}_j, \dots, \mathbf{v}_k) \in \mathbb{Z}^{k-j+1}$ , 满足  $\left\| \pi_j \sum_{i=j}^k \mathbf{v}_i \mathbf{b}_i \right\| = \lambda_1(L_{[j,k]})$ 

5.   if  $\mathbf{v} \neq (1, 0, \dots, 0)$  then

6.      $z \leftarrow 0, \text{LLL}\left(\mathbf{b}_1, \dots, \sum_{i=j}^k \mathbf{v}_i \mathbf{b}_i, \mathbf{b}_j, \dots, \mathbf{b}_n, \mathbf{u}\right)$  //在当前区块中插入新的向量  $\sum_{i=j}^k \mathbf{v}_i \mathbf{b}_i$ , 并去除一个向量使向量
     之间独立, 更新矩阵  $\mu$ 

7.   else

8.      $z \leftarrow z + 1; \text{LLL}(\mathbf{b}_1, \dots, \mathbf{b}_n, \mathbf{u})$  //在下一轮 Enum 算法之前, 使用 LLL 对区块进行约化

9.   end if

10. end while

```

最近 BKZ 算法的几个优化变种被提出^[15-19], 这些算法在一定程度上提升了 BKZ 算法的性能。

3 全同态加密常见困难假设

下面着重介绍全同态加密常见的困难假设, 包括 LWE 问题、RLWE 问题、NTRU 问题等。

3.1 LWE 假设

LWE (Learning With Errors) 假设^[20]是很多格密码方案构造的基础假设, LWE 假设的困难性可以归约到格上困难问题。

定义 3-1 (LWE 分布) 给定秘密向量 $\mathbf{s} \in \mathbb{Z}_q^n$, 在 $\mathbb{Z}_q^n \times \mathbb{Z}_q$ 上的 LWE 分布 $A_{q,s,\chi}$ 定义为 $(\mathbf{a}, b = \langle \mathbf{s}, \mathbf{a} \rangle + e \bmod q)$, 其中 $\mathbf{a} \in \mathbb{Z}_q^n$ 是均匀选取的, 噪声 e 从分布 χ 中选取。

定义 3-2 (LWE 搜索问题^[20]) 对于均匀选取的 $\mathbf{s} \in \mathbb{Z}_q^n$, 噪声分布 χ 和模数 q , 给定 LWE 分布 $A_{q,s,\chi}$ 上任意多的实例 $(\mathbf{a}_i, b_i) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$, $\text{LWE}_{n,q,\chi}$ 问题就是搜索秘密向量 $\mathbf{s}$ 。如果将 LWE

问题中, 实例个数限定为最多有 $m \in N$ 个实例, 可以得到 LWE 的变种 $\text{LWE}_{n,m,q,\chi}$ 。

密码学中使用更加广泛的是 LWE 问题的另一个版本: 判定性 LWE 问题。

定义 3-3 (LWE 判定问题^[20]) 对于均匀选取的 $\mathbf{s} \in \mathbb{Z}_q^n$, 噪声分布 χ 和模数 q , 给定任意多的实例 $(\mathbf{a}_i, b_i) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$, 要求判定实例是取自 $A_{q,\mathbf{s},\chi}$ 分布还是 $\mathbb{Z}_q^{n+1}$ 上的均匀分布。

这两个版本的问题可以相互归约。2005 年, Regev 证明了只要参数满足 $\alpha q \geq 2\sqrt{n}$, 那么在平均情况下解决 LWE 问题, 其困难性至少与使用量子归约算法求解近似格上标准困难问题相同, 其中近似因子是 $\tilde{O}(n/\alpha)$ ^[20]。2009 年, Peikert 给出了一个相同近似因子的经典归约而非量子归约, 但是需要满足 $q \geq 2^{n/2}$ ^[21]。2013 年, Brakerski 等人通过对 LWE 实例的模数和维数进行折中, 证明了模数为多项式级别时, 存在 LWE 问题到格上最坏困难问题的经典归约^[22]。

同态加密方案中需要对密文进行运算, 这会导致噪声的增长。为了控制噪声增长、提升方案效率, 在实际中我们常常使用 LWE 的各种变种, 如 binLWE 问题等。

binLWE 问题: 同态加密方案中为了控制噪声增长, 经常使用私钥均匀取自 $\{0,1\}^n$ 时的特殊 LWE 问题, 称为 binLWE 问题。

定义 3-4 (binLWE 问题)^[22] 假设 $k \geq 1$, $q \geq 1$ 且 $m \geq n \geq 1$ 是整数, 并令 $\varepsilon \in (0, 1/2)$, $\alpha > 0$, $\delta > 0$, 使得 $n \geq (k+1)\log_2 q + 2\log_2(1/\delta)$, $\alpha \geq \sqrt{\ln(2n(1+1/\varepsilon))/\pi}/q$, 则存在从 $\text{binLWE}_{n,m,q \leq \sqrt{10n\alpha}}$ 到 $\text{LWE}_{k,m,q,\alpha}$ 的高效归约算法。

进一步地, 2016 年, Cheon 等人^[23]证明了秘密 $\mathbf{s}$ 的分量取自 $\{0,1\}$ 并且汉明重量固定的情况时, LWE 问题仍然是困难的。

Hermite 正规型 LWE 问题 原始 LWE 问题要求秘密 $\mathbf{s}$ 取自 $\mathbb{Z}_q^n$ 上的均匀分布, 而 Hermite 正规型 LWE 问题中秘密 $\mathbf{s}$ 取自与噪声 $\mathbf{e}$ 相同的分布。2014 年, Bai 和 Galbraith 等人^[24]中给出了原始 LWE 问题到 Hermite 正规型 LWE 问题的归约。

3.2 RLWE 问题

虽然 LWE 问题的安全性基础被研究的比较充分, 但是基于 LWE 问题构造的密码方案往往密文和密钥尺寸过大。所以, Lyubashevsky 等人在 2010 年, 提出 LWE 问题在多项式环上的一个变种 RLWE 问题。基于 RLWE 问题构造的格密码方案, 不仅密文和密钥尺寸显著减少, 而且计算效率可以得到大幅提升。

定义 3-5 (RLWE 分布) 令 $f(x)$ 为分圆多项式, $R = \mathbb{Z}[x]/(f(x))$, $R_q = R/qR$, 对于均匀选取的秘密多项式 $s \in R_q$, 定义 $(a, b = as + e) \in R_q^2$, 并将其分布记为 $A_{s, \chi}$ (也称为 RLWE 分布), 其中 $a \in R_q$ 是均匀选取的多项式, 噪声 $e \in R$ 的系数是从分布 χ 中选取。

定义 3-6 (RLWE 搜索问题) RLWE 问题 $\text{RLWE}_{q, \Psi_\alpha}$ 为给定 RLWE 分布 $A_{s, \chi}$ 的任意多实例, 搜索秘密多项式 $s \in R_q$ 。

定义 3-7 (RLWE 判定问题) 令 $f(x)$ 为分圆多项式, $R = \mathbb{Z}[x]/(f(x))$, $R_q = R/qR$, 对于均匀选取的秘密多项式 $s \in R_q$, 给定任意多的实例 $(a, b) \in R_q^2$, 判定性 $\text{RLWE}_{n, q, \chi}$ 问题是判定实例是取自 R_q^2 上的均匀分布, 还是 RLWE 分布 $A_{s, \chi}$ 。

2010 年, Lyubashevsky 等人^[2]中证明了关于 RLWE 问题的困难性, 说明了 RLWE 问题到 SIVP 或 SVP 问题的归约, 以及存在 RLWE 问题的平均情况到最坏情况的归约。

定义 3-8 令 K 是维度为 $n = \varphi(m)$ 的 m 次分圆域而 $R = \mathcal{O}_K$ 是其中的代数整数环。令 $\alpha = \alpha(n) > 0$ 是噪声分布 Ψ_α 的参数, $q = q(n) \geq 2$, $q = 1 \bmod m$ 是一个不超过 $\text{poly}(n)$ 大小的素数, 并且使得 $\alpha q \geq \omega(\sqrt{\log_2 n})\alpha$, 则存在一个多项式时间的量子归约算法从 $\text{RLWE}_{q, \Psi_\alpha}$ 问题归约到近似因子为 $\tilde{O}(\sqrt{n}/\alpha)$ 的 SIVP 或 SVP 问题。

上述定义只限于分圆域, 而 Peikert 给出了任意数域上的 RLWE 的平均情况——最坏情况归约^[2]。不过, 为了实现方便, 很多方案使用环 $\mathbb{Z}(x)/(x^n + 1)$, 其中 n 为 2 的幂。

在分析具体困难性时, 我们将 Ring-LWE 的难度作为 LWE 问题进行分析, 因为到目前为止, 已知的最佳攻击并没有利用环结构。一般来说, 有许多算法需要考虑 (参见文献[25]中的总结), 但其中许多与我们的参数集无关。特别是, 因为只有 $m=n$ 个样本可用, 所以可以排除 BKW 类型的攻击^[26]和线性化攻击^[27]。这基本上只留下了两种 BKZ^[14-15]攻击, 通常称为原始攻击和对偶攻击, 下面将简要回顾它们。

原始攻击包括从 LWE 问题构建唯一 SVP 实例并使用 BKZ 解决它。我们研究了需要多大的块维度 b 才能使 BKZ 找到唯一的解。给定矩阵 LWE 实例 $(A, b = As + e)$, 我们构建了一个 $d = m + n + 1$ 维的格 $A = \{x \in \mathbb{Z}^{m+n+1} : (A | -I_m | -b)x = 0 \bmod q\}$, 体积为 q^m , 并且具有唯一的 SVP 解 $v = (s, e, 1)$, 其范数 $\lambda \approx \zeta \sqrt{n+m}$, ζ 为 s 和 e 每个分量的标准差。并且在具体分析中可以任意选择在 0 和 $2n$ 之间样本数量 m 来优化攻击效果。

对偶攻击包括在对偶格 A' 中找到一个短向量 $w \in A' = \{(x, y) \in \mathbb{Z}^m \times \mathbb{Z}^n : A^t x = y \bmod q\}$ 。假设我们已经找到了一个长度为 l 的向量 (x, y) , 并计算了 $z = v^t \cdot b = v^t As + v^t e = w^t s + v^t e \bmod q$,

如果 $(\mathbf{A}, \mathbf{b})$ 确实是一个 LWE 样本（否则它是模 q 的均匀分布），那么 $\mathbf{z}$ 的分布是标准差为 $l\zeta$ 的高斯分布。这两个分布的最大统计距离至少为 $\varepsilon = 4\exp(-2\pi^2\tau^2)$ ，其中 $\tau = l\zeta/q$ 。因此，对于长度为 l 的向量 $(\mathbf{x}, \mathbf{y})$ ，LWE 判定性具有优势 ε 。

在具体的方案安全评估中，可以利用开源软件工具来估计 LWE 实例的安全级别。

3.3 NTRU 问题

同态加密方案除了基于 LWE/RLWE 问题，还可能基于 NTRU 问题构造^[28-30]。

定义 3-9 (NTRU 问题) 令 n 为一个素数， q 为正整数， $f, g \in (\mathbb{Z}/q\mathbb{Z})[X]$ 为取自分布 χ 的 n 次小系数的多项式。其中 f 在环上 $R_q := (\mathbb{Z}/q\mathbb{Z})[X](X^n - 1)$ 。 f, g 两个多项式组成私钥，公钥为 $h := g / f \bmod R_q$ 。NTRU 问题定义为公钥 h 中恢复任意私钥的旋转 $(X^i f, X^i g)$ 。

NTRU 问题最早由 Hoffstein、Pipher 和 Silverman 提出^[31]。直到最近，对于 NTRU 格的格缩减攻击被认为，与具有相同参数的（环）LWE 格表现相似。然而，最近的研究^[32-33]表明，在大模数 q 下 NTRU 与 RLWE 问题安全性存在显著差距。Ducas 和 Woerden 指出^[34]，根据广泛的实验和分析，当 $q \ll 0.004 \cdot n^{2.484}$ 时 NTRU 问题仍然是足够安全的。

3.4 近似最大公因子问题

近似最大公因子（Approximate Greatest Common Divisor, AGCD）问题是数论困难问题。由 Howgrave-Graham 提出^[35]，之后被应用于构造一系列全同态加密方案^[36-41]。

定义 3-10 (近似最大公因子 AGCD 假设) 设 γ, n, ρ 是给定的正整数， p 是 ρ 位大奇数 ($2^{\eta-1} < p < 2^\eta$)，给定 AGCD 分布 $D_{\gamma, \rho}(p) = \{pq + r \mid q \leftarrow \mathbb{Z} \cap [0, 2\gamma/p), r \leftarrow (-2^\rho, 2^\rho)\}$ 。AGCD 假设是：给定分布 $D_{\gamma, \rho}(p)$ 中的多组实例 x_i ，计算 p 是困难的。

4 全同态加密的安全模型

全同态加密在现实中往往应用于云服务场景。在此类场景中，通常是用户将询问加密后提交服务提供商，服务提供商通过对密文进行计算并将密态的结果返回给用户。在这个过程中，用户处于隐私保护的目，希望询问和返回值都不会被服务提供商知晓。这就需要全同态加密可以保障消息的机密性。而服务提供商计算的函数通常是某些统计或人工智能推理模

型, 因此处于商业利益的考虑, 通常希望计算的函数不被用户获取。这就需要全同态加密可以保障电路的隐私性。下面将从机密性和电路隐私性两个角度讨论全同态加密的模型。

4.1 机密性

通常而言, 全同态方案必须满足基本的语义安全性, 即攻击者无法由密文得到明文的任何信息。在公钥密码学中, 语义安全通常由选择明文安全性来刻画。在选择明文安全性中, 攻击者可以选择若干明文, 并将其加密后获得相应的密文, 然后根据这些选择的明文和对应的密文来分析加密算法, 试图推断其他密文。选择明文安全性旨在确保密码方案在面对主动攻击时仍然保持机密性。需要注意的是, 确定性加密方案无法抵抗选择明文攻击, 因为相同的明文会加密成相同的密文, 攻击者可以通过分析重复的密文获得有关明文和密钥的信息。

1. IND-CPA 安全性

直观上, 如果一个全同态方案公开其加密算法和公钥, 攻击者可以获得任意明文的加密结果, 但仍然无法区分给定密文是对哪个消息的加密, 我们称该方案在选择明文攻击下满足不可区分性, 即 IND-CPA 安全性。

定义 4-1 (IND-CPA 安全性) 设 $E = (\text{KeyGen}, \text{Enc}, \text{Dec}, \text{Eval})$ 为全同态加密方案。定义以比特 $b \in \{0, 1\}$ 和概率多项式时间敌手 A 为参数的实验 $\text{Expr}_b^{\text{cpa}}[A]$:

$\text{Expr}_b^{\text{cpa}}[A](1^\kappa)$:

$(sk, pk, ek) \leftarrow \text{KeyGen}(1^\kappa)$

$(x_0, x_1) \leftarrow A(1^\kappa, pk, ek)$

$CT \leftarrow \text{Enc}_{pk}(x_b)$

$b' \leftarrow A(CT)$

返回 b'

如果对于任意概率多项式敌手 A , 有以下等式成立, 那么该方案是 IND-CPA 安全的:

$$\text{Adv}_{\text{cpa}}[A](\kappa) = \left| \Pr\{\text{Expr}_0^{\text{cpa}}[A](1^\kappa) = 1\} - \Pr\{\text{Expr}_1^{\text{cpa}}[A](1^\kappa) = 1\} \right| = \text{negl}(\kappa)$$

上面的定义表明, 敌手不能以不可忽略的概率区分 x_0 的加密和 x_1 的加密。注意, 如果加密方案是确定性的, 敌手可以加密消息后与 $\text{Enc}_{pk}(x_b)$ 进行比较, 从而获胜。因此, IND-CPA 安全性只有在加密方案是随机的情况下才能实现。

值得注意的是, 当前全同态加密方案的自举过程都需要用到秘密密钥的加密, 因此所有已知的全同态加密方案构造中都需要额外假设循环安全性。循环安全性意味着加密方案在攻

击者观察到该方案的秘密密钥在其公开密钥下的加密后仍然是安全的。

2. IND-CPA^D 安全性

Li 和 Daniele^[42]发现 CKKS 满足上述经典 IND-CPA 安全性, 但是敌手如何能够知道特定密文的解密结果, 由于解密结果包含密文中引入的噪声, 因此可能能够通过解密结果恢复密钥。为了应对这一严重的安全隐患, 加强全同态加密方案在近似加密情况下的安全性, 他们提出了 IND-CPA^D 安全性, 允许敌手进行有限制的解密查询。

定义 4-2 (IND-CPA^D 安全性) 设 $E = (\text{KeyGen}, \text{Enc}, \text{Dec}, \text{Eval})$ 为全同态加密方案 (可以是近似加密算法), 其公钥空间为 K , 明文空间为 M , 密文空间为 C 。定义如下实验 $\text{Expr}_b^{\text{indcpa}^D}[A]$, 其中参数 $b \in \{0, 1\}$, 敌手 A 可访问如下谕言机, 这些谕言机共享一系列由消息-消息-密文三元组组成的状态 $S \in (M \times M \times C)$:

(1) 加密谕言机 $E_{pk}(m_0, m_1)$, 给定一对明文消息 m_0, m_1 , 计算 $c \leftarrow \text{Enc}_{pk}(m_b)$, 扩展状态 $S := [S; (m_0, m_1, c)]$ 并将密文 c 返回给对手。

(2) 给定函数 $g: M^k \rightarrow M$ 和一个指标序列 $J = (j_1, \dots, j_k) \in \{1, \dots, |S|\}^k$, 计算密文 $c \leftarrow \text{Eval}_{pk}(g, S[j_1].c, \dots, S[j_k].c)$, 扩展状态 $S := [S; (g(S[j_1].m_0, \dots, S[j_k].m_0), g(S[j_1].m_1, \dots, S[j_k].m_1), c)]$, 将密文 c 返回给对手。这里和下面的 $|S|$ 表示序列 S 中三元组的数目, $S[j].m_0$ 、 $S[j].m_1$ 和 $S[j].c$ 表示 S 的第 j 个元素的 3 个分量。

(3) 解密谕言机 $D_{sk}(j)$, 给定一个索引 $j \leq |S|$, 检查是否有 $S[j].m_0 = S[j].m_1$, 若是, 则返回 $\text{Dec}_{sk}(S[j].c)$ 给敌手 (如果检查失败, 将返回错误符号 $\perp$)。

这个实验被定义为

$$\text{Expr}_b^{\text{indcpa}^D}[A](1^\kappa) : (sk, pk, ek) \leftarrow \text{KeyGen}(1^\kappa)$$

$$S := []$$

$$b' \leftarrow A^{E_{pk}, H_{ek}, D_{sk}}(1^\kappa, pk, ek)$$

返回 b'

敌手 A 关于该方案的 IND-CPA^D 安全性的优势是

$$\text{Adv}_{\text{cpa}^D}[A](\kappa) = \left| \Pr\{\text{Expr}_0^{\text{cpa}^D}[A](1^\kappa) = 1\} - \Pr\{\text{Expr}_1^{\text{cpa}^D}[A](1^\kappa) = 1\} \right| = \text{negl}(\kappa)$$

如果对于任何概率多项式时间敌手 A , 优势 $\text{Adv}_{\text{cpa}^D}[A](\kappa)$ 关于 κ 是可忽略的, 那么方案 E 是 IND-CPA^D 安全的。在定义 4-2 中, 用来询问解密查询的密文必须是由正确的加密算法

或同态评估算法（由谕言机 E 和 H 输出）诚实计算的密文。此外，为了避免频繁攻击，在询问解密谕言机时要求 $S[j].m_0 = S[j].m_1$ 。

精确加密方案可以被看作是近似加密的一种增加了正确性要求的特殊情况。定义 4-2 可以用来刻画精确的和近似的同态加密方案的安全性。相关文献证明了对于精确加密方案的情况，IND-CPA^D 与 IND-CPA 安全概念一致。

除此之外，公钥密码中常用的安全模型还包括 IND-CCA1 和 IND-CCA2 安全性。若敌手可以在有限时间内使用解密谕言机，则为 IND-CCA1；若敌手可以随时使用解密谕言机，则为 IND-CCA2。众所周知，如果一个加密方案具有同态性质，那么它就无法达到 IND-CCA2 安全性（因为敌手可以通过同态性质来“定向”改变挑战密文然后询问解密谕言机，并掌握解密后结果和挑战消息之间的关系）。下面讨论 IND-CCA1 安全性。

目前大部分 FHE 方案均起始于一个近乎同态加密方案（SHE 或 LHE），因此无法达到 IND-CCA1 安全性。因为 SHE/LHE 的密文在经过同态计算后噪声会放大，当噪声达到一定阈值会使得解密失败，所以仅能执行有限数量的同态操作。为了把基础方案扩展为 FHE 方案，需要通过自举（Bootstrapping）技术来减少密文的噪声。这个过程需要发布一些加密的秘密密钥材料，否则无法达到 IND-CCA1 安全性。其原因在于敌手可以选择密文来询问解密谕言机，也可以通过解密谕言机来解密加密的秘密密钥材料，从而解密他选择的其他任何密文。同理，如果 LHE 和 SHE 方案依赖于其他发布加密秘密密钥材料的技术，如密钥切换（如 BGV^[43]）或线性化（如 BV11a^[44]和 BV11b^[45]）等，那么也无法达到 IND-CCA1 安全性。此外，这些 SHE 和 LHE 方案还容易遭受自适应密钥恢复攻击：如果敌手可以访问解密谕言机，就能够恢复秘密密钥。Canetti 等人^[46]提出使用 zkSNARK 的通用构造方法来实现 IND-CCA1 安全性，但需要进一步研究使用更弱原语的可能性，如指定验证者的 zkSNARK，确定使用 Canetti 等人的构造方法实现 IND-CCA1 安全性所需的最小零知识特性是一个有趣的开放问题。此外，设计不可自举的 SHE 或 LHE 方案可能是实现 IND-CCA1 安全性或对抗自适应密钥恢复攻击的一种途径，因为这样可以产生一个足够“复杂”的解密过程。

4.2 电路隐私性

同态加密的方案一个重要的性质是电路隐私性。电路隐私性可以保障经过同态计算后的密文不会泄露具体操作的过程。此类安全性在隐私保护的代理计算中非常重要。例如，在隐私保护的机器学习推理场景，服务器拥有机器学习的模型，用户提供待推理的询问。用户将自己的询问通过同态加密方案加密后提交给服务器，服务器根据机器学习的模型对用户的询问进行同态推理计算，并将计算结果返回给用户。用户解密后得到推理结果。这里同态加密方案的机密性足以保障服务器无法获知用户的推理询问与推理结果；但是机密性自身不足

以保障用户无法获知机器学习模型的相关知识，因而可能会导致机器学习模型信息的泄露，给服务提供方带来损失。

严格来说，完美（或统计或计算）意义上的电路隐私性定义如下。

定义 4-3（电路隐私性） 若某个同态加密方案针对电路族 $\mathcal{C}$ 满足电路隐私性，则对于任意公私钥 (sk, pk) 和计算密钥 evk ，对任意电路 $C \in \mathcal{C}$ 和满足解密结果为 m_i 的密文 c_i ，由算法随机数产生的两个分布

$$D_1 = \text{Eval}(evk, C, c_1, \dots, c_n) \text{ 和 } D_2 = \text{Enc}(pk, C(m_1, \dots, m_n))$$

完美（或统计或计算）意义上不可区分。

上述定义声明了在密文上对特定电路进行运算的输出看起来像是由电路和相应明文生成的明文值 v 的加密输出。由于 v 只是另一个明文（ $v = C(m_1, m_2, \dots, m_n)$ ），因此很难确定它是如何生成的（难度程度从完美到计算层次逐渐递减）。

实现严格的电路隐私性往往并不容易。在具体构造电路隐私的全同态加密方案过程中，实现同态加密电路隐私的主要挑战在于避免从计算的密文中的噪声项中泄露信息。在所有现有的同态加密方案中，密文中都存在一个“噪声”项，这对于安全性是必要的。随着进行同态操作，噪声会增长和变化。然而，噪声同时可能会泄露有关函数 f 的信息，因此需要通过额外的手段抵消密文中的噪声项来实现电路隐私。由于电路隐私要求同态计算后的密文与新加密的密文不可区分，这意味着需要解密并重新加密同态计算后的密文，即对密文进行自举。在不知道解密密钥的情况下，可以通过两种方式之一安全地实现：①使用混淆电路^[8]，但这往往会牺牲同态加密方案的紧致性，即导致密文加公共参数的尺寸与计算电路规模相关；②借助解密密钥的密文对解密电路进行同态的解密操作^[8]，但这需要额外的循环安全性假设，同时自举操作的计算复杂度较高。

可见，在实际构造中，如果需要满足 Gentry^[47]给出的电路隐私的准确定义需要付出较高的代价。而在实际场景中往往更倾向于使用类似但略弱的安全定义，即函数隐私^[48-50]。重要的区别是函数隐私仅要求对加密数据进行不同电路的计算所产生的分布在统计上接近、计算上接近或相同。与之不同的是，电路隐私还要求这些分布与新加密的密文分布相同。函数隐私的安全性要求往往更贴合实际应用需求，因为电路规模和深度等信息往往并不属于隐私保护的范畴。

函数隐私的具体定义基于构造一个模拟器。其定义的内在逻辑是，如果模拟器在仅知道最终计算结果而不知道具体电路结构时，也能模拟出整个交互过程中所有的输出值，那么即可确定同态加密方案没有泄露同态计算函数的额外信息。

定义 4-4（函数隐私） 如果一个同态加密方案满足函数隐私，即存在一个多项式时间的模拟器使得对任意深度为 L 的电路 C ，对于其任意的 l 个明文输入 $x_1, \dots, x_l$ 和其被加密的密文 $c_1, \dots, c_l$ （其中 $c_i = \text{Enc}(sk, x_i)$ ），满足：

$$(\text{Eval}(\text{evk}, C, (c_1, \dots, c_l)), c_1, \dots, c_l, sk) \approx (\text{Sim}(C(x_1, \dots, x_l), (c_1, \dots, c_l)), c_1, \dots, c_l, sk)$$

即两个分布（统计意义上）不可区分。

值得注意的是，上述定义中考虑不可区分的输出值时也包括用于加解密的私钥 sk 。如此定义是为了保证代理计算应用场景中，用户即使拥有私钥也无法获知服务器对询问密文所进行的操作函数。除了通过自举实现函数隐私的方法，一种高效且通用的选择是使用“噪声掩盖”，它是由 Gentry 在其博士论文中引入的^[51]。其方法是在计算结束时添加一个更大的噪声；由于添加的噪声需要超多项式地大于在同态操作中积累的噪声，因此这要求新加密的密文中模数-噪声比率为超多项式，然而这会显著降低底层 LWE 问题的困难性，从而影响方案的安全性。此外，针对 GSW 类型的全同态加密方案，还可以结合自身具体设计的代数特点，基于多项式模数-噪声比的标准 LWE 问题以极低的代价实现函数隐私^[50]。

5 总结

全同态加密方案在实际应用中，需要同时考虑方案的安全性、方案的效率、同态计算能力。因此，在方案设计中，首先需要保证方案在给定的安全模型下可以基于标准的困难问题给出安全证明，其次需要结合应用中的同态计算能力，确定方案中密文模数、维度、私钥分布、噪声分布等参数。通常而言，同态加密算法中噪声会随着同态运算而增长，而当噪声增长到一定的幅度将导致密文无法正确解密。因此，需要根据噪声增长的情况给出密文模数、维度需要满足的条件（通常是密文模数的下限与维度的上限）。另外，底层问题的困难性确定了能够保障指定安全强度的同态加密方案中密文模数的上限、维度的下限。此外满足方案的安全性和功能性参数可能不止一组。在实际应用中，通常需要对多组满足条件的参数进行测试，找到性能最优的参数组合，达到功能性、安全性和效率性 3 个方面的最优平衡。

参考文献

- [1] MICCIANCIO D, REGEV O. Lattice-Based Cryptography[M]. Post-quantum cryptography: Springer, 2009: 147-191.
- [2] LYUBASHEVSKY V, PEIKERT C, REGEV O. On ideal lattices and learning with errors over rings[C]. Annual international conference on the theory and applications of cryptographic techniques, 2010: 1-23.
- [3] LENSTRA A K, LENSTRA H W, LOVÁSZ L. Factoring polynomials with rational coefficients[J]. Mathematische annalen, 1982, 261(ARTICLE): 515-534.

- [4] SCHNORR C P. A hierarchy of polynomial time lattice basis reduction algorithms[J]. Theoretical computer science, 1987, 53(2-3): 201-224.
- [5] KANNAN R. Improved algorithms for integer programming and related lattice problems[C]. Proceedings of the fifteenth annual ACM symposium on Theory of computing, 1983: 193-206.
- [6] AJTAI M, KUMAR R, SIVAKUMAR D. A sieve algorithm for the shortest lattice vector problem[C]. Proceedings of the thirty-third annual ACM symposium on Theory of computing, 2001: 601-610.
- [7] MICCIANCIO D, VOULGARIS P. A deterministic single exponential time algorithm for most lattice problems based on Voronoi cell computations[C]. Proceedings of the forty-second ACM symposium on Theory of computing, 2010: 351-358.
- [8] AGGARWAL D, DADUSH D, REGEV O, et al. Solving the shortest vector problem in 2^n time using discrete Gaussian sampling[C]. Proceedings of the forty-seventh annual ACM symposium on Theory of computing, 2015: 733-742.
- [9] AJTAI M. The shortest vector problem in L_2 is NP-hard for randomized reductions[C]. Proceedings of the thirtieth annual ACM symposium on Theory of computing, 1998: 10-19.
- [10] MICCIANCIO D. The shortest vector in a lattice is hard to approximate to within some constant[J]. SIAM journal on Computing, 2001, 30(6): 2008-2035.
- [11] KHOT S. Hardness of approximating the shortest vector problem in lattices[J]. Journal of the ACM (JACM), 2005, 52(5): 789-808.
- [12] HAVIV I, REGEV O. Tensor-based hardness of the shortest vector problem to within almost polynomial factors[C]. Proceedings of the thirty-ninth annual ACM symposium on Theory of computing, 2007: 469-477.
- [13] AHARONOV D, REGEV O. Lattice problems in $NP \cap coNP$ [J]. Journal of the ACM (JACM), 2005, 52(5): 749-765.
- [14] SCHNORR C P, EUCHNER M. Lattice basis reduction: Improved practical algorithms and solving subset sum problems[C]. International Symposium on Fundamentals of Computation Theory, 1991: 68-85.
- [15] CHEN Y, NGUYEN P Q. BKZ 2.0: Better lattice security estimates[C]. International Conference on the Theory and Application of Cryptology and Information Security, 2011: 1-20.
- [16] BAI S, STEHLÉ D, WEN W. Measuring, simulating and exploiting the head concavity phenomenon in BKZ[C]. Advances in Cryptology-ASIACRYPT 2018: 24th International Conference on the Theory and Application of Cryptology and Information Security, Brisbane, QLD, Australia, December 2-6, 2018, Proceedings, Part I 24, 2018: 369-404.

- [17] ALBRECHT M R, DUCAS L, HEROLD G, et al. The general sieve kernel and new records in lattice reduction[C]. Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2019: 717-746.
- [18] ALBRECHT M R, BAI S, LI J, et al. Lattice reduction with approximate enumeration oracles: practical algorithms and concrete performance[C]. Annual International Cryptology Conference, 2021: 732-759.
- [19] ALBRECHT M R, BAI S, FOUQUE P A, et al. Faster Enumeration-Based Lattice Reduction: Root Hermite Factor Time[C]. Annual International Cryptology Conference, 2020: 186-212.
- [20] REGEV O. On lattices, learning with errors, random linear codes, and cryptography[J]. Journal of the ACM (JACM), 2009, 56(6): 1-40.
- [21] PEIKERT C. Public-key cryptosystems from the worst-case shortest vector problem[C]. Proceedings of the forty-first annual ACM symposium on Theory of computing, 2009: 333-342.
- [22] BRAKERSKI Z, LANGLOIS A, PEIKERT C, et al. Classical hardness of learning with errors[C]. Proceedings of the forty-fifth annual ACM symposium on Theory of computing, 2013: 575-584.
- [23] CHEON J H, HAN K, KIM J, et al. A Practical Post-Quantum Public-Key Cryptosystem Based on[C]. International Conference on Information Security and Cryptology, 2016: 51-74.
- [24] BAI S, GALBRAITH S D. An improved compression technique for signatures based on learning with errors[C]. Topics in Cryptology-CT-RSA 2014: The Cryptographer's Track at the RSA Conference 2014, San Francisco, CA, USA, February 25-28, 2014. Proceedings, 2014: 28-47.
- [25] ALBRECHT M R, PLAYER R, SCOTT S. On the concrete hardness of learning with errors[J]. Journal of Mathematical Cryptology, 2015, 9(3): 169-203.
- [26] KIRCHNER P, FOUQUE P-A. An improved BKW algorithm for LWE with applications to cryptography and lattices[C]. Advances in Cryptology-CRYPTO 2015: 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I 35, 2015: 43-62.
- [27] ARORA S, GE R. New algorithms for learning in presence of errors[C]. International Colloquium on Automata, Languages, and Programming, 2011: 403-415.
- [28] ROHLOFF K, COUSINS D B. A scalable implementation of fully homomorphic encryption built on NTRU[C]. International Conference on Financial Cryptography and Data Security, 2014: 221-234.

[29] KLUCZNIAK K. NTRU-v-um: Secure Fully Homomorphic Encryption from NTRU with Small Modulus[C]. Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, 2022: 1783-1797.

[30] LÓPEZ-ALT A, TROMER E, VAIKUNTANATHAN V. On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption[C]. Proceedings of the forty-fourth annual ACM symposium on Theory of computing, 2012: 1219-1234.

[31] HOFFSTEIN J, PIPHER J, SILVERMAN J H. NTRU: A ring-based public key cryptosystem[C]. International algorithmic number theory symposium, 1998: 267-288.

[32] ALBRECHT M, BAI S, DUCAS L. A subfield lattice attack on overstretched NTRU assumptions: Cryptanalysis of some FHE and graded encoding schemes[C]. Annual International Cryptology Conference, 2016: 153-178.

[33] KIRCHNER P, FOUQUE P A. Revisiting lattice attacks on overstretched NTRU parameters[C]. Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2017: 3-26.

[34] DUCAS L, VAN WOERDEN W. NTRU fatigue: how stretched is overstretched[C]. Advances in Cryptology-ASIACRYPT 2021: 27th International Conference on the Theory and Application of Cryptology and Information Security, Singapore, December 6-10, 2021, Proceedings, Part IV 27, 2021: 3-32.

[35] HOWGRAVE-GRAHAM N. Approximate integer common divisors[C]. International Cryptography and Lattices Conference, 2001: 51-66.

[36] VAN DIJK M, GENTRY C, HALEVI S, et al. Fully homomorphic encryption over the integers[C]. Advances in Cryptology-EUROCRYPT 2010: 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, French Riviera, May 30-June 3, 2010. Proceedings 29, 2010: 24-43.

[37] CORON J-S, MANDAL A, NACCACHE D, et al. Fully homomorphic encryption over the integers with shorter public keys[C]. Annual Cryptology Conference, 2011: 487-504.

[38] CORON J S, LEPOINT T, TIBOUCHI M. Scale-invariant fully homomorphic encryption over the integers[C]. Public-Key Cryptography-PKC 2014: 17th International Conference on Practice and Theory in Public-Key Cryptography, Buenos Aires, Argentina, March 26-28, 2014. Proceedings 17, 2014: 311-328.

[39] CHEON J H, STEHLÉ D. Fully homomophic encryption over the integers revisited[C]. Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2015: 513-536.

[40] CHEON J H, KIM J, LEE M S, et al. CRT-based fully homomorphic encryption over the integers[J]. Information Sciences, 2015, 310: 149-162.

[41] CHEON J H, CORON J S, KIM J, et al. Batch fully homomorphic encryption over the integers[C]. Advances in Cryptology-EUROCRYPT 2013: 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings 32, 2013: 315-335.

[42] LI B, MICCIANCIO D. On the security of homomorphic encryption on approximate numbers[C]. Advances in Cryptology-EUROCRYPT 2021: 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, October 17-21, 2021, Proceedings, Part I 40, 2021: 648-677.

[43] BRAKERSKI Z, GENTRY C, VAIKUNTANATHAN V. (Leveled) fully homomorphic encryption without bootstrapping[J]. ACM Transactions on Computation Theory (TOCT), 2014, 6(3): 1-36.

[44] BRAKERSKI Z, VAIKUNTANATHAN V. Efficient fully homomorphic encryption from (standard) LWE[J]. SIAM Journal on computing, 2014, 43(2): 831-871.

[45] BRAKERSKI Z, VAIKUNTANATHAN V. Fully homomorphic encryption from ring-LWE and security for key dependent messages[C]. Annual cryptology conference, 2011: 505-524.

[46] CANETTI R, RAGHURAMAN S, RICHELSON S, et al. Chosen-ciphertext secure fully homomorphic encryption[C]. IACR International Workshop on Public Key Cryptography, 2017: 213-240.

[47] GENTRY C. Fully homomorphic encryption using ideal lattices[C]. Proceedings of the forty-first annual ACM symposium on Theory of computing, 2009: 169-178.

[48] ISHAI Y, PASKIN A. Evaluating branching programs on encrypted data[C]. Theory of Cryptography: 4th Theory of Cryptography Conference, TCC 2007, Amsterdam, The Netherlands, February 21-24, 2007. Proceedings 4, 2007: 575-594.

[49] ARMKNECHT F, BOYD C, CARR C, et al. A guide to fully homomorphic encryption[J]. Cryptology ePrint Archive, 2015.

[50] BOURSE F, DEL PINO R, MINELLI M, et al. FHE circuit privacy almost for free[C]. Advances in Cryptology-CRYPTO 2016: 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part II, 2016: 62-89.

[51] GENTRY C. A fully homomorphic encryption scheme[M]. Stanford university, 2009.

全同态加密技术综述

戴怡然^{*1,2,3}, 张江², 向斌武^{1,2,3}, 邓熹^{1,3}

1. 中国科学院信息工程研究所信息安全国家重点实验室, 北京, 100093

2. 密码科学技术全国重点实验室, 北京, 100878

3. 中国科学院大学网络空间安全学院, 北京, 100049

通讯作者: 戴怡然, E-mail: daiyiran@iie.ac.cn

摘要: 随着人工智能、云计算、区块链等技术的广泛应用, 数据的安全性和隐私保护问题日益突出。全同态加密技术可以实现对密文数据的任何同态运算, 是一种功能强大的加密技术, 也是一种能够解决用户隐私安全性问题的有效方法, 有着广阔的应用前景。该文对2009年以来已有的全同态密码方案进行了综述, 根据其核心技术分为4条路线, 分别对每条路线下方案的基本结构、算法的优化过程、发展历程进行了详细的分析与探讨。该文首先对全同态加密的类型进行了详细的阐述, 并对其安全性和数学原理进行了叙述。然后依不同的技术路线, 对各种典型的全同态加密方案进行重点分析, 总结出加密方案的一般结构, 归纳了自举算法的核心步骤, 并对其最新的研究进展进行探讨, 进而对各种不同加密方案的存储效率和计算速度进行全面的分析与对比。最后在前面分析的基础上分别从不同的角度对全同态加密方案的实际应用特点进行分析与对比, 并对各个全同态加密方案的发展趋势做了展望。

关键词: 全同态加密; 自举; BGV; GSW; CKKS

A Survey of Fully Homomorphic Encryption Technology

DAI Yiran^{1,2,3}, ZHANG Jiang², XIANG Binwu^{1,2,3}, DENG Yi^{1,3}

1.State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, 100093, China

2.State Key Laboratory of Cryptology, Beijing, 100878, China

3.School of Cyber Security, University of Chinese Academy of Sciences, Beijing, 100049, China

Corresponding author: DAI Yiran, E-mail: daiyiran@iie.ac.cn

Abstract: With the widespread application of technologies such as artificial intelligence, cloud computing, and blockchain, data security and privacy protection issues have become increasingly prominent. Fully homomorphic encryption technology can realize any homomorphic operation on ciphertext data. It is a powerful encryption technology and an effective method to solve user privacy and security problems. It has broad application prospects. This paper summarizes the existing fully homomorphic encryption schemes since 2009, divides them into four routes according to their core technologies, and makes a detailed analysis of the basic structure, algorithm optimization process, and development history of each route. and discuss. This paper first elaborates the types of fully homomorphic encryption, and describes its security and mathematical principles. Then, according to different technical routes, various typical fully homomorphic encryption schemes are mainly analyzed, the general structure of the encryption scheme is summarized, the core steps of the bootstrap algorithm are summarized, and the latest research progress is discussed. Comprehensive analysis and comparison of the storage efficiency and computing speed of various encryption schemes. Finally, on the basis of the previous analysis, the practical application characteristics of fully homomorphic encryption schemes are analyzed and compared from different angles, and the development trend of each fully homomorphic encryption scheme is forecasted.

Keywords: Fully Homomorphic Encryption; Bootstrapping; BGV; GSW; CKKS

1 引言

传统的密码体制往往涉及加密信息的交换,并在交换的节点中分享密钥,但是这样容易引发隐私泄露的问题,特别是需要将敏感数据交给第三方的场景,如云计算中服务器、提供商和云运营商都可以接触到用户的敏感信息。同态加密,作为密码学里一种特殊的加密方式,可以解决这个问题^[1-2]:在不提供解密密钥的情况下,同态密文发给任意的第三方进行计算,第三方发回密文形式的计算结果再由用户自行下载解密。这里用一个黄金加工的场景解释同态加密:小红有一块黄金想加工成首饰,她雇用小明来进行黄金加工,但小红又担心小明加工黄金的过程中会昧下黄金的边角料,于是她设计了一个密封的手套箱:一个密封的巨大透明箱,一侧有两个孔洞连接着一副非常结实的手套,箱子外的人可以通过将双手伸入手套中对箱内的物体进行操作。这个箱子还有一个上锁的开口,钥匙在小红手里,她可以通过这个开口将黄金放入箱子,也可以将加工后的金饰取出箱子。在整个加工过程中,小明只能隔着手套对黄金进行加工,不会直接接触到黄金,因此杜绝了昧下边角料的可能。这个场景中的

手套箱可以看成同态加密方案, 钥匙为加密方案中用户持有的密钥, 黄金为需要处理的消息, 金饰则为消息处理后的结果。小明不直接接触黄金就完成了对黄金的加工, 同态加密可类比于此过程: 第三方不知道消息(消息为密文)也可以完成对消息的处理。

全同态加密的概念最早于 20 世纪 70 年代末被提出, 1978 年, Rivest、Adleman 和 Dertouzos^[3]第一次提到了对密文进行计算, 来获得密文对应明文相同操作结果的系统构想, 后人把这个想法重新总结命名为全同态加密。全同态加密在私有计算外包方面具有很大的实用意义。例如, 在云计算^[8-9]、金融^[10]、医疗^[11]等领域, 由于不需要数据的解密就可直接运算, 它可以由不受信任的一方运行且不会暴露其输入和内部状态, 用户的数据安全得到有效保障。除此以外, 全同态加密作为格密码的一种, 具有抵抗量子算法攻击的能力, 特别是在大整数素因子分解问题(Prime Factorization)和离散对数假设(Diffie Hellman)等传统数论上的困难问题相继被量子计算攻破的今天, 抗量子加密方案的研究显得尤为重要, 全同态加密方案就是其中之一。

2009 年, Gentry^[7]提出了第一个基于理想格(Ideal Lattice)构造的全同态方案, 它可以支持任意电路的求值。Gentry 不仅提出了一个全同态方案, 而且提供了一种由类同态方案构造全同态方案的方法——自举。自举是目前全同态加密方案构造最通用途径, 它可以理解为加密方案同态的运行自己的解密算法, 将一个噪声接近临界值的密文“刷新”成一个噪声很低的新密文, 进而得以继续新一轮的运算。因为对称同态加密方案对云计算的适用性有限, 且存在通用转换可以将按比特加密的对称全同态转换成公钥全同态加密, 所以全同态加密方案的研究工作主要基于公钥加密体制。在 Gentry 思想的影响下, 涌现了许多新结构的同态加密方案, 除了基于格上困难学习假设(Learning With Errors, LWE)的方案, 还有基于 AGCD 的方案(如 Pereira)^[59], 基于 NTRU 算法的有 FINAL^[57]、Kamil^[58]和国内向斌武等人^[60]的最新研究成果, 其中基于 AGCD 和 NTRU 的方案目前均从基于(R)LWE 的全同态加密方案“移植”得到, 因此本文不对基于这两个困难问题的全同态加密方案做重点讨论。

本文提供了全同态加密自 2009 年 Gentry 取得突破性进展以来的一个全面的视角, 探究不同技术路线下全同态加密方案的发展脉络和相关应用。本文的第二节给出了全同态方案的概念和必要的安全性定义, 补充了格的数学定义, 以及基于格的密码学中用到的主要数学问题, 便于更好地理解后续不同全同态加密方案的具体构造; 第三节根据全同态方案的关键技术特点, 分成 4 个部分详细讲述了每类代表性全同态方案的主要结构、自举算法和对应的优化迭代, 以及具体方案实现的相关数据; 第四节根据前文提到的代表性全同态方案, 介绍了一些基于这些加密方案设计的全同态算法库及其支持的优化算法; 最后一节讨论了全同态领域发展的现状和目前存在的一些问题与挑战, 总结回顾了本文内容, 对未来全同态加密算法的改进优化方向做出了展望。

2 背景知识

本节的第一部分介绍同态加密方案的分类，重点说明全同态加密方案中的一些基本算法模型和同态加密相关的安全性定义；第二部分介绍格的数学定义，并描述基于格的同态加密方案的安全性所依赖的主要数学问题。

2.1 全同态加密算法模型及安全性质

根据密文数据上允许进行运算操作的种类和次数，同态加密方案可分为单同态加密、类同态加密、全同态加密三大类。

单同态加密（Partially Homomorphic Encryption, PHE）：仅能支持一种同态运算（同态乘法或同态加法），但是可以执行无限次该同态运算。例如，用 ElGamal^[4]函数相乘指数做加法 $g^{m_0} \cdot g^{m_1} = g^{m_0+m_1}$ 的性质可以构造支持加法同态的单同态加密方案，利用 RSA^[5]算法相乘底数做乘法 $m_0^e \cdot m_1^e = (m_0 \cdot m_1)^e$ 的性质则可以构造支持乘法同态的单同态加密方案。它们的局限性在于只能进行单一的加法或乘法运算，无法支持同时需要多种同态运算的计算场景。

类同态加密（Somewhat Homomorphic Encryption, SWHE）：可以支持多种同态运算，但是允许的运算次数有限。例如，基于配对的循环群加密算法^[6]可以同时支持加法和乘法的同态运算，但是乘法只能进行至多一次。相比于单同态加密仅支持单一运算，类同态加密方案已经初步实现支持混合电路同态运算，但对于一个任意逻辑和深度的函数仍无法实现完整的同态计算。此外，还有一种层次型同态加密（Leveled Fully Homomorphic Encryption），它可以在给定深度的电路上完成有限次的函数同态运算，执行同态乘法或加法的次数与电路深度呈正相关，本质上仍属于类同态加密的一种特殊情况。

全同态加密（Fully Homomorphic Encryption, FHE）：顾名思义，可以支持无限次、所有种类的电路同态运算。全同态加密方案支持对密文进行任意次数同态计算的特性，可以为任何想要的功能（函数）构建程序，这样的程序可以直接在密文输入上运行，产生密文为相应明文运算结果的加密。

对全同态密文的同态操作可以这样描述：给定一些密文，对明文的任何加法乘法操作都可以通过直接对密文进行操作且不需要解密。一个全同态加密方案 ϵ 共有 4 个算法：密钥生成算法 KeyGen、加密算法 Enc、解密算法 Dec、同态运算算法 Eval，加密方案的主要工作流程如下。

(1) 给定安全参数 λ ，运行密钥生成算法 KeyGen(1^λ)，生成公钥 pk 、私钥 sk ，同态计算公钥 evk 。

(2) 给定明文 $m_1, \dots, m_k$ 和公钥 pk , 对 $i \in [1, k]$, 运行加密算法 $\text{Enc}(pk, m_i)$, 生成密文 ct_i 。

(3) 给定函数 f 和一些密文 $ct_1, \dots, ct_k$, 同态计算公钥 evk , 运行同态运算算法 $\text{Eval}(f, evk, ct_1, \dots, ct_k)$, 获得密文 ct_f 在经过函数 f 同态运算的密文结果 ct_f 。

(4) 用私钥 sk 对密文 ct_f 运行解密算法 $\text{Dec}(sk, ct_f)$, 获得相应的明文 $m_f = f(m_1, \dots, m_k)$ 。

其中第(3)步的同态运算算法就是同态加密方案的核心, 对 Eval 输入密文可以实现任意函数的计算, 还可以计算解密函数, 这也是构造全同态加密方案的思想。

一个安全且合理的全同态加密方案具有 4 个安全性质: 正确性、语义安全性、同态性、紧凑性。

①正确性: 对于任意的消息 $m \in \{0, 1\}$, 经同态加密方案加密的消息都能正确解密, 即 $m = \text{Dec}(sk, \text{Enc}(pk, m))$ 。

②语义安全性: 一个任意多项式时间敌手 A 成功区分两个不同消息加密密文的优势是可忽略的, 也可以称这个加密方案是不可区分选择明文安全的 (IND-CPA 安全)。

$$\Pr[A(pk, \text{Enc}(pk, m=0)=1)] - \Pr[A(pk, \text{Enc}(pk, m=1)=1)] = \text{negl}(\lambda)$$

③同态性: 给定一个电路 f , 任意一组消息 $m_1, \dots, m_k$, 有

$$\text{Dec}(sk, \text{Eval}(f, evk, \text{Enc}(pk, m_1), \dots, \text{Enc}(pk, m_k))) = f(m_1, \dots, m_k)$$

在同态加密密文上所做运算解密后得到的明文, 与给直接密文对应明文做相同运算得到的结果相同。

④紧凑性: 对于任意电路 f , 任意一组密文 $ct_1, \dots, ct_k$, 经过同态运算后得到的密文长度与电路 f 无关, 即

$$|\text{Eval}(f, evk, ct_1, \dots, ct_k)| = \text{poly}(\lambda)$$

其中, 正确性和语义安全性与公钥加密方案中的定义相同, 这是一个健全的加密体系的必要条件; 同态性和紧凑性则是同态加密方案特有的安全性定义, 同态性是同态加密方案独特属性, 紧凑性是同态算法有意义的前提^[18], 如果没有紧凑性的限制, 假设将原始密文输入运算函数 $f(ct_1, \dots, ct_k)$ 直接作为新密文输出, 解密时就需要先对原始密文 $ct_1, \dots, ct_k$ 解密再将解密后的内容 $m_1, \dots, m_k$ 放入函数 f 中运算才能得到经过同态运算后的消息 $f(m_1, \dots, m_k)$, 导致的后果是密文会随着同态运算次数的增加越来越大, 这样不仅毫无意义, 而且对于一些需要非公开计算的函数 f , 最后输出与 f 强相关的密文结果也会泄露 f 的信息。

自举是构造全同态加密方案中核心的步骤, 它涉及公开加密密钥的密文, 为了保证密钥安全性需要满足循环安全 (Circular Security), 这也是 Gentry 在方案中提出的一个额外假设, 下面给出在公钥密码体制 (PKE) 下的定义。

定义 2-1 循环安全假设^[67] 令 $n = n(\lambda)$ 为多项式, $\text{PKE} = (\text{KeyGen}, \text{Enc}, \text{Dec})$ 为一个比特公钥加密方案, 其私钥长度为 l_{sko} 。对任意多项式时间的敌手 A , 若此公钥加密方案是 n -key 循环安全的, 则有

$$\text{Adv}_{\text{PKE}, A, n}^{\text{circ}}(\lambda) := 2 \left| \Pr \left[\text{Expt}_{\text{PKE}, A, n}^{\text{circ}}(\lambda) = 1 \right] - \frac{1}{2} \right| = \text{negl}(\lambda)$$

其中, $\text{negl}(\lambda)$ 是可忽略函数, 因此敌手获胜的概率是可忽略的。

该定义说明循环安全假设保证了敌手无法区分随机消息的密文与加密私钥的密文, 满足了公钥加密方案中语义安全的经典安全性要求。

2.2 格定义及格上困难问题

一个 k 维格是 R^n 的离散加性子群。令 $B = (b_1, \dots, b_k)$ 为 R^n 中线性无关的向量集合, 则 B 中元素的所有整数线性组合的集合可以定义为 B 生成的格 $\mathcal{L}(B)$:

$$\mathcal{L} = \mathcal{L}(B) = \left\{ \sum_{i=1}^k \gamma_i b_i : \gamma_i \in \mathbb{Z}, b_i \in B \right\}$$

其中, B 为格 $\mathcal{L}(B)$ 的基。

大多数已知格上的定义都是基于距离的概念。具体来说, 对于 R^n 中的任意向量 $\mathbf{t}$ 和格 $\mathcal{L}$ 中的任意元素 $\mathbf{v}$, 这两个向量之间的距离定义为 $\text{dist}(\mathbf{t}, \mathbf{v}) = \|\mathbf{t} - \mathbf{v}\|_2$ 。因此, $\mathbf{t}$ 与 $\mathcal{L}$ 中任意元素之间的最小距离为

$$\text{dist}(\mathbf{t}, \mathcal{L}) = \min \|\mathbf{t} - \mathbf{v}\|_2 : \mathbf{v} \in \mathcal{L}$$

定义格 $\mathcal{L}$ 的最小距离为 $\lambda_1(\mathcal{L})$, 即 $\mathcal{L}$ 中最短的非零向量的长度, 即

$$\lambda_1(\mathcal{L}) = \min \{\|\mathbf{v}\|_2 : \mathbf{v} \in \mathcal{L}, \mathbf{v} \neq \mathbf{0}\}$$

许多基于格的全同态方案的安全性依赖于最短向量问题 (Shortest Vector Problem, SVP) 及其变体的困难性。SVP 表示在给定格中寻找最短的非零向量。若将给定的格限制为理想格, 则可以得到 IdealSVP^[13]。

定义 2-2 SVP 问题^[13] (1) γ -近似最短向量问题 (SVP_γ): 确定一个近似最短的向量。给定 $\gamma > 1$, 在格上找到一个非零向量 $\mathbf{v} \in \mathcal{L}$ 使得 $\|\mathbf{v}\| \geq \gamma \cdot \lambda_1(\mathcal{L})$ 。

(2) 判定最短向量问题 ($\text{GapSVP}_{\gamma, r}$): 判定最短向量的边界是否正确。给定 $\gamma > 1, r > 0$, 判断是否满足 $\lambda_1(\mathcal{L}) \leq r$ 或 $\lambda_1(\mathcal{L}) \geq \gamma \cdot r$ 。

注意, 如果最短向量问题是可解的, 那么判定最短向量问题也是可解的, 即 $\text{GapSVP}_{\gamma, r} \leq \text{SVP}$ 。

最近向量问题 (Closest Vector Problem, CVP) 是最短向量问题的一个推广。给定非零向量 $\mathbf{t} \in R^n$, 在格 $\mathcal{L}$ 中寻找一个最接近 $\mathbf{t}$ 的向量 $\mathbf{v}$, 即 $\text{dist}(\mathbf{t}, \mathbf{v}) = \text{dist}(\mathbf{t}, \mathcal{L})$ 。

定义 2-3 CVP 问题^[14] (1) γ -近似最短向量问题(CVP $_{\gamma}$): 在格中找到一个与目标向量最接近的向量。给定 $\gamma \geq 1, \mathbf{t} \in R^n$, 找到向量 $\mathbf{v} \in \mathcal{L}$ 使得 $\text{dist}(\mathbf{t}, \mathbf{v}) \leq \gamma \cdot \text{dist}(\mathbf{t}, \mathcal{L})$ 。

(2) 判定最近向量问题(DCVP $_{\gamma, r}$): 给定 $\gamma \geq 1, r > 0, \mathbf{t} \in R^n$, 判断是否满足 $\text{dist}(\mathbf{t}, \mathcal{L}) \leq r$ 或 $\text{dist}(\mathbf{t}, \mathcal{L}) \geq \gamma \cdot r$ 。

错误学习问题 (LWE) 是由 Regev^[16]在 2005 年引入的, 作为 Blum 等人^[49]的“奇偶校验学习”问题的扩展, 在密码学的发展中起到关键作用。

定义 2-4 LWE 问题^[16] 给定一个向量 $\mathbf{b} \in \mathbb{Z}_q^m$ 和矩阵 $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$, LWE 问题即为寻找一个向量 $\mathbf{s} \in \mathbb{Z}_q^n$, 使得 $\mathbf{A} \cdot \mathbf{s} + \mathbf{e} = \mathbf{b}$, 其中 $\mathbf{e} \in \mathbb{Z}_q^m$ 为误差分布 χ 的抽样。

2005 年, Regev^[16]通过量子归约将格中最坏情况下的判定最短向量 GapSVP 归约为 LWE 问题。也就是说, 如果能找到一个算法在多项式时间内解决 LWE 问题, 那么也能在多项式时间内解决 GapSVP 问题。因此, 基于 LWE 的同态加密方案的安全性与 SVP 及其变体的格上困难问题密切相关。

定义 2-5 RLWE 问题^[24] 给定多个抽样 $(a, b = a \cdot s + e) \in R_q \times R_q$, 寻找 $s \in R_q$, 其中 a 为 R_q 中均匀随机的抽样, $e \in R_q$ 为误差分布 χ 的抽样。

环上错误学习 (Ring Learning With Errors Problem, RLWE) 问题是 LWE 问题的环版本。从实际应用角度来看, 基于 RLWE 问题的同态加密方案比基于 LWE 问题的同态加密方案计算效率更高, 密文大小更紧凑。

3 全同态加密的技术路线

根据代表性全同态方案核心技术, 可以按照技术路线划分为以下几类。第一类的代表是 Gentry^[7]在 2009 年提出的全同态加密方案, 它找到了“理想格”这个既支持同态加法又支持同态乘法的工具, 还开创性地提出了自举的思想, 使得类同态加密方案都可以通过自举构造全同态加密方案。第二类是 Brakerski 和 Vaikuntanathan^[27]在 2011 年提出的类同态加密方案, 它率先提出了密钥切换操作控制密文乘法的维数扩展, 模切换操作控制噪声增长速度的想法。在此基础上, 后续的加密方案对 BV11 的效率和噪声控制进行了优化, 并实现了相应的自举算法。第三类是 Gentry、Sahai 和 Waters^[42]提出的基于矩阵的近似特征向量技术的同态加密方案, 它的技术特点是使用了矩阵的密文形式, 矩阵密文做乘法避免了向量密文的维数扩张问题, 同时将乘法噪声的增长速度由指数级降低到线性级。第四类是 Jung Hee Cheon 团队^[32]

在 2017 年提出近似同态加密方案，它适用于一些不需要精确计算的场景，在自举操作中用多项式函数近似计算替代了精确的比特提取。下面将对每类代表性全同态加密方案的理论基础、加密方案的结构通式、核心的技术特点、算法的发展过程、目前方案存在的局限性以及实现效果等方面进行叙述分析。

全同态加密的不同技术路线分支及其代表性方案如图 1 所示。

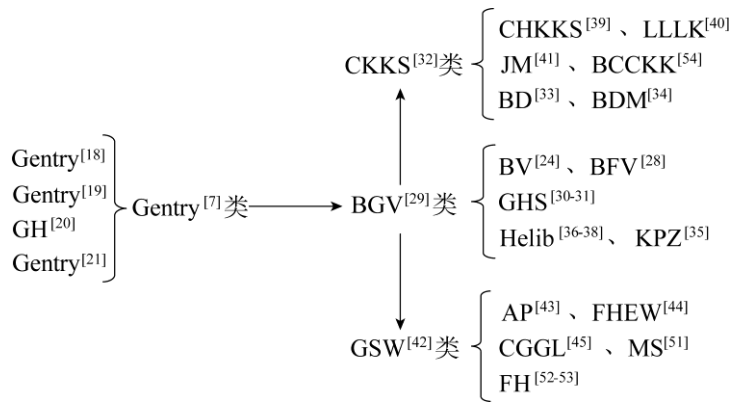


图 1 全同态加密的不同技术路线分支及其代表性方案

3.1 理想格和自举理论

以 2009 年 Gentry 的工作为代表的同态加密方案，这类方案构造的安全性是基于理想格上的困难问题^[12]。格上的两个著名困难问题是最近向量问题（CVP）^[13]和最短向量问题（SVP）^[14]，分别在格上找一个已知格点的最近向量和最短向量，后续又在此基础上引申出格基约减问题^[15]，为给定的格找一组尽可能接近正交的、既正交叉短的“好”格基。若已知一组“好”格基，则求解 CVP 和 SVP 问题只需要多项式的时间；反之，则需要指数级的时间。在 Gentry 的方案中，密钥就相当于一组“好”格基，求解密文可以归约到求解 CVP 和 SVP 问题。

3.1.1 具体方案

因为理想格本身支持加法和乘法，所以是用于构建完全同态加密的自然数学对象。Gentry 在方案中定义的理想格具有两个性质：首先在理想格中假设一个随机元素来屏蔽消息，保证加密的安全性；其次可以生成一个消除理想格的秘密陷门。依据这两个属性可以实现密文的同态加法和同态乘法。方案的主要思想是先构造一个支持有限次加法和乘法同态计算的类同态加密方案，然后借助稀疏子集和问题（Sparse Subset Sum Problem，SSSP）压缩解密电

路^[19]，降低电路深度，使得这个解密电路可以被同态计算，接着又基于循环安全假设(Circular Security)进行自举操作，最终实现全同态加密。这里给出以 Gentry 方案为代表的同态加密方案的一般结构：

密钥生成：对于给定的环 R 和理想 I 的基 B_I ， $\text{IdealGen}(R, B_I) = (B_J^{\text{sk}}, B_J^{\text{pk}})$ ，满足 $I + J = R$ ，其中 B_J^{sk} 、 B_J^{pk} 分别为理想 J 的基 B_J 组成的私钥和公钥。

加密：选择随机向量 $\mathbf{r}$ 、 $\mathbf{g}$ ，消息 $m \in \{0,1\}$ ，生成密文

$$ct = \text{Enc}(m) = m + \mathbf{r} \cdot B_I + \mathbf{g} \cdot B_J^{\text{pk}}$$

解密： $\text{Dec}(ct, B_J^{\text{sk}}) = (ct - B_J^{\text{sk}} \cdot (B_J^{\text{sk}})^{-1} \cdot ct) \bmod B_I$

密文加法：

$$ct_1 + ct_2 = \text{Enc}(m_1) + \text{Enc}(m_2) = m_1 + m_2 + (\mathbf{r}_1 + \mathbf{r}_2)B_I + (\mathbf{g}_1 + \mathbf{g}_2) \cdot B_J^{\text{pk}}$$

密文乘法：

$$\begin{aligned} ct_1 \cdot ct_2 &= \text{Enc}(m_1) \cdot \text{Enc}(m_2) \\ &= m_1 m_2 + (m_1 \mathbf{r}_2 + m_2 \mathbf{r}_1 + \mathbf{r}_1 \mathbf{r}_2 B_I)B_I + (m_1 \mathbf{g}_2 + m_2 \mathbf{g}_1 + \mathbf{g}_1 \mathbf{g}_2 B_J^{\text{pk}})B_J^{\text{pk}} + (\mathbf{r}_1 \mathbf{g}_2 + \mathbf{r}_2 \mathbf{g}_1)B_I B_J^{\text{pk}} \end{aligned}$$

此类同态方案正确的要求是噪声量 $ct \bmod B_J^{\text{pk}}$ 要小于 $B_J^{\text{pk}} / 2$ ，因此乘法操作后剧烈增长的噪声限制了同态操作的继续进行。

3.1.2 自举算法发展

在 Gentry 的著作中提出了一个关键技术——自举，这个术语表示刷新一个密文的操作过程：自举电路由同态方案的解密电路组成，旧密文输入这个自举电路产生一个新的密文，该密文加密相同的消息，但具有较低的噪声水平，可以对其进行更多的同态操作。这个操作是迄今为止大部分全同态加密方案的核心，更通俗地讲，就是把一个噪声马上达到临界点的同态密文二次加密，并且用加密的私钥同态计算解密算法，把内层的密文解密成明文，只留外层的加密，这样就获得了一个全新的低噪声同态密文且加密的内容不变。为了实现自举，Gentry 设计了一种压缩类同态方案解密函数的方法。这种方法将原来的类同态方案 ε 转换成另一个具有相同的同态容量和更简单的允许自举的解密函数全同态加密方案 ε^* 。同时为了降低解密算法的复杂性，Gentry^[50]证明了在同态计算公钥中添加一些关于私钥的“额外信息”就足够了。这需要针对私钥设置一个额外的安全假设，称为循环安全假设，这个假设意味着公开发布用公钥加密私钥的密文不会影响私钥安全性，即它可以防止敌手观察到该方案在公钥下加密的私钥。目前的全同态加密方案结构需要私钥的加密是可自举的，因此所有已知的全同态结构都要求循环安全性。但是目前循环安全性并没有严谨的数学证明，因此它通常被作为方案基础安全假设之上的附加假设。

接下来将给出自举的具体算法流程^①。考虑一个类同态加密方案 ε , 两个公私钥对 (pk_1, sk_1) 、 (pk_2, sk_2) , 加密算法 Enc , 用公钥 pk_1 加密消息 m 得到的密文 $ct = \text{Enc}(pk_1, m)$ 记为 $\text{Enc}_{pk_1}(m)$ 。

对密文 ct 的自举分为以下 3 步。

(1) 用公钥 pk_2 加密私钥 sk_1 , 生成自举密钥 $\boxed{sk_1}$: $\text{Enc}_{pk_2}(sk_1) \rightarrow \boxed{sk_1}$

(2) 用公钥 pk_2 加密密文 ct (非必须步骤): $\text{Enc}_{pk_2}(ct) \rightarrow \boxed{ct}$

(3) 使用自举密钥同态解密新密文: $\text{Dec}_{\boxed{sk_1}}(\boxed{ct})$ 。

通过这种方式, 就可以在公钥 pk_2 下获得同一消息 m 的加密 $ct_{\text{new}} = \text{Enc}_{pk_2}(m)$ 。 $\boxed{\cdot}$ 表示对方框中内容的加密。

自举的目的是降低密文的误差噪声。从概念上讲, 自举调用解密函数并同时执行重加密, 这些操作生成了新密文, 这个新密文的误差包含了重加密的误差加上解密电路同态计算导致的误差。因此, 自举得到的新密文误差大于用加密算法得到的初始密文, 但小于用比解密电路深度更高的同态求值函数得到的密文误差。自举可以应用于任何密文, 因此可以在多次同态求值后使用, 以降低噪声。这使得具有任意大深度的函数的同态求值成为可能, 但解密电路必须是可同态运算的。2011 年, Gentry 和 Halevi^[20]尝试用类同态加密和乘法同态加密的混合体构造全同态加密方案, 用 Decision Diffie-Hellman 替换 SSSP 假设, 这个方案仍然依赖于自举。Gentry^[21]的研究结果证明, 构造一个全同态方案的问题足以转移到构造一个只能对特定的小次函数集求值的方案, 即类同态方案, 且当该方案用公钥加密的私钥密文公开时仍具有安全性(满足循环安全假设)。如果是这种情况, 那么可以应用自举(必要时还可以压缩)来获得全同态方案。值得强调的是, 自举是获得全同态加密方案的唯一已知方法。不幸的是, 自举在计算上很复杂, 同时需要很大的内存空间。

3.1.3 研究现状

Gentry 的工作做出了许多理论贡献, 但方案中包含的一些数学概念过于复杂难以实现, 而且计算成本过高, 单个比特的自举就需要 30 分钟^[22], 虽然这个全同态方案巨大的开销制约了其实用性, 但是自举思想的提出让类同态加密构造全同态加密提供了实现途径, 后续全同态加密方案的设计都据此思想先构造一个高效的类同态加密方案, 再研究适配此方案的高效自举算法, 最后令类同态加密方案调用自举算法实现全同态加密。此时 Gentry 方案的性能表明投入实际应用还有一段距离, 简化方案结构、降低成本、提升效率都是后续全同态加密方案设计需要解决的问题。

^① 通常自举算法只需自举密钥与密文做乘法, 即可同时完成密文的解密与重加密。

3.2 密钥切换和模切换

以 BFV/BGV 为代表的这类全同态加密方案主要是基于错误学习问题（LWE）和环上的错误学习问题（RLWE）构造的。LWE 问题是 2005 年由 Regev 提出的，具体是指给定一个矩阵 $\mathbf{A} \in \mathbb{Z}_q^{m \times n}$ 和向量 $\mathbf{b} = \mathbf{A} \cdot \mathbf{s} + \mathbf{e} \pmod{q} \in \mathbb{Z}_q^n$ ，求解向量 $\mathbf{s} \in \mathbb{Z}_q^n$ ，已知解决 LWE 问题不比解决格上的 CVP 和 SVP 简单^[23]。RLWE 问题则是 LWE 问题的代数变体，其安全性可以归约到理想格上最坏情况问题，在 2011 年被 Brakerski 和 Vaikuntanathan^[24]提出。同一时期人们发现，基于 RLWE 构造全同态加密方案^[25]似乎比基于 LWE 更有优势。在 BGV 同态加密方案中，LWE 型密文的明文空间为 $\mathbb{Z}_t$ ，RLWE 型密文的明文空间为 $R_t = \mathbb{Z}_t[X]/X^n + 1$ ，但一个 LWE 密文仅能存储一个明文，而一个 RLWE 密文最多能存储 n 个明文，空间利用率大幅度提高。进一步，在引入了单指令多数据（Single Instruction Multiple Data, SIMD）操作^[26]之后，可以将 RLWE 密文对应明文的 n 位全都有效使用，因此后续以 BGV 为基础的全同态加密方案的设计大多基于 RLWE 问题。

3.2.1 具体方案

接下来以 BGV 为例，给出这类全同态加密方案的一般结构。

密钥生成： 给定安全参数 λ ， $\text{KeyGey}(1^\lambda) \rightarrow pk, sk$ ，生成公钥 pk ，私钥 $sk = (l, s)$ 。

加密： $\text{Enc}(m, pk) = ct = (c_0, c_1) \in R_q^2$

解密： $\text{Dec}(ct, sk) = ((c_0 + c_1 \cdot s \bmod q) \bmod t) = m \in R_t$

密文加法： $ct_1 + ct_2 = (c_{10}, c_{11}) + (c_{20}, c_{21}) = (c_{10} + c_{20}, c_{11} + c_{21}) = ct_{\text{add}}$

密文乘法：

$$ct_1 \cdot ct_2 = (c_{10}, c_{11}) \cdot (c_{20}, c_{21}) = (c_{10}c_{20}, c_{10}c_{21} + c_{11}c_{20}, c_{11}c_{21}) = (c_0c_1c_2) = \overline{ct}$$

$$\overline{ct} = (c_0, c_1, c_2) \xrightarrow{\text{keyswitch}} (c_{0'}, c_{1'}) = ct', \quad ct' \bmod q' \xrightarrow{\text{modswitch}} ct_{\text{mult}} \bmod q$$

2011 年，Brakerski 和 Vaikuntanathan^[27]基于 LWE 问题设计了一个不使用自举技术但能增加计算轮次的类同态加密方案（BV），它用重线性化技术来控制密文的维数膨胀，用模切换技术来控制噪声增长，将密文的尺寸缩小到 $O(\lambda \log_2 \lambda)$ 。但由于 BV 方案是基于 LWE 问题构造的，空间利用率并不高，因此在 BV 方案的基础上，2012 年就出现了基于 RLWE 的优化 BV 同态加密方案^[28-29]，分别从不同角度改进拓展了 BV 的核心技术，使得 BV 类同态加密方案性能得到进一步提升。

2012 年, Junfeng Fan 和 Vercauteren^[28]将基于 LWE 问题的 Brakerski 全同态方案移植到 RLWE 中构造了 BFV 方案, 提出了两个优化版本的重线性化算法, 用来解决乘法后密文维数增加的问题。与 BV 方案相比, BFV 的算法缩小了重线性化密钥, 令计算速度得到提升。第一种重线性化方法是选一组合适的基 T , 将重线性密钥 rlk 在这组基上进行分解, 来减少重线性化过程中由于密文乘法增加的噪声。

$$rlk \rightarrow \sum_i^l rlk[i], \quad \bar{ct} = (c_0, c_1, c_2) \rightarrow \left(c_0, c_1, \sum_i^l T^i c_2[i] \right),$$

$$ct_{\text{mult}} = (c_0, c_1) + c_2 \cdot rlk = \left(\left[c_0 + \sum_i^l rlk[i][0] \cdot c_2[i] \right]_q, \left[c_1 + \sum_i^l rlk[i][1] \cdot c_2[i] \right]_q \right)$$

另一种重线性化方法则用到了模切换的思想, 先将密文切换到大模数, 做完重线性化之后再切换回小模数, 利用模约减缩小重线性化乘法中引入的密文噪声。

$$rlk \bmod pq, \quad \bar{ct} \bmod q \rightarrow \bar{ct} \bmod pq$$

$$ct_{\text{mult}} \bmod q = (c_0, c_1) + [c_2 \cdot rlk]_p = \left(c_0 + \frac{c_2 \cdot rlk[0]}{p}, c_1 + \frac{c_2 \cdot rlk[1]}{p} \right) \bmod q$$

BFV 方案中从有限同态方案转变为全同态方案的方法继承了 Gentry 的自举思想, 即在密文噪声达到最大噪声界前降低噪声, 通过简化解密电路, 令解密算法可以在数据加密的情况下同态运行, 这样得到的新密文噪声只与同态解密电路相关, 与初始噪声无关。

BGV 方案是 Brakerski、Gentry 和 Vaikuntanathan^[29]在 2012 年提出的一种基于 RLWE 无自举的全同态方案, 主要贡献是优化了密钥切换技术 (keyswitch) 和其核心思想的模切换技术 (modswitch) 来控制密文的噪声增长速度。

$$\text{keyswitch} : \bar{ct} = (c_0, c_1, c_2) \rightarrow (c'_0, c'_1) = ct'$$

$$\text{modswitch} : ct' \bmod q^l \rightarrow \frac{ct'}{q} = ct_{\text{mult}} \bmod q^{l-1}$$

密文的每次同态乘法都会造成密文维度的增加, 此时就需要密钥切换技术来降低维数确保密文能正确解密, 同时乘法后的密文噪声会有所增加, 用模切换技术将密文模数缩小可以同步地缩小噪声, 在后续计算中减小噪声的增长速度, 增加密文的运算次数。

基于同一基础假设构造的这些全同态方案, 在算法上有一些相似性, 但同时有它们各自的特点。当 BGV、BFV 的明文空间都是 R_t , 明文模 t , 密文模 q 时, BGV 会将消息放在低位

加密 $\text{Dec}(ct_{\text{BGV}}) = m + te$, 而 BFV 会把消息放在高位加密 $\text{Dec}(ct_{\text{BFV}}) = \frac{q}{tm} + e$, 这样的结果是在

t 较大时 BGV 噪声的增长速度会比 BFV 小。对于较大的明文模数, BGV 方案的噪声增加速度较小, 计算效率较高, 而在较小明文模数的情况下, BFV 的运算效率较高, 表现更好。

此时 BGV 方案一次只能对单个密文进行运算, 运算效率低下, 为了降低 BGV 方案的开销、提升计算效率, 人们开始思考如何同时处理多个密文。2010 年, Smart 和 Vercauteren^[26]在 BGV 方案中引入了 SIMD 操作, 通过在多项式环上使用中国剩余定理, 令密文呈现多槽 (Slots) 结构, 允许把明文数组插进槽中, 这样对两个密文进行同态操作时, 可以对数组向量按分量进行同态操作, 提高了空间利用率和计算效率, 降低了密文转换率, 使得基于 RLWE 的同态加密方案相较于 LWE 方案呈现了巨大的优势。2012 年, Gentry 团队^[30-31]针对 SIMD 算法的特点优化了计算电路, 将输入消息打包到 l 个数组中, 每层使用 SIMD 操作, 再将输出作为到下一层的输入, 对于浅层算数电路, 可以实现 polylog 的低开销; 对于安全参数 λ 、任意 t 个门、深度 L 、平均宽度 $\Omega(\lambda)$ 的电路, 可以在时间 $t \cdot \text{polylog}(\lambda)$ 内同态计算。虽然 BGV 方案中模数会随着同态计算逐渐变小, 但上述算法对深度的依赖可以通过使用自举操作来规避。若模数小到无法进行进一步的计算时, 则自举刷新密文, 用模数较大的新密文来支持后续计算。

3.2.2 自举算法发展

BFV/BGV 类同态加密方案从有限同态转变为全同态的方法使用了 Gentry 的自举思想, 即在密文噪声达到最大噪声界前刷新密文降低噪声。先对密文用一个同态计算公钥进行二次加密, 通过简化解密电路, 令解密算法可以同态运行, 将“双重加密”的密文进行同态解密, 这样得到的新密文噪声只与同态解密电路相关, 与初始噪声无关。转化成全同态加密方案的自举流程如下。

- (1) 模切换: 将密文切换到大模数上。
- (2) 内积操作: 密文与同态计算公钥做内积, 对密文进行重加密的同时, 开始进行密文的内层解密。
- (3) 线性变换: 将明文的系数转换到明文槽中, 便于下一步对明文系数进行操作。
- (4) 比特提取: 提取明文系数的最高和最低比特位并相加, 完成重加密密文的内层解密。
- (5) 逆线性变换: 第三步的逆过程, 得到一个噪声减小的新密文。

模切换操作的目的是便于后续计算的顺利进行, 所以将密文切换到更大的密文模数可以容纳更多噪声。比特提取的作用是提取系数的主比特完成同态解密的模 2 运算, 同时去掉之前同态运算引入的噪声, 自举过程中复杂度最高的部分就是比特提取。2016 年, IBM 基于 BGV 开发了一个开源的全同态运算库 HELib^[36-38], Halevi 和 Shoup 详细描述了如何在 HELib

中实现 BGV 的自举算法。在对消息进行稀疏打包的情况下，平均每比特自举仅需 1.3ms，具有实际应用价值。

3.2.3 研究现状

根据 BFV 和 BGV 的这些特点，Andrey Kim 团队^[35]在 2021 年对这两个经典全同态加密方案做出了进一步优化（表 1）。有趣的是，这两个经典方案的改进方法大都源于对方技术特点的提炼转化。经过上述优化之后，明文模较小时（如取 $t=2$ ）改进 BFV 的速度占优，但随着明文模的增大，改进 BGV 的速度更快：对于所有明文模量，改进 BFV 方案运算速度均有提升，当明文模取 $t=2$ 电路深度为 20 时，需要运算时间仅需 50s，相比原始方案的 150s 快了 3 倍，这归功于在 BFV 加密过程中将部分内容进行了预计算；当进行 7 层乘法深度的二叉树计算时，改进 BGV 仅需 1.27s，快于改进的 BFV 1.54s，这两个结果都优于原始的 BFV 2.48s。

表 1 Andrey Kim 团队对 BFV、BGV 方案的算法优化

BFV	BGV	优化效果
$\left\lfloor \frac{q}{t} \right\rfloor m \rightarrow \left\lfloor \frac{qm}{t} \right\rfloor$ 预计算 qm	—	降低初始噪声
将乘法中的模切换步骤置于重线性化之后	—	降低模切换维数
将 BFV 的模数分层，乘法之后切换到小模数	—	减缓噪声增长速度
—	静态噪声管理	增强方案的实用性

3.3 密文矩阵和非对称噪声

2013 年，Gentry、Sahai 和 Waters^[42]提出的以近似特征向量为核心思想，代数运算为同态运算的 GSW 方案标志着全同态方案引入了新的密文结构——矩阵，因为矩阵的特征值与特征向量具有线性关系 $Ax = \lambda x$ ，在原本体系中引入随机噪声，就可以得到一个 LWE 型的加密方案 $C \cdot s = \mu s + e$ ，其中 C 为密文， s 为密钥， e 为随机噪声。GSW 方案的安全性可以归约到与 BFV/BGV 方案相同的 LWE 问题和 RLWE 问题，不同的是，以 BGV、BFV 为代表的加密方案的密文和密钥都是以向量形式表示的，需要额外的密钥切换技术来控制乘法运算后密文的维数扩展，而以 GSW 为代表加密方案的密文是以矩阵形式表示，密文的同态计算可以直接用矩阵加法和乘法运算，避免了密钥交换技术需要用到的大量密钥交换的矩阵而导致公钥长度的增长。

3.3.1 具体方案

GSW 方案继承了 BGV 方案的许多优点，并在此基础上加入了自己的核心思想——近似特征向量技术，这是一个无须同态计算公钥 evk 就可以进行同态运算的“层次型”同态加密方案。GSW 类方案的特点是同态密文运算为简单的矩阵运算，矩阵运算支持加法乘法同态运算，相比于向量结构的密文，矩阵结构虽然存在不支持密文打包的局限性，但避免了维数膨胀问题和模切换运算。此外，根据 GSW 方案的特点还可以构造基于身份的同态加密方案和基于属性的同态加密方案。GSW 类同态加密方案的一般结构如下。

密钥生成：随机抽取一个向量 $\bar{s} \in \mathbb{Z}_q^{n-1}$ ，生成私钥 $s \rightarrow (\bar{s}, -1)^T \in \mathbb{Z}_q^n$ 。

加密：给定消息 u 和 n 阶单位矩阵 I_n ，随机选取矩阵 $A \in \mathbb{Z}_q^{n \times (n-1)}$ ，噪声向量 $e \rightarrow \chi_e^n$ ，生成密文矩阵 $C = (A, A \cdot \bar{s} + e) + u \cdot I_n \in \mathbb{Z}_q^{n \times n}$ 。

解密： $\text{Dec}(C, s) = C \cdot s = u \cdot s - e$ ，这里噪声向量 e 的正负都表示引入的 LWE 噪声，后文统一取正号。

密文加法：矩阵加法 $C_1 + C_2$ ， $\text{Dec}(C_1 + C_2) = (u_1 + u_2) \cdot s + (e_1 + e_2)$ 。

密文乘法：矩阵乘法 $C_1 C_2$ ， $\text{Dec}(C_1 \cdot C_2) = (u_1 \cdot u_2) \cdot s + u_2 \cdot e_1 + C_1 \cdot e_2$ 。

可以观察到密文乘法中噪声并不是对称增长的，也就是 $C_1 + C_2$ 与 $C_2 + C_1$ 引起的噪声增长并不相同。因为密文矩阵 C 远大于消息 u ，所以 $C \cdot e$ 在乘法噪声中占主导地位。因为 GSW 密文正确解密的噪声上界为 $q/4$ ，所以为了控制乘法中产生的噪声项 $C \cdot e$ ，就需要限制密文 C 的大小，GSW 方案中使用二进制分解的思想定义了如下函数。

$$\text{BitDecomp}(a) = (a_{1,1}, \dots, a_{1,m}, a_{2,1}, \dots, a_{n,1}, \dots, a_{n,m})$$

为向量 $a = (a_1, \dots, a_n)$ 从最低位到最高位每项系数的比特分解。

$$\text{BitDecomp}^{-1}(b \in \mathbb{Z}_q^{nm}) = \left(\sum_j^{m-1} 2^j b_{1,j} \bmod q, \dots, \sum_j^{m-1} 2^j b_{n,j} \bmod q \right)$$

为 $\text{BitDecomp}(\cdot)$ 的逆函数。

$$\text{Flatten}(b \in \mathbb{Z}_q^{nm}) = \text{BitDecomp}(\text{Bitcomp}^{-1}(b))$$

生成一个系数为 $\{0,1\}$ 的向量。

$$\text{Powersof } 2(s) = (s_1, 2s_1, \dots, 2^{m-1}s_1, s_2, 2s_2, \dots, 2^{m-1}s_n) \bmod q$$

对于任意 $a, s \in \mathbb{Z}_q^n$ ，满足

$$\langle a, s \rangle = \langle \text{BitDecomp}(a), \text{Powersof } 2(s) \rangle$$

对于任意 $b \in \mathbb{Z}_q^{mn}, s \in \mathbb{Z}_q^n$ ，满足

$$\langle b, \text{Powersof } 2(s) \rangle = \langle \text{BitDecomp}^{-1}(b), s \rangle = \langle \text{Flatten}(b), \text{Powersof } 2(s) \rangle$$

使用上述函数, 对密文矩阵 C 进行二进制分解, 即可实现降低乘法项的密文尺寸, 控制噪声增长的目的。取 $m = n \cdot \log_2 q$, 根据上述函数定义一个 $m \times n$ 维的二进制重组矩阵:

$$G = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 2 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 2^{\log_2 q - 1} & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 2^{\log_2 q - 1} \end{pmatrix}$$

在加密时, 生成密文 $C = (A, A \cdot \bar{s} + e) + u \cdot G$, 然后不是直接输出 C , 而是输出密文分解后的二进制矩阵 $\hat{C} = G^{-1}(C)$ 。密文 C_1 与 C_2 进行乘法运算之后 $\text{Dec}(C_1 \cdot C_2) = (u_1 \cdot u_2) \cdot G \cdot s + u_2 \cdot e_1 + C_1 \cdot e_2$, 这里的 C_1 为二进制矩阵, 当噪声界取 δ 时, $C_1 \cdot e_2$ 的上界为 $n \cdot \log_2 q \cdot \delta$, 远小于 $C_1 \cdot e_2$ 的上界 $m \cdot q \cdot \delta$ 。

除此以外, 观察到 G 的第一行 $G_1 = (1, 0, \dots, 0)$ 只有第一项是 1, 其余为 0。在解密阶段, $\text{Dec}(\hat{C}, s) = \hat{C} \cdot G \cdot s = u \cdot G \cdot s + e$, 则 $u \cdot G \cdot s$ 的第一行有 $(u \cdot G \cdot s)_1 = u \cdot G_1 \cdot s = u \cdot s[1]$, 因为 $s[1] \in \mathbb{Z}_q$ 大概率是个远大于噪声上限的数字, 所以只需计算观察结果向量的第一项, 就可以得到消息 u 。

3.3.2 自举算法发展

目前 GSW 类方案主要为单比特的加密, 因此自举过程不涉及密文的打包与解包, 自举过程即为同态解密的过程, 主要涉及以下两个步骤。

(1) 同态累加: 构造同态累加器完成同态解密过程, 包括内积操作和模约减。

(2) 比特提取: 进行同态的四舍五入操作, 得到自举后的新密文。

对 GSW 类加密方案来说, 调整方案构造, 使用比较短的参数就可以满足安全性, 这样便于引入一些非常高效的自举技术, 因此后续的一系列研究都专注于提升方案的自举性能。Jacob Alperin-Sheriff 和 Chris Peikert^[43]在 2014 年利用对称群和置换矩阵的思想, 设计了一个 GSW 方案的单比特快速自举算法, 为了完全重新随机化误差, 在构造格陷门时把使用的矩阵 $G^{-1} = \text{BitDecomp}(I_N)$ 从一个确定性的比特分解, 重新定义为一个随机抽样的矩阵。此外, 他们还使用中国剩余定理将大的加法群嵌入小的对称群, 这样自举需要的同态操作次数就从

$\tilde{O}(\lambda^3)$ 降到 $\tilde{O}(\lambda)$ 次。2015 年, Ducas 和 Micciancio^[44] 将 AP14 方案的思想扩展到环上, 用模 4 操作来同态计算两个密文的 NAND 与非门, 构造了一个基于环 GSW 的同态累加器来执行密文的自举过程, 将自举速度提升到 1 秒/次, 并得到自举一个比特只需要 0.69 秒/次的优秀结果。Ilaria Chillotti 团队^[45] 在 2016 年对 GSW 密文与 LWE 密文、GSW 密文与 GSW 密文的乘法做出了进一步的优化, 设计了一种新的一般化乘积机制, 并在此基础上改进了自举步骤, 将单个门自举时间减少到 0.1s。

3.3.3 研究现状

GSW 类的全同态加密方案以其优秀的自举速度被认为是目前最理想的全同态加密方案, 目前主流实现的 GSW 类加密方案仍使用单比特加密自举, 如何设计具有实用价值的密文打包方法和并行计算算法, 进一步提升可实现的自举算法效率, 是近年研究的重点。GSW 类不同于 BGV/BFV 类加密方案的矩阵密文结构, 无法自然地令一个密文向量打包加密多条消息或多个密文并行运算。2018 年, Steven Myers 和 Adam Shull^[51] 提出了一种打包密文的自举算法, 通过在 FHEW 的基础上使用递归的计算方法, 将计算复杂度降低到 $O(3^\rho \cdot \lambda^{1/\rho})$, $\rho = O(1)$ 。2023 年 Feng-Hao Liu 和 Han Wang^[52-53] 在 MS18 方案打包算法的基础上, 进一步引入了适用 FHEW 的 SIMD 算法, 将计算复杂度直接优化到 $\tilde{O}(1)$ 。虽然上述方案对提升自举效率具有很大的理论价值, 但是素数幂的明文环的设置, 在盲旋转算法中 test 向量的构造等实例化参数的选择是该方案目前的实现瓶颈。

3.4 非精确解密场景的适应性方案

为了配合一些不需要精确解密的应用场景, Jung Hee Cheon 团队^[32] 在 2017 年提出了一种近似同态加密方案——CKKS, 该方案脱胎于 BGV, 它有能支持浮点数消息加密的特性, 同时采用了正则嵌入的消息编码方式。

3.4.1 具体方案

CKKS 主要思想是将加密噪声视为消息的一部分, 与 BFV、BGV 相比, CKKS 的特点有两个: 一是对引入的噪声与明文加密时并未进行高低比特位上的分离, 即 $\langle ct, sk \rangle = m + e$; 二是解密后的明文是原始明文的近似值, 即 $\text{Dec}(\text{Enc}(m)) \approx m$ 。这些特点让 CKKS 方案的使用更加灵活, 也出现了许多改进优化版本适应更多领域的应用。

密钥生成: 给定安全参数 λ , $\text{KeyGen}(1^\lambda) \rightarrow pk, sk$, 抽取噪声 $e, e' \rightarrow \mathcal{X}$, 生成公钥 $pk =$

$(b, a) = (-as + e, a)$, 私钥 $sk = (1, s)$ 。

加密: 给定明文 m , 生成密文 $\text{Enc}(m, pk) = ct = (c_0, c_1) \in R_q^2$, 其中 $(c_0, c_1) = (-as + e + m, a)$ 。

解密: $\text{Dec}(ct, sk) = ((c_0 + c_1 \cdot s \bmod q) \bmod t) \approx m \in R_t$ 。

CKKS 的密文加法和乘法与 BGV 类方案计算过程基本相同, 故不再赘述。CKKS 需要额外关注的是它的编码过程, 因为其编码消息是浮点数, 不同于其他全同态方案的整数消息, CKKS 用正则嵌入矩阵将浮点数消息编码成整数明文。给定精度因子 Δ , 正则嵌入矩阵 σ , 函数 π , 消息 $z = (a_1 + ib_1, \dots, a_{n/2} + ib_{n/2}) \in C^{n/2}$, 明文 $m \in R_t^n$, ζ 为多项式环 $\Phi_M = X^n + 1$ 的本源单位根, 满足 $\zeta^n = 1 \bmod M$ 。

编码: $m = \text{Encode}(z, \Delta) = \sigma^{-1} \lceil \Delta \pi^{-1}(z) \rceil$

解码: $z = \text{Decode}(m, \Delta) = \pi \circ \sigma(m / \Delta)$

其中, 函数 π 的作用是将复数向量取共轭再连接到原向量后面, 即 $\pi^{-1}(z) = (a_1 + ib_1, \dots, a_{n/2} + ib_{n/2}, a_1 - ib_1, a_{n/2} - ib_{n/2})$ 。CKKS 方案加密的是复数域上消息向量, 所以需要先通过正则嵌入打包成环上的明文向量, 这个过程的实质是快速数论变换 (Number Theory Transform, NTT) 操作。下面给出了正则嵌入矩阵的示例:

$$\sigma = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \zeta^1 & (\zeta^1)^2 & \dots & (\zeta^1)^{n-1} \\ 1 & \zeta^3 & (\zeta^3)^2 & \dots & (\zeta^3)^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \zeta^{2n-1} & (\zeta^{2n-1})^2 & \dots & (\zeta^{2n-1})^{n-1} \end{pmatrix}$$

3.4.2 自举算法发展

因为 CKKS 方案主要针对于无须精确解密场景的应用, 所以自举过程中的比特提取刷新密文的步骤可以用多项式近似计算模函数直接将密文从小模数切换到大模数实现密文的刷新, 无须进行同态的精确解密。CKKS 方案的自举主要有以下几个步骤。

- (1) 升模: 将密文从小模数直接提升到大模数。
- (2) 线性变换: 将明文系数转换到明文槽中便于下一步对明文系数进行操作。
- (3) 同态模约减: 为了执行完整的模切换, 需要明文系数在升模的模数下进行模约减。
- (4) 逆线性变换: 第二步过程的逆变换, 得到一个大模数下的新密文。

CKKS 用正则嵌入的方式打包密文, 将消息存储在明文槽中, 因为模切换是对明文系数进行操作, 这种操作不会改变消息, 所以它在自举过程中不能直接进行模切换, 而是需要先进行线性变换, 将系数向量放入明文槽中, 再通过一个从小模数切换到大模数的换模操作来

完成全部自举过程，实现减小密文噪声扩大密文模数的目标。

CKKS 类全同态加密方案自举的不同近似函数及自举精度对比如表 2 所示。

表 2 CKKS 类全同态加密方案自举的不同近似函数及自举精度对比

方案名称	近似函数（方法）	自举精度（bit）
CHKKS ^[39]	$\sin\left(\frac{2\pi x}{q}\right) \approx x \bmod q$	24
LLLK ^[40]	$\left(\frac{2\pi x}{q}\right) \arcsin\left(\sin\left(\frac{2\pi x}{q}\right)\right) = x \bmod q$	40.5
JM ^[41]	$\frac{q}{2\pi} \sum_{k=1}^n \beta_k \sin\left(\frac{2\pi k x}{q}\right) \approx x \bmod q$	44
BCCKK ^[42]	META BTS	255

CKKS 的换模操作需要用多项式来近似模约减函数，这个近似过程中产生的近似误差是自举误差的主要来源。目前主流的思想是选择一个在原点附近与模约减函数形状相近的函数作为模约减函数和近似多项式之间的桥梁，如 $\sin$ 函数^[39]。但是近似函数与模约减函数之间同样会产生近似误差，因此需要找到更接近模约减函数的近似函数。2021 年 Joon-Woo Lee 团队^[40]用复合函数的方法在原 $\sin$ 函数近似的基础上添加了一个逆 $\sin$ 函数，去掉了 $\sin$ 函数与模约减函数之间的近似误差，但是需要额外计算一个逆 $\sin$ 函数的多项式近似。总体而言，此方法提高了自举精度，特别是在小明文模数的情况下具有优异表现，因为在小明文模数情况下近似函数与模约减函数之间的近似误差占主导地位，此时自举精度可以达到 40.5bit，计算时间仅需 94.7s。2022 年，Jutla 和 Manohar^[41]使用了一个新的近似函数——正弦级数近似，可以实现几乎任意精度的 CKKS 自举，使用 n 阶正弦级数近似时误差约为 $O(\epsilon^{2n+1})$ ，优于正弦函数近似误差 $O(\epsilon^3)$ ，自举精度也达到 44bit。CKKS 自举过程中由于不可避免的模切换操作引入了误差导致自举精度始终受限，2022 年，Youngjin Bae 团队^[54]针对该问题提出了一种新的自举模式——META BTS，是一个类似比特提取的算法，通过重复两次给定的任意自举算法，可以获得一个精度加倍的自举结果，即迭代低精度自举来获得高精度自举。在目标精度 n 足够大的情况下， n 位精度自举的密文模数近似为 $O(n^{5/4})$ 。该方案的优势在于将自举算法作为黑盒，可以用于 CKKS 类的所有自举算法，因此如果未来出现了更先进的自举算法，那么同样可以令 META BTS 在精度方面得到更好的结果。

3.4.3 研究现状

理想状态下只要噪声 e 与消息 m 相比足够小，则该噪声不会破坏消息 m 的有效数字，但人们一直对 CKKS 的安全性有所质疑。Baiyu Li 和 Daniele Micciancio^[33]在 2021 年发表的一

篇论文中提出，虽然 CKKS 方案满足传统的 IND-CPA 安全，但是对于更强的带有特殊解密 Oracle 敌手——IND-CPA^D 安全性，可以用恢复密钥攻击打破。这一结论使得人们的目光又聚焦到 CKKS 方案的安全性改进上，2022 年 Baiyu Li^[34] 团队利用 noise flood 技术，通过给解密后的明文添加噪声的方法使 CKKS 方案满足 IND-CPA^D 安全性，但这一方法同时会令密文损失将近一半的精度。不论是自举算法中的精度改进，还是近似解密相关安全性研究，CKKS 方案都与噪声联系密切，噪声控制分析也将是 CKKS 类同态方案未来研究的重点。

3.5 对比分析

表 3 列出了 4 条不同技术路线全同态加密方案的一些特点对比，表中的 s 为稀疏子集的大小， S 为密钥长度， d 为格的行列式， N 为理想格维数， $f(\lambda)$ 为安全参数 λ 的函数， n 为环维数， q 为密文模， t 为明文模。首先，因为 Gentry 类方案基础假设中的理想格参数设置难度较大，而 BGV 类、GSW 类、CKKS 类方案都是基于 (R)LWE 困难问题的，(R)LWE 假设相较于理想格假设更加简单易用，所以这三条技术路线下的全同态加密方案都具有实用性且被广泛应用，并且这些方案都由多个算法库支持代码实现，但 Gentry 类方案由于实现存在困难逐渐退出人们的研究视野。接下来将主要对基于 (R)LWE 的方案进行对比分析。

表 3 Gentry 类、BGV 类、GSW 类、CKKS 类全同态加密方案对比

方案类别	Gentry 类	BGV 类		GSW 类	CKKS 类
基础假设	理想格	LWE	RLWE	(R)LWE	RLWE
密文空间	$\mathbb{Z}_d$	$\mathbb{Z}_q^n$	R_q^n	$\mathbb{Z}_q^{n \times n}$	R_q^n
密文大小	$f(\lambda)$	$(n+1)\lceil \log_2 q \rceil$	$2n\lceil \log_2 q \rceil$	$(n+1)^2\lceil \log_2 q \rceil^3$	$2n\lceil \log_2 q \rceil$
明文空间	$\mathbb{Z}_2$	$\mathbb{Z}_t$	R_t	$\mathbb{Z}_2$	—
单个密文存储明文数	1	1	n	1	n
公钥大小	$s \cdot \lceil 2\sqrt{S} \rceil \cdot \log_2 d$	$2n(n+1)\log_2 q$	$2n\lceil \log_2 q \rceil$	$2n(n+1)\log_2 q$	$2n\lceil \log_2 q \rceil$
是否支持消息打包	—	否	是	暂不支持	是
是否支持 SIMD	—	是	是	暂不支持	是
同态乘法噪声增长速度	$N^{O(1)}$	$O(n^2)$	$O(n^2)$	$O(n)$	$O(n^2)$
(平均) 单比特自举时间	30min	—	1.3ms	13ms	6.43s

BGV 类方案基于 LWE 和 RLWE 不同困难问题的效率有所不同, 得益于 RLWE 的环结构, 可以令单个密文加密多个消息, 提高了密文利用率, 同时降低了公钥的尺寸, 这样打包密文的平均单比特的自举时间可以实现约 1ms, 与以自举速度著称的 GSW 类方案同样做到毫秒级。虽然文中提到的一些最新成果给了 GSW 类方案的 SIMD 和打包操作理论支撑, 但目前因为自举效率和技术局限性, 在代码实现上仍不支持上述两种操作。

表中数据显示, GSW 类方案的密文要大于其他方案, 是因为密文呈矩阵结构, 这样的好处是在进行密文同态乘法时做的是矩阵乘法, 不会有向量乘法的维数扩张问题, 而且噪声增长的速度也会因噪声向量被扩张成二进制矩阵而降低, 由指数级降低为线性级。GSW 类方案的经典方案构造基于 LWE 假设, 仅在自举步骤用到 RLWE 的环结构, 并不涉及密文的打包, 所以表中数据除自举时间之外都是基于 LWE 的方案参数。

CKKS 因为近似加密的特点会把加密引入的 LWE 噪声看作消息的一部分, 解密的精度由精度因子 Δ 决定, 所以只有密文空间, 没有其他全同态加密方案所需的明文空间。

在方案效率方面, Gentry 类方案^[22]代码实现的效果并不理想, 在 72bit 的安全强度下, 公钥大小约为 2.3GB, 自举刷新但比特密文需要 30min。相较于基于(R)LWE 的全同态加密方案中公钥最小仅需几十兆字节, 自举速度已经提升到毫秒级, Gentry 类方案的自举速度和内存需求明显缺乏竞争力。BGV 类方案在算法库 HELib^[37]的实现中设定安全参数 128bit, 稀疏自举^①可以做到平均单比特自举仅用时 1.3ms。GSW 类方案在 TFHE 库中实现取安全参数为 128bit 可以做到 13ms 的单比特自举。CKKS 类方案的自举时间与精度呈正相关, 所需精度越高, 自举过程中用来近似的多项式维数越高, 自举时间越长。目前 CKKS 类方案的实现效果^[54]聚焦的重点是自举后的密文精度, 以及可以支持的乘法深度。在算法库 HEAAN 中选择 128bit 安全参数, 精度因子 $\Delta = 2^{36}$, 当取环维度 2^{15} 时, 自举所需时间为 6.43s, 自举后支持的最大乘法深度为 14; 当取环维度 2^{16} 时, 自举所需时间为 16.3s, 自举后支持的最大乘法深度为 29。

4 应用实现

随着数据安全的普及, 各类云服务的广泛应用, 全同态加密算法拥有广阔的应用场景和发展空间, 目前已经有一些团队根据上述全同态方案构造了相应的算法库, 将理论成果转化为高质量实现。

HELib^[46]由 IBM 开发设计, 是一个用 C++ 语言编写的同态加密开源软件库, 可以在 Windows、macOS、Ubuntu、CentOS 等操作系统平台上进行安装部署, 底层依赖于 NTL 数论

^① 每个明文槽都只包含字段（或环）的一个元素，而不是扩展字段（或环）。

运算库和 GMP 多精度运算库，支持 BGV 和 CKKS 方案的实现。此外，HElib 还提供许多提升算法运行效率的代码，如 Smart-Vercauteren 密文打包技术和 Gentry-Halevi-Smart 优化算法，支持“set”“add”“multiply”“shift”等基本操作指令。2018 年 IBM 发布的新版本 HElib 库，优化了重线性化算法，使效率进一步提升了 15~75 倍。

Microsoft SEAL^[47]由 Microsoft 的密码学和隐私研究小组开发，是一个用 C++编写的开源同态加密库，可以在多种环境中运行，支持 BFV、BGV 和 CKKS 3 种同态加密方案。SEAL 库的使用要求用户理解许多同态加密的特定概念，它可以进行密文的同态加法和乘法操作，但无法进行密文的比较排序等操作。

PALISADE 是一个开源项目，提供格密码构建实例和最新同态加密方案的高效实现。PALISADE 专为可用性而设计，支持 BGV、BFV、CKKS 和 FHEW 方案及 TFHE 方案的变体，包括对应的自举算法。PALISADE 还提供后量子公钥加密、代理重加密、多方计算阈值同态、基于身份的加密、基于属性的加密和数字签名等加密方案。目前 PALISADE 的主创团队已将该项目合并到后继者 OpenFHE^[56]库中。

TFHE 库由 Chillotti 团队^[48]开发设计，它实现了 GSW 方案的类同态和带自举的全同态。与其他库不同，TFHE 库对电路门的数量或其组成没有限制，因此它的电路可由用户自定义，或者用带有自动电路生成的工具生成，可以实现对任意电路的密文同态运算。

HEAAN^[55]是基于 CKKS 类方案由 CKKS 作者团队开发的一个开源算法库，支持定点算术和有理数的近似计算，支持自举算法基于 CHKKS18^[39]，引入的近似误差基本等同浮点运算误差，取决于具体参数的设置。HEAAN 库支持 C++和 NTL library 语言，目前更新到 V2.1 版本。

OpenFHE^[56]是 Duality Technologies 团队在 2022 年发布的一个支持目前大部分全同态算法的开源算法库。OpenFHE 库继承了 PALISADE 库的所有功能，并融合了 HElib 和 HEAAN 的特定功能，它支持用户友好模式和编译器友好模式，可以由库或外部编译器直接调用同态密文的运算算法，提供了更简单的 API、模块化、跨平台支持和硬件加速器集成。

除上述主流同态实现库之外，还有主要用于探索使用低级处理器原语的高性能同态加密的 NFLlib 库，探索使用 GPU 来加速同态加密的 cuFHE 库，用 Go 语言编写的 Lattigo 库等。表 4 针对一些基本的使用需求对比了这些实现库的应用特点。

表 4 不同实现库的应用特点对比

实现库名称	实现语言	支持加密方案	自举	特色功能
HElib	C++	BGV、CKKS	√	支持 HE 汇编语言
SEAL	C++	BFV、CKKS	√	跨平台编译

续表

实现库名称	实现语言	支持加密方案	自举	特色功能
PALISADE	C++	BGV、BFV、CKKS、FHEW、TFHE	√	跨平台编译，BGV/BFV/CKKS 的多方扩展
TFHE	C/C++	GSW、FHEW、TFHE	√	可编程自举
HEAAN	C++	CKKS	√	GPU 加速
NFLlib	C++	—	—	基于 NTT 的快速格计算
cuFHE	C++	TFHE	√	CUDA 加速
Lattigo	Go	BGV、BFV、CKKS	√	跨平台编译，剩余数系统（RNS）
OpenFHE	C++	BGV、BFV、CKKS、FHEW、TFHE	√	剩余数系统（RNS），BGV/BFV/CKKS 的多方扩展

5 总结与展望

自从 Gentry 开创了全同态的方案构造，经过十多年的发展，全同态加密技术正在从理论走向实际，从一个理论上的方案转变成一种实际可用的工具。多方安全计算、密码数据库、区块链隐私保护、机器学习隐私等多种应用领域不断涌现同态密码技术的身影，并将在今后的发展中发挥出巨大的作用。

一方面，为了适应云环境下多用户快速、稳定地存储和计算密文，全同态加密的方案设计开始面向多用户：门限全同态^[61]（ThresholdFHE），多密钥的全同态^[62]（MKFHE）、多方参与的全同态^[63]（MPFHE），这些带交互的方案设计进一步拓宽了全同态加密的应用场景。

另一方面，随着计算机算力的提升，数据处理能力的增强，针对全同态方案基础结构的许多实验性改进将更有实际意义：非 2 次幂的明文环次数^[64]， $t = X^m + b$ 的明文模数^[65]，比 SIMD 更复杂、高效的域指令多数据（Field Instruction Multiple Data, FIMD）打包^[66]等，这些技术都具有推广到所有全同态加密方案的潜力。

目前已有的 4 条同态加密方案技术路线各有不同的挑战和机遇。Gentry 基于理想格做出的一系列全同态相关领域研究，开创发展了全同态加密的理论基础，但是理想格等数学概念难以用代码实现，所以在(R)LWE 假设被广泛使用后这条技术路线下的相关研究就渐渐遇冷。BGV 类全同态方案基于简单易用的(R)LWE 假设，便于用计算机编程验证方案的具体性能，此外还擅长多组打包密文的并行处理，因此即使自举时间相对 GSW 类方案较长，单比

特的平均自举时间基本为同一水平。目前 BGV 类方案的优化方向一是加速自举算法,改进自举中最耗时的比特提取步骤,提升密文同态的打包解包速度;二是优化消息打包方式,提升密文空间的使用率。GSW 类全同态加密方案利用矩阵的代数性质加速了密文乘法,特有的非对称噪声增长也将噪声增长速度从 BGV 类的指数级降低到线性级,密文的自举性能相比其他技术路线的同态方案具有极大优势。但 GSW 类加密方案如何实现消息的打包和密文的并行处理仍是亟待解决的难题,目前的理论成果往往需要一些严苛的参数要求,距离应用实现仍有差距。CKKS 类方案的主要结构与 BGV 类相似,因此除 BGV 类方案存在的问题之外,CKKS 类方案选择更好的近似函数提升自举精度,噪声泄露的信息对方案安全性的影响,这些都是未来研究的重点。总体来说,上述全同态加密方案目前仍无法满足实际应用中云端对大量数据的密文快速处理需求,如何进一步加速理论成果转化,改进全同态加密方案实用性能,这将是未来很长一段时间需要解决的难题。

参考文献

- [1] BRICKELL E F. On privacy homomorphism[C]. In: Advances in Cryptology-EUROCRYPT 1987. Springer Berlin Heidelberg, 1988: 117-125.
- [2] CRAMER R, DAMGÅRD I, NIELSEN J. Multiparty computation from threshold homomorphic encryption[C]. In: Advances in Cryptology-EUROCRYPT 2001. Springer Berlin Heidelberg, 2001: 280-300.
- [3] RIVEST R L, ADLEMAN L, and DERTOUZOS M L. On data banks and privacy homomorphisms[J]. Foundations of Secure Computation, 1978, 4(11): 169-180.
- [4] ELGAMAL T. A public key cryptosystem and a signature scheme based on discrete logarithms[C]. In: Advances in Cryptology-CRYPTO 1984. Springer Berlin Heidelberg, 1985: 10-18.
- [5] RIVEST R L, SHAMIR A, ADLEMAN L. A method for obtaining digital signatures and public-key cryptosystems[J]. Communications of the ACM, 1978, 21(2): 120-126.
- [6] BONEH D, GOH E, NISSIM K. Evaluating 2-DNF formulas on ciphertexts[C]. In: Theory of Cryptography-TCC 2005. Springer Berlin Heidelberg, 2005: 325-341.
- [7] GENTRY C. Fully homomorphic encryption using ideal lattices[C]. In: Proceedings of the 41st Annual ACM Symposium on Symposium on Theory of Computing-STOC 2009. ACM, 2009: 16.

[8] BEHERA S, PRATHURI J R. Design of Novel Hardware Architecture for Fully Homomorphic Encryption Algorithms in FPGA for Real-Time Data in Cloud Computing. in IEEE Access, vol. 10, pp. 131406-131418, 2022.

[9] LIJING Z, ZIYU W, HONGRUI C, et al. HEAD: an FHE-based Privacy-preserving Cloud Computing Protocol with Compact Storage and Efficient Computation. Cryptology ePrint Archive, 2022.

[10] CARSTEN B, JAMES H C, BERNARDO D, et al. SoK: Privacy-Enhancing Technologies in Finance. Cryptology ePrint Archive, Paper 2023/122.

[11] LI T, FANG B K, MA Z. Homomorphic encryption-based ciphertext anomaly detection method for e-health records [J]. SCIENTIA SINICA Informationis, 2022.

[12] REGEV O. Lattice-based Cryptography[C]. In: Berlin, Heidelberg: Springer, 2009.

[13] STEHLÉ D, STEINFELD R, TANAKA K, et al. Efficient publickey encryption based on ideal lattices. In: International Conference on the Theory and Application of Cryptology and Information Security, 2009.

[14] SVP challenge, <https://www.latticechallenge.org/svp-challenge/>.

[15] AJTAI M. Generating hard instances of lattice problems (extended abstract). in: Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing. New York, NY, USA: Association for Computing Machinery, 1996.

[16] REGEV O. On lattices, learning with errors, random linear codes, and cryptography. In: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing, 2005.

[17] STEHLE D, STEINFELD R, TANAKA K, et al. Efficient public key encryption based on ideal lattices. In: International Conference on the Theory and Application of Cryptology and Information Security, 2009.

[18] GENTRY C, HALEVI S, VAIKUNTANATHAN V. i-Hop Homomorphic Encryption and Rerandomizable Yao Circuits. In: Advances in Cryptology-CRYPTO 2010, T. Rabin, Ed. Berlin, 2010.

[19] GENTRY C. Computing arbitrary functions of encrypted data. In: Communications of the ACM, vol. 53, no. 3, pp. 97-105, 2010.

[20] GENTRY C, HALEVI S. Fully homomorphic encryption without squashing using depth-3 arithmetic circuits[C]. In: IEEE 52nd Annual Symposium on Foundations of Computer Science—FOCS, 2011. IEEE, 2011.

[21] GENTRY C, HALEVI S, SMART N P. Better bootstrapping in fully homomorphic encryption[C]. In: Public Key Cryptography—PKC. Springer Berlin Heidelberg, 2012: 1-16.

- [22] GENTRY C, HALEVI S. Implementing Gentry's fully-homomorphic encryption scheme[C]. In: Advances in Cryptology-EUROCRYPT 2011. Springer Berlin Heidelberg, 2011: 129-148.
- [23] LYUBASHEVSKY V, PEIKERT C, REGEV O. On ideal lattices and learning with errors over rings. In: Advances in Cryptology-EUROCRYPT 2010, H. Gilbert, Ed. Berlin, 2010.
- [24] BRAKERSKI Z, VAIKUNTANATHAN V. Fully homomorphic encryption from ring-LWE and security for key dependent messages[C]. In: Advances in Cryptology-CRYPTO 2011.
- [25] DUCAS L, DURMUS A. Ring-LWE in Polynomial Rings. In: Public Key Cryptography-PKC 2012. Berlin, Heidelberg: Springer, 2012, pp. 34-51.
- [26] SMART N P, VERCAUTEREN F. Fully homomorphic SIMD operations[J]. Designs, Codes and Cryptography, 2014, 1(71): 57-81.
- [27] BRAKERSKI Z, VAIKUNTANATHAN V. Efficient fully homomorphic encryption from (standard) LWE[C]. In: IEEE 52nd Annual Symposium on Foundations of Computer Science-FOCS 2011. IEEE, 2011: 97-106.
- [28] FAN J, VERCAUTEREN F. Somewhat practical fully homomorphic encryption[J/OL]. In: IACR Cryptology ePrint Archive, 2012: 2012/144.
- [29] BRAKERSKI Z, GENTRY C, VAIKUNTANATHAN V. (Leveled) fully homomorphic encryption without bootstrapping[C]. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. ACM, 2012.
- [30] GENTRY C, HALEVI S, SMART N P. Fully Homomorphic Encryption with Polylog Overhead. In: Advances in Cryptology-EUROCRYPT 2012.
- [31] GENTRY C, HALEVI S, SMART N P. Homomorphic Evaluation of the AES Circuit. In: Advances in Cryptology-CRYPTO 2012. Lecture Notes in Computer Science, vol 7417.
- [32] CHEON J H, KIM A, KIM M, et al. Homomorphic encryption for arithmetic of approximate numbers[C]. In: International Conference on the Theory and Applications of Cryptology and Information Security, HongKong, China, 2017.
- [33] LI B, MICCIANCIO D. On the Security of Homomorphic Encryption on Approximate Numbers. In: Advances in Cryptology-EUROCRYPT 2021. Lecture Notes in Computer Science, vol 12696.
- [34] LI B, MICCIANCIO D, SCHULTZ M, et al. Securing Approximate Homomorphic Encryption Using Differential Privacy. In: Advances in Cryptology-CRYPTO 2022. Lecture Notes in Computer Science.

[35] KIM A, POLYAKOV Y, ZUCCA V. Revisiting Homomorphic Encryption Schemes for Finite Fields. In: *Advances in Cryptology-ASIACRYPT 2021. Lecture Notes in Computer Science*, vol 13092.

[36] HALEVI S, SHOUP V. Algorithms in HELib. In: *Advances in Cryptology-CRYPTO 2014. Lecture Notes in Computer Science*, vol 8616.

[37] HALEVI S, SHOUP V. Bootstrapping for HELib[J]. In: *Cryptol* 34, 7 (2021).

[38] HALEVI S, VICTOR S. Design and implementation of HELib: a homomorphic encryption library. In: *IACR Cryptol. ePrint Arch*, 2020 (2020): 1481.

[39] CHEON J H, HAN K, KIM A, et al. Bootstrapping for Approximate Homomorphic Encryption. In: *Advances in Cryptology-EUROCRYPT 2018*.

[40] LEE J W, LEE E, LEE Y, et al. NO J S. High-Precision Bootstrapping of RNS-CKKS Homomorphic Encryption Using Optimal Minimax Polynomial Approximation and Inverse Sine Function. In: *EUROCRYPT 2021*.

[41] JUTLA C S, MANOHAR N. Sine Series Approximation of the Mod Function for Bootstrapping of Approximate HE. In: *Advances in Cryptology-EUROCRYPT 2022. Lecture Notes in Computer Science*.

[42] GENTRY C, SAHAI A, WATERS B. Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based[C]. In: *Advances in Cryptology-CRYPTO 2013*.

[43] ALPERIN-SHERIFF J, PEIKERT C. Faster bootstrapping with polynomial error[C]. In: *Advances in Cryptology-CRYPTO 2014. Springer Berlin Heidelberg*, 2014: 297-314.

[44] DUCAS L, MICCIANCIO D. FHEW: Bootstrapping homomorphic encryption in less than a second[C]. In: *Advances in Cryptology-EUROCRYPT 2015. Springer Berlin Heidelberg*, 2015: 617-640.

[45] CHILLOTTI I, GAMA N, GEORGIEVA M, et al. Faster Fully Homomorphic Encryption: Bootstrapping in Less Than 0.1 Seconds. In: *Advances in Cryptology-ASIACRYPT 2016*.

[46] homenc. HELib. <https://github.com/homenc/HElib>.

[47] Microsoft. SEAL. <https://github.com/microsoft/SEAL>.

[48] tfhe. TFHE. <https://github.com/tfhe/tfhe>.

[49] BLUM A, FURST M, KEARNS M, et al. Cryptographic primitives based on hard learning problems. In: *Advances in Cryptology CRYPTO' 93, Berlin, Heidelberg: Springer*, 1994.

[50] GENTRY C. Computing arbitrary functions of encrypted data. In: *Communications of the ACM*, 2010.

- [51] MYERS S, SHULL A. Practical Revocation and Key Rotation . In: CT-RSA, 2018.
- [52] LIU F H, WANG H. Batch Bootstrapping II: Bootstrapping in Polynomial Modulus Only Requires $O(1)$ FHE Multiplications in Amortization. In: Eurocrypt, 2023: 32909.
- [53] LIU F H, WANG H. Batch Bootstrapping I: A New Framework for SIMD Bootstrapping in Polynomial Modulus. In: Eurocrypt, 2023:32907.
- [54] BAE Y J, CHEON J H, CHO W, et al. META-BTS: Bootstrapping Precision Beyond the Limit. In: CCS 2022: 223-234.
- [55] snucrypto.HEAAN. <https://github.com/snucrypto/HEAAN>.
- [56] openfheorg. openfhe-development. <https://github.com/openfheorg/openfhe-development>
- [57] BONTE C, ILIASHENKO I, PARK J, et al. FINAL: Faster FHE instantiated with NTRU and LWE. In: Cryptology ePrint Archive, 2022: 074.
- [58] KAMIL K. NTRU-u-um: Secure Fully Homomorphic Encryption from NTRU with Small Modulus. Cryptology ePrint Archive, Paper 2022/089.
- [59] PEREIRA H V L. Bootstrapping Fully Homomorphic Encryption over the Integers in Less than One Second. In: Garay, J.A. (eds) Public-Key Cryptography - PKC 2021.
- [60] XIANG B W, ZHANG J, DENG Y, et al. Fast Blind Rotation for Bootstrapping FHEs. In: Crypto 2023.
- [61] Siddhartha C , Sayani S , Animesh S, et al. Efficient Threshold FHE with Application to Real-Time Systems. Cryptology ePrint Archive, Paper 2022/1625.
- [62] YUAN M, WANG D, ZHANG F, et al. An Examination of Multi-Key Fully Homomorphic Encryption and Its Applications. Mathematics 2022, 10, 4678.
- [63] JEONGEUN P , SERGI R . Efficient TFHE Bootstrapping in the Multiparty Setting. Cryptology ePrint Archive, Paper 2023/759.
- [64] ANDREA D G , CHIARA M. Breaking the power-of-two barrier: noise estimation for BGV in NTT-friendly rings. Cryptology ePrint Archive, Paper 2023/783.
- [65] CHEN H, ILIASHENKO I, LAINE K. When HEAAN Meets FV: A New Somewhat Homomorphic Encryption with Reduced Memory Overhead. In: Paterson, M.B. (eds) Cryptography and Coding. IMACC 2021.
- [66] KHIN M M A , ENHUI L , JUN J S, et al. Field Instruction Multiple Data. Cryptology ePrint Archive, 2022.
- [67] KITAGAWA F, MATSUDA T. Circular Security Is Complete for KDM Security. ASIACRYPT 2020.

全同态加密构造发展报告

路献辉¹, 王安宇²

1. 中国科学院信息工程研究所, 北京, 100085

2. 清华大学高等研究院, 北京, 100090

通讯作者: 路献辉, E-mail: luxianhui@iie.ac.cn

摘要: 全同态加密允许在加密数据上直接进行计算, 是实现数据要素价值流转, 打破数据壁垒的关键技术。自 2009 年 Gentry 构造第一个全同态加密方案以来, 全同态加密在原理和构造上均取得了长足的进展。本文将从 Gentry 的经典构造开始, 详细介绍全同态加密的设计思想与关键理论, 并回顾目前最主要的几类全同态加密方案的具体构造及最近进展。

关键词: 全同态加密; BGV; BFV; CKKS; GSW; TFHE/FHEW

Report on the Development of Fully Homomorphic Encryption Constructions

LU Xianhui¹, WANG Anyu²

1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing, 100085, China

2. Institute for Advanced Study, Tsinghua University, Beijing, 100090, China

Corresponding author: LU Xianhui, E-mail: luxianhui@iie.ac.cn

Abstract: Fully homomorphic encryption allows direct computation on encrypted data, serving as a key technology to realize the transaction of data elements and break down data barriers. Since Gentry constructed the first fully homomorphic encryption scheme in 2009, there has been significant progress in the theory and construction of fully homomorphic encryption. This paper will start from Gentry's classic construction, providing a detailed introduction to the design philosophy

and key theories of fully homomorphic encryption, and reviewing the constructions and recent advancements of several of the most significant fully homomorphic encryption schemes.

Keywords: Fully Homomorphic Encryption; BGV; BFV; CKKS; GSW; TFHE/FHEW

1 全同态加密

在当今的信息时代,数据已经成为一种重要的资源和财富。许多传统的行业,如金融、医疗和教育,都在利用人工智能、物联网、云计算和大数据等先进的科技手段,为用户提供更优质、更高效、更便捷的产品和服务。这些技术通过从海量的数据中挖掘有价值的信息,打破了原有的行业壁垒,实现了与信息产业的跨界融合。然而,数据的价值也引来了一些风险和挑战。用户、企业乃至国家的数据隐私和安全可能会面临各种威胁与攻击。如何在保证数据可用性的同时,保护数据不被泄露或滥用,是一个亟待解决的问题。全同态加密是一种能够有效解决这一问题的技术。它可以让人们在不解密数据的情况下,对密文进行各种计算和分析,从而实现数据可用不可见的目标。这样即使密文数据被窃取或泄露,也不会造成任何损失。因此,全同态加密是一种在保护数据隐私的基础上,利用数据价值的理想方案。

全同态加密起源于 1978 年由 Rivest 等人以银行为应用背景提出的隐私同态,旨在构造一种支持密文检索的加密机制^[1]。在此后的 30 年里,密码学家们对同态加密展开了大量的研究,相继提出了单同态加密方案和部分同态加密方案。单同态加密方案(Partial Homomorphic Encryption, PHE)仅支持对密文的任意次加法同态或乘法同态,代表方案有支持乘法同态的 GM82 方案^[2]和 ElGamal^[3]方案,以及支持加法同态的 Paillier^[4]方案。部分同态加密方案(Somewhat Homomorphic Encryption, SHE)能够同时支持对密文的有限次加法同态和乘法同态运算,代表方案是 BGN05^[5],其基于双线性映射中的困难问题构造得到,能够支持任意多次的同态加法和一次同态乘法。然而,这些同态加密方案都不能做到同态计算任意深度的电路。直到 2009 年,Gentry 基于理想格提出了第一个全同态加密方案^[6],被誉为密码学界的“圣杯”。

Gentry 的论文是全同态加密研究的奠基之作。这不仅是因为他的方案对全同态加密有着从无到有的划时代意义,更重要的是他提出了一种通用的全同态加密构造思想,为后来的研究者提供了指导和启发。Gentry 的全同态蓝图可以概括为 3 个步骤:第一步,设计一个部分同态加密方案,即只能支持有限深度的同态计算;第二步,将解密过程压缩成一个简单的电路;第三步,在一定次数的同态运算后有序地执行同态解密操作(自举),从而实现无限深度的同态计算,得到了一个全同态加密方案。

自 Gentry 的突破性进展以来,全同态加密方案的设计和构造已经取得了显著的成果。首先是密码底层困难问题的选择,2011 年,Brakerski 和 Vaikuntanathan 首次将格上的误差学习

问题 (LWE) 以及其在环结构上的变体 (RLWE) 引入 FHE 方案中^[10]。基于 LWE/RLWE 构造的密码系统具有解密电路深度较低的特点, 因此这类问题适合应用于全同态加密算法设计。

目前实用的 FHE 方案主要基于这两个困难问题, 这些方案可以根据它们的明文空间和计算方法被划分为 3 个类别: BGV/BFV 方案针对有限域上的明文执行算术运算; CKKS 方案在实数/复数上进行近似的算术运算, 其解密结果与明文运算结果存在一定的误差; GSW/FHEW/TFHE 方案针对布尔明文计算逻辑电路。

FHE 理论发展路线如图 1 所示。

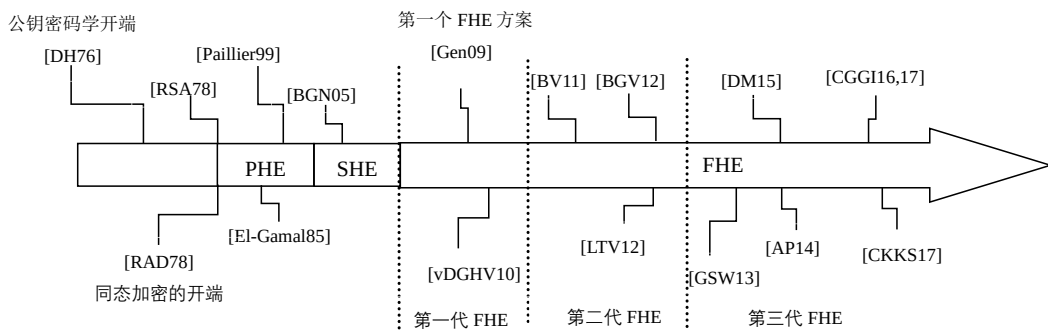


图 1 FHE 理论发展路线

本文旨在详细介绍全同态加密设计思想和构造实例的发展进程。首先, 在第 1 节的后续部分介绍全同态加密的基本概念和性质。其次, 在第 2 节中, 将详细阐述 Gentry 提出的全同态设计蓝图, 并在第 3 节介绍他在理想格上构造的第一个全同态加密方案。最后, 在第 4~7 节中, 将分别介绍 BGV/BFV、CKKS、GSW 及 FHEW/TFHE 4 种主流全同态加密方案的设计思路、关键技术和具体构造。

1.1 全同态加密方案的定义

全同态加密密钥方案由 4 个算法组成: 密钥生成算法 (KeyGen)、加密算法 (Encrypt)、解密算法 (Decrypt) 和密文计算算法 (Evaluate), 详情如下。

- $\text{KeyGen}(1^\lambda) \rightarrow (pk, sk, evk)$: 密钥生成算法, 该算法输入安全参数 λ , 生成用于加密的公钥 pk 、用于解密的私钥 sk 和用于密文计算的密钥 evk , 一般来说, evk 也属于公开密钥的一部分。
- $\text{Encrypt}(pk, m) \rightarrow c$: 加密算法, 该算法输入公钥 pk 和明文 m , 输出对应的密文 c 。
- $\text{Decrypt}(sk, c) \rightarrow m$: 解密算法, 该算法输入私钥 sk 和密文 c , 输出对应的明文 m 。

• $\text{Evaluate}(\text{evk}, c_1, \dots, c_l, f) \rightarrow c_f$: 密文计算算法, 该算法输入计算密钥 evk 、一组密文 $c_1, \dots, c_l$ 和函数 $f: \mathcal{M}^l \rightarrow \mathcal{M}$, 输出对应的计算密文 c_f 。

一个有意义的 (non-trivial) 全同态加密方案需要满足 3 种基本性质: 安全性、正确性、紧凑性。

安全性: 所有加密方案的基本要求, 全同态加密方案一般是在选择明文攻击下不可区分 (IND-CPA) 安全的。

正确性: 对于任意的明文 $m = (m_1, m_2, \dots, m_n)$ 、任意的密文 $c = (c_1, c_2, \dots, c_n)$, 其中 $c_i = \text{encrypt}(m_i)$, 以及任意的 KeyGen 算法生成的密钥对 (pk, sk) 、任意电路 C , 需要保证 $c' = \text{Evaluate}(\text{evk}, C, c)$ 且 $\text{Decrypt}(sk, c') = C(m)$ 成立, 则称该同态加密方案满足正确性要求。

紧凑性: 如果在同态加密算法中, Evaluate 步骤输出的大小独立于计算函数 f , 那么称该同态方案是紧凑的, 此时密文长度不受计算电路深度的影响。

如果一个方案对于某个函数集合中所有的电路都满足安全性、正确性、紧凑性这 3 种要求, 那么称该方案关于该函数集合是同态的。如果对于所有函数, 该方案均满足同态性, 那么称该方案是全同态的。

1.2 全同态加密方案中的基本运算和代表方案

对于任意函数都存在一个与之等价的电路, 在全同态方案中, 通常会将要计算的函数转化成电路模型进行研究。电路模型可分为逻辑电路和算术电路两种。逻辑电路是由一些基本的逻辑门组成的, 如“与门”“或门”“非门”等。算术电路是由一些基本的算术运算组成的, 如加法、乘法等。不同的电路模型需要不同的同态计算单元。

对于逻辑电路来说, 我们需要考虑图灵完备的基础运算单元。例如, NAND 门自身就是一个图灵完备的逻辑门集合, 任何逻辑函数都可以用 NAND 门来表示。常见的完备的电路门集合还包括“与门”和“异或门”; Gentry 的方案就是通过有限域 $\mathbb{F}_2$ 上的加乘运算实现了与门和异或门, 最终实现全同态加密。另外, 三代加密方案 (TFHE、FHEW) 可以实现各种逻辑门的同态计算^[7-9], 从而也可以实现全同态加密。此类方案主要是针对单比特运算, 在非线性函数运算中也有较好的表现。

对于算术电路, 需要同态地实现加法和乘法。正如上面提到的, 有限域 $\mathbb{F}_2$ 上的加乘运算和门电路是等价的, 因为逻辑电路中的“与门”和“异或门”等同于比特乘法与比特加法。在其他有限域上则需要考虑一些代数细节和近似问题。有限域上的加乘代表方案是 BGV^[10-11]、BFV 方案^[12]。此外, CKKS 实现了近似的加乘计算^[13], 与 BGV 和 BFV 方案不同, CKKS 将噪声视为消息的一部分。由于在现实世界的某些计算需求下, 都允许数据与真实值之间存在一定误差 (如神经网络), 因此支持浮点数运算的 CKKS 方案也具有较高的实用价值。

2 Gentry 的全同态加密方案设计蓝图

目前，在全同态加密方案设计中，为了保证加密算法的安全性，需要引入噪声来掩盖消息，并且只有当密文中的噪声小于临界值时才可以恢复得到正确的消息。但在进行密文运算时，该噪声会随着同态运算的进行而不断地累积。一旦该噪声的尺寸超出某个阈值，就会出现解密错误。Gentry 解决噪声问题，构造全同态方案最重要的思想是自举操作，其实际是同态地执行解密函数，即 $\text{Eval}(\text{evk}, \text{Decrypt}(\text{sk}, c))$ 。如图 2 所示，粗框表示一个噪声比较大的密文，自举过程首先将该密文与自举密钥进行组合得到一个新的密文，然后对这个新的密文进行关于 g 函数的同态运算，此时 g 函数为同态方案的解密函数，最终得到一个噪声比较小的密文。

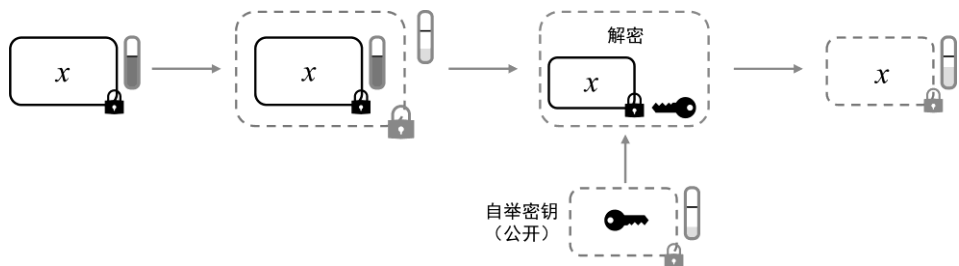


图 2 自举操作过程

为了保证方案是可自举的 (Bootstrappable)，有两个前提条件。首先，解密电路的深度应该小于部分同态加密方案所能计算的电路深度；否则就需要设计压缩解密电路的方法，如使用稀疏子集和假设等。其次，为了满足同态加密算法的紧凑性要求，自举密钥的尺寸不能受运算电路的影响，如图 2 所示。当虚框和粗框部分使用不同的密钥加密时，自举密钥为新密钥加密原密钥产生的密文。由于每次自举的自举密钥不同，因此自举密钥的尺寸与需要运算的电路的深度相关。为了解决这个问题，我们引入了循环安全的概念，即私钥由自身公钥加密作为自举密钥，这样的密钥公开后，不影响整体方案的安全性，从而整个同态运算的过程仅需要一把自举密钥，满足紧凑性要求。

综上所述，Gentry 的全同态加密方案设计蓝图如图 3 所示。

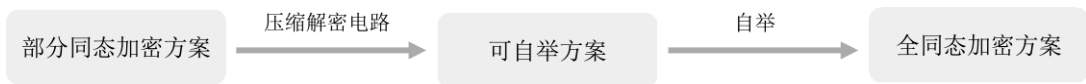


图 3 Gentry 的全同态加密方案设计蓝图

(1) 构建一个部分同态加密方案，这个方案可以支持一定深度的加法和乘法运算。在执行有限次同态操作之后，噪声达到上限。

(2) 压缩解密电路。判断该方案的解密电路的深度是否小于部分同态所支持的最大深度。如果大于, 需要压缩函数的解密电路, 从而获得一个可自举的 (Bootstrappable) 部分同态加密方案。

(3) 执行自举操作获得全同态加密方案, 即同态地执行部分同态加密方案的解密电路(或者压缩后的解密电路), 该步骤需要使用自举密钥, 且需要循环安全假设。该过程输出一个噪声相对低的密文, 且保持明文消息没有改变。

3 基于理想格的全同态加密方案

Gentry 的设计蓝图表明, 如果部分同态加密方案能够同态计算自己的解密电路, 就可以通过同态解密来降低密文的噪声, 使其支持更多的同态运算能力, 从而能够实现全同态加密。这个思想为全同态加密方案的设计提供了全新的角度: 一方面, 应该平衡算法设计中的“结构性”和“随机性”, 选取和研究基于特定代数结构的加密方案, 使得它同时具备加法同态性和乘法同态性, 并拥有尽可能大限度的同态计算能力; 另一方面, 应该关注和研究那些解密电路复杂度较低的方案, 最小化解密电路的复杂度, 弥合自举设计中同态计算能力和解密电路深度之间的差距, 使得方案的解密电路能够被自身同态计算, 在循环安全的假设下实现全同态加密方案。

这种思想引导着 Gentry 在具体的方案设计中使用了理想格这一代数结构。其设计理论: 首先基于理想格上的理想陪集问题构造一个部分同态加密方案, 由于密文中噪声的存在, 进行加法同态和乘法同态时会使噪声增加, 因此该方案只能支持有限次数的同态运算操作; 在考虑自举运算时, 由于基于理想格的方案解密电路复杂度超过了它自身能够计算的上限, 因此引入了压缩加密电路, 降低解密算法的计算复杂度, 最后通过自举实现全同态加密。

后续一系列的主流方案^[7-14]也都是基于格密码进行设计的, 常用的困难问题包括 LWE/RLWE^[15-16]和 NTRU^[17]等。一方面, 格具有良好的代数性质; 另一方面, 格密码具有解密电路复杂度较低的特点, 符合 Gentry 提出的自举思想的要求, 适合构造全同态加密算法。接下来详细介绍 Gentry 基于理想格设计的全同态加密方案。

3.1 基于理想格的部分同态加密方案

首先, Gentry 利用环和理想的代数性质, 构造了一个初始的部分同态加密方案。简单来说, 他把消息看作一个环元素, 密文形式可以被简单表示为 $m+xI$, x 是环中的元素, I 是环的理想。这种编码天然具有同态加乘的性质。其次, 我们给出具体的方案构造。

该方案的构造选定环 R ，取它的理想 $I \subset R$ 和理想的一组基 B_I ，算法 $\text{IdealGen}(R, B_I)$ 对满足 $I + J = R$ 的理想 J (I 与 J 互素) 生成其公开基 B_J^{pk} 和私有基 B_J^{sk} 。算法 $\text{Samp}(x, B_I, R, B_J)$ 表示从陪集 $x+I$ 中取样。

- $\text{KeyGen}(R, B_I)$: 输入环 R 和理想 I 的基 B_I ，令 $(B_J^{sk}, B_J^{pk}) \xleftarrow{R} \text{IdealGen}(R, B_I)$ ，明文空间 P 为 $R \bmod B_I$ (的子集)。公钥包含 R 、 B_I 、 B_J^{pk} 和 Samp ，私钥包含 B_J^{sk} 。
- $\text{Encrypt}(pk, m)$: 输入公钥 pk 和明文 $m \in P$ ，令 $c' \leftarrow \text{Samp}(m, B_I, R, B_J^{pk})$ ，输出 $c \leftarrow c' \bmod B_J^{pk}$ 。
- $\text{Decrypt}(sk, c)$: 输入公钥 sk 和密文 c ，输出 $m \leftarrow (c \bmod B_J^{sk}) \bmod B_I$ 。
- $\text{Evaluate}(pk, C, c)$: 输入公钥 pk ，电路 C (由 Add_{B_I} 和 Mult_{B_I} 门组成)，以适当的顺序调用下面给出的加法和乘法来计算输出的密文。

$\text{Add}(pk, c_1, c_2)$: 输出 $c_1 + c_2 \bmod B_J^{pk}$ 。

$\text{Mult}(pk, c_1, c_2)$: 输出 $c_1 \times c_2 \bmod B_J^{pk}$ 。

该方案的正确性和紧凑性是基于环与理想的代数性质。其安全性基于理想陪集问题 (Ideal Coset Problem) 的困难性。

理想陪集问题 (ICP): 选定环 R 和格基 B_I ，理想生成算法 IdealGen ，抽样算法 Samp 。挑战者随机选取挑战比特 $b \xleftarrow{R} \{0, 1\}$ ，运行理想生成算法 $(B_J^{sk}, B_J^{pk}) \xleftarrow{R} \text{IdealGen}(R, B_I)$ 。当 $b = 0$ 时，调用抽样算法 $r \xleftarrow{R} \text{Samp}(R)$ 并计算 $t \leftarrow r \bmod B_J^{pk}$ ；否则从 $R \bmod B_J^{pk}$ 中均匀随机地选取 t 。挑战者将对应的 (t, B_J^{pk}) 发送给敌手，敌手需要推测出正确的挑战比特 b 。

简单来说，ICP 要求敌手区分 t 是从 B_J^{pk} 中均匀随机选取的，还是通过抽样算法 Samp 抽取的，满足特定的抽样分布。该问题是计算困难的，Gentry 证明了部分同态加密方案的安全性可以被归约到 ICP。

接下来需要考虑这个抽象方案的具体实现。令环 $R = \mathbb{Z}[x] / f(x)$ ， $f(x) \in \mathbb{Z}[x]$ 是一元 n 阶多项式，则该环中的元素 $v \in R$ 同样对应一个由其多项式系数组成的向量，理想对应由列向量 $\{v \times x^i \bmod f(x) : i \in [0, n-1]\}$ 生成的格，称为理想格。对应的实现为理想格上的部分同态加密方案。

3.2 压缩解密电路

回顾部分同态加密方案的解密电路： $\text{Decrypt}(sk, c) : m \leftarrow (c \bmod B_J^{sk}) \bmod B_I$ 。在全同态加密方案设计中，同态计算模函数 $\bmod$ 往往复杂度较高，是同态解密算法设计的重点。主要原因是：模函数是一个周期的非线性函数，很难使用加法和乘法进行表达。后续的一系列方案使

用了不同的方法来实现同态求模：在 BGV/BFV 方案中，模函数被等价于比特提取，并借助有限域的代数性质，通过有限域上的多项式函数进行实现；在 CKKS 方案中，由于该方案本身是一个近似方案，因此模函数也可以通过拉格朗日插值、Remez 算法等进行多项式近似；在 TFHE 方案中，解密运算所需要的加法群被同构到 $2N$ 阶多项式乘法群，加法群中的模加运算可对应到多项式幂次上的 $\text{mod } 2N$ 加法。

这里我们只关注理想格上的同态加密方案，由于理想格良好的代数性质，模函数的实现是自然的：在理想格上， $t \bmod B_I$ 可以通过 $t - B \cdot \lfloor B^{-1} \cdot t \rfloor$ 有效计算，其中 $\lfloor \cdot \rfloor$ 为舍入操作。因此，同态解密需要计算 $(c - B_J^{sk_1} \cdot \lfloor B_J^{sk_2} \cdot c \rfloor) \bmod B_I$ （在循环安全假设下 sk_1, sk_2, sk ），其中 $c \in \mathbb{Z}^n$ ， $B_J^{sk_1} = B_J^{sk} \in \mathbb{Z}^{n \times n}$ ， $B_J^{sk_2} = (B_J^{sk})^{-1} \in \mathbb{Q}^{n \times n}$ ， B_I 是 $R = \mathbb{Z}[x] / f(x)$ 的理想 I 的基。具体的解密分为以下 3 步。

步骤 1：从 $B_J^{sk_2} \cdot c$ 中生成 n 个向量。

步骤 2：利用 n 个向量 $x_1, \dots, x_n$ ，计算 $\lfloor \sum x_i \rfloor$ 生成整数向量 $y_1, \dots, y_{n+1}$ 。

步骤 3：计算 $m \leftarrow c - B_J^{sk_1} \cdot (\sum y_i) \bmod B_I$ 。

为了达到方案可自举的目的，需要提供一种压缩解密电路的方法。整体的压缩思想是将解密的计算量分担一部分给加密算法。也就是说，由加密者对自己的初始密文进行预处理，为解密者留下更少的工作。这种两个阶段的解密方法是在服务器辅助密码学中发现的，具体来说，设 E^* 为原方案，构造新的部分同态加密方案如下。

- $\text{KeyGen}_E(\lambda)$ ：运行 $\text{KeyGen}_{E^*}(\lambda)$ ，得到密钥 (pk^*, sk^*) ， $(sk, \tau) \leftarrow \text{SplitKey}_E(pk^*, sk^*)$ ，私钥为 sk ，公钥为 (pk^*, τ) 。

- $\text{Decrypt}_E(pk, m)$ ：运行原来的加密算法， $c^* \leftarrow \text{Encrypt}_{E^*}(pk^*, \pi)$ ，输出 c ，包括 c^* 与 $\text{ExpandCT}_E(pk, c^*)$ 。

- $\text{Decrypt}_E(sk, c)$ ：使用 sk 和扩展的密文解密。

$\text{Add}_E(pk, c_1, c_2)$ ：从 (c_1, c_2) 中提取 (c_1^*, c_2^*) ，计算 $c^* \leftarrow \text{Add}_{E^*}(pk^*, c_1^*, c_2^*)$ ，输出 c ，包含 c^* 和 $\text{ExpandCT}_E(pk, c^*)$ 。 $\text{Mult}_E(pk, c_1, c_2)$ 同理。

给定 $\gamma_{\text{set}}(n)$ 、 $\gamma_{\text{subset}}(n)$ 两个参数，预处理算法 SplitKey_E 、 ExpandCT_E ，以及新的解密算法 Decrypt_E 的具体构造如下。

- $\text{SplitKey}_E(sk^*, pk^*)$ ：输入 (sk^*, pk^*) ，输出 (sk, τ) ，其中， τ 为 $\gamma_{\text{set}}(n)$ 个向量 $t_1, t_2, \dots, t_{\gamma_{\text{set}}(n)}$ 的集合，其中 $t_i \in J^{-1} \bmod B_I$ 均匀随机， τ 中存在子集和 $S \subseteq \{1, \dots, \gamma_{\text{set}}(n)\}$ ，大小为 $\gamma_{\text{subset}}(n)$ ，满足 $\sum_{i \in S} t_i \in \mathbf{v}_J^{sk^*} + I$ （短向量 $\mathbf{v}_J^{sk^*} \in J^{-1}$ ，满足：存在 $u = 1 + I$ 使得 $u \times (\mathbf{v}_J^{sk})^{-1} \in 1 + I$ ）； sk 为 $\gamma_{\text{subset}}(n) \times \gamma_{\text{set}}(n)$ 的矩阵 M ， $M_{ij} = 1$ 当且仅当 j 是 S 的第 i 个成员。

- $\text{ExpandCT}_E(pk, c^*)$: 输出 $c_i \leftarrow t_i \times c^* \bmod B_l$, 其中 $i \in [1, \gamma_{\text{set}}(n)]$ 。
- $\text{Decrypt}_E(sk, c)$: 输入私钥 sk 和密文 c , 进行以下步骤。

步骤 0: 计算 $w_{ij} \leftarrow M_{ij} \cdot c_j$ 。

步骤 1: 计算 $x_i \leftarrow \sum_{j=1}^{\gamma_{\text{set}}(n)} w_{ij}$ 。

步骤 2: 从 $x_1, \dots, x_{\text{subset}(n)}$ 中产生整数向量 $y_1, \dots, y_{\text{subset}(n)+1}$, 和为 $\lfloor \sum x_i \rfloor$ 。

步骤 3: 计算 $m \leftarrow c^* - (\sum y_i) \bmod B_l$ 。

在该实现方案中, 引入了 SplitKey_E 算法, 这个算法由加密者执行, 生成了大小为 $\gamma_{\text{set}}(n)$ 的向量集合, 其中包含大小为 $\gamma_{\text{subset}}(n)$ 的稀疏子集。解密阶段的步骤 2 只需要利用稀疏子集来计算加和即可, 可以通过调整参数 $\gamma_{\text{set}}(n)$ 、 $\gamma_{\text{subset}}(n)$, 大幅度减少该步骤的计算量。值得注意的是, 由于新算法使用了稀疏子集求和, 因此新方案的安全性还必须基于稀疏子集和假设, 即假设下述稀疏子集和问题 (Sparse Subset Sum Problem) 是困难的。

稀疏子集和问题 (SSSP):

给定 $\gamma_{\text{set}}(n)$ 、 $\gamma_{\text{subset}}(n)$ 两个参数, 设 q 是一个正整数, 挑战者随机选取挑战比特 $b \xleftarrow{R} \{0, 1\}$, 当 $b=0$ 时, 生成 τ , 其为 $\gamma_{\text{set}}(n)$ 个整数 $\{a_1, \dots, a_{\gamma_{\text{set}}(n)}\}$ 的集合, 其中每个数在 $[-q/2, q/2]$ 中均匀随机, 且集合 τ 中存在大小为 $\gamma_{\text{subset}}(n)$ 的稀疏子集, 指标集 $S \subseteq \{1, \dots, \gamma_{\text{set}}(n)\}$, 满足: $\sum_{i \in S} a_i = 0 \bmod q$; 当 $b=1$ 时, 对集合不加约束。挑战者将对应的 τ 发送给敌手, 敌手需要推测出正确的挑战比特 b 。

综上, 本文介绍了 Gentry 提出的利用稀疏子集和来压缩解密电路的方法, 通过调整参数 $\gamma_{\text{set}}(n)$ 、 $\gamma_{\text{subset}}(n)$, 可以使解密电路更加简化, 为实现理想格上部分同态加密方案的自举提供了理论工具。

3.3 自举

想要实现自举, 该方案的同态计算能力必须超过同态计算解密电路的复杂度。因此, 这里分为 3 个步骤来介绍 Gentry 提出的自举算法: 首先提出允许电路 (Permitted Circuits) 的概念来衡量部分同态加密算法的同态计算能力; 其次评估压缩解密电路前后的解密电路复杂度; 最后通过调整压缩解密电路的参数来实现自举。

对于部分同态加密方案, 其所能处理的计算函数的电路是有限的。为了刻画部分同态加密方案可以计算的电路范围, 引入如下定义。

定义 3-1 (广义电路) 设 C 为 $\text{mod } B_I$ 电路, C 的广义电路 $g(C)$ 为密文环 R 中的加法 “+” 和乘法 “ $\times$ ” 替代 C 的 Add_{B_I} 和 Mult_{B_I} 运算形成的新电路。

定义 3-2 (X_{Enc} 与 X_{Dec}) 设 X_{Enc} 为 Samp 的像 (Encrypt 输出的所有密文都在 $X_{\text{Enc}}+J$ 中)。令 X_{Dec} 等于 $R \bmod B_J^{p_k}$ 。

定义 3-3 (r_{Enc} 与 r_{Dec}) 令 r_{Enc} 为能使得 $X_{\text{Enc}} \subseteq B(r_{\text{Enc}})$ 满足的最小值, 其中 $B(r)$ 为半径 r 的球; 令 r_{Dec} 为能使得 $X_{\text{Dec}} \subseteq B(r_{\text{Dec}})$ 满足的最大值。

基于这些定义, 可以刻画出部分同态加密方案能够同态计算的电路集合, 称为允许电路。

定义 3-4 (允许电路) 假设

$$C'_E = \{C : \forall (x_1, \dots, x_t) \in B(r_{\text{Enc}})^t, g(C)(x_1, \dots, x_t) \in B(r_{\text{Dec}})\}$$

也就是说, C'_E 是一组 $\text{mod } B_I$ 电路, 广义化时, 若输入在 $B(r_{\text{Enc}})$ 中, 则输出总在 $B(r_{\text{Dec}})$ 中。若 $C_E \subseteq C'_E$, 则称 C_E 是允许电路。

接下来考虑允许电路的最大深度, 当计算电路深度小于等于 $\log_2 \log_2 r_{\text{Dec}} - \log_2 \log_2 (\gamma_{\text{Mult}}(R) \cdot r_{\text{Enc}})$ 时, 容易计算得到对于所有 $x_1, \dots, x_t \in B(r_{\text{Enc}})$, 满足 $C(x_1, \dots, x_t) \in B(r_{\text{Dec}})$ 。其中 $\gamma_{\text{Mult}}(R)$ 是环 R 的扩张因子。于是, 为了最大化方案可以计算的电路深度, 应该最小化 $\gamma_{\text{Mult}}(R)$ 与 r_{Enc} , 并最大化 r_{Dec} 。然而, 加密算法的正确性和语义安全性会限制 $r_{\text{Dec}}/r_{\text{Enc}}$ 的大小: 一方面 r_{Dec} 不能过大, 需要保证 $r_{\text{Dec}} < \lambda_1(J)$, 后者是 J 中最小非 0 向量的长度, 否则无法正确解密; 另一方面 r_{Enc} 不能过小, 否则 ICP 问题的安全级别会降低。为了保证解密正确性和语义安全性, Gentry 最终选取了 $r_{\text{Dec}} = 2^{n^{\alpha_1}}$, $\gamma_{\text{Mult}}(R) \cdot r_{\text{Enc}} = 2^{n^{\alpha_2}}$, 使得允许电路的最大深度为 $(c_1 - c_2) \log_2 n$ 。

接下来评估解密电路在压缩前后的计算复杂度。压缩解密电路前, 计算代价最大的是步骤 2, 其包含 n 个向量的加法, 需要被表达为比特加法。 n 个 k bit 的数相加, 需要的线路深度为 $\Omega(\log_2 n + \log_2 k)$ 。注意, 经过计算, 允许电路的最大深度为 $O(\log_2 n)$, 因此压缩加密电路前的方案不能进行自举。具体来说:

对于 $i \in [1, t]$, 令 $a_i = (\dots, a_{i,1}, a_{i,0}, a_{i,-1}, \dots)$ 为 $\text{mod } B_I$ 下的二进制表示的实数, 并满足 $\sum_i a_i \bmod 1 \in [-1/4, 1/4]$ 。若存在一个 $\text{mod } B_I$ 电路 C 产生 $t+1$ 个整数 $z_1, \dots, z_{t+1}$, 满足 $\sum_i z_i = \lfloor \sum a_i \rfloor$, 且当 $g(C)$ 的输入在 $B(r_{\text{in}})$ 中时, 其输出在 $B(r_{\text{out}})$ 中, 需要满足如下条件:

$$r_{\text{out}} \leq (\gamma_{\text{Mult}}(R) \cdot n \cdot \|B_I\| \cdot (1 + \gamma_{\text{Mult}}(R) \cdot r_{\text{in}})^t \cdot t)^{\text{polylog}(t)}$$

对于 $\|B_I\| \leq r_{\text{in}}$, $t \leq n$ 与 $\gamma_{\text{Mult}}(R) = n^{\Omega(1)}$, 则有

$$r_{\text{out}} \leq (\gamma_{\text{Mult}}(R) \cdot r_{\text{in}})^{t + \text{polylog}(t)}$$

步骤 2 中的 $t = n$, 意味着 $r_{\text{Dec}}/r_{\text{Enc}} \geq r_{\text{out}}/r_{\text{in}} \geq 2^n$, $r_{\text{Dec}}/r_{\text{Enc}}$ 超出了允许范围。

压缩解密电路后，解密步骤 2 中的加法个数由 n 个降低到 $\gamma_{\text{subset}}(n)$ 个，则可以通过调整 $\gamma_{\text{subset}}(n)$ 的大小，实现自举。 $\gamma_{\text{subset}}(n)$ 需要满足的条件为

$$\gamma_{\text{subset}(n)} \cdot \log_2^{\epsilon_1} \gamma_{\text{subset}(n)} \leq \left(\frac{\log_2(r_{\text{Dec}}/2)}{2^{\epsilon_2} \cdot \log_2(\gamma_{\text{Mult}}(R) \cdot r_{\text{Enc}})} \right)$$

此时，当输入小于加密半径 r_{Enc} 时，电路的输出为 $(\gamma_{\text{Mult}}(R) \cdot r_{\text{Enc}})^{2^{\epsilon_2} \cdot (\gamma_{\text{subset}}(n) + \text{polylog}(\gamma_{\text{subset}}(n)))}$ ，将上述条件代入可得输出小于解密半径 r_{Dec} 。综上，通过合理设置参数，我们得到了一个理想格上的可自举的部分同态加密方案。通过同态解密，即可实现全同态加密。

4 BGV/BFV 方案

4.1 BGV 方案

2011 年，Brakerski 和 Vaikuntanathan 基于 LWE/RLWE 问题提出了 BV 方案^[10,18]，标志着第二代全同态加密的诞生。BGV 方案^[11]是由 Brakerski、Gentry、Vaikuntanathan 在 2011 年基于 BV 方案提出的层级全同态方案（Leveled Fully Homomorphic Scheme）。该方案使用批处理（Batching）和模变换（Modulus Switch）技术，避免使用高成本开销的自举。同时，Brakerski 等人也在文献[11]中给出了对应的自举方法，可将 BGV 方案转换为全同态加密方案。BGV 方案在全同态加密算法库 HELib^[19]中实现。BGV 方案流程如下：

参数初始化：

d 为 2 的幂， q 为奇整数模， χ 是 $R = \mathbb{Z}[x]/\langle x^d + 1 \rangle$ 上的噪声分布。 B 为 χ 输出元素的长度（为了保持安全性， B 需要设置足够小）。对任意自然数 p ，记 $R_p = \mathbb{Z}_p[x]/\langle x^d + 1 \rangle$

密钥生成：

输入安全参数 λ ，随机选择小密钥 $s \in \chi$ ，设置私钥 $sk = s = (1, s)$ 。均匀随机生成 $a' \in R_q$ ，计算 $b = a's + 2e$ （其中 e 是 χ 中的随机噪声）

输出 sk 和 $pk = a = (b, -a')$

加密：

输入公钥 $pk \in R_q^2$ 和消息 $m \in R_2$ 。该过程将 m 转换为向量 $\mathbf{m} = (m, 0) \in R_q^2$ ，并随机选取 $r, e_0, e_1 \in \chi$

输出密文 $\mathbf{c} = \mathbf{m} + 2(e_0, e_1) + ar$ ，即 $\mathbf{c} = (c_0, c_1) = (m + 2e_0 + br, 2e_1 - a'r) \in R_q^2$

解密：

输入私钥 $s \in R_q^2$ 和密文 $\mathbf{c} \in R_q^2$ 。该过程计算 $\langle \mathbf{c}, s \rangle = c_0 + c_1 s = m + 2e_0 + 2e_1 s + 2er$

输出 $((m + 2(e_0 + e_1s + er)) \bmod q) \bmod 2 = m$

同态加法:

即逐元素相加

同态乘法:

由 $\langle \mathbf{c}, \mathbf{s} \rangle \cdot \langle \mathbf{c}', \mathbf{s} \rangle = (c_0 + c_1s)(c'_0 + c'_1s) = c_0c'_0 + (c_0c'_1 + c_1c'_0)s + c_1c'_1s^2 = d_0 + d_1s + d_2s^2$ ，得扩展私钥 $(1, s, s^2)$ 加密的扩展密文 (d_0, d_1, d_2) 。每次乘法都会使解密密钥扩张。因此 BGV 方案使用密钥转换 (Key Switching) 技术，利用 s^2 在 s 下的加密将密文项转换为 $\overline{c}_0 + \overline{c}_1s$ 。实际上， $\text{Enc}_s(s^2) = (\beta, -\alpha)$ ，其中 $(\beta, -\alpha) = (s^2 + 2e + d'rs, 2e_1 - d'r) \approx (s^2 + \alpha s, -\alpha)$ 。因此， $s^2 \approx \beta - \alpha s$ ，扩展密文 $d_0 + d_1s + d_2s^2$ 得以转换为对应于相同明文的常规密文 $c_0 + c_1s$

在 CRYPTO 2012 中，Gentry、Halevi、Smart^[20]给出了新的密钥交换技术和双 CRT 表示 (double-CRT Representation) 下的模交换技术，并基于 BGV 方案同态实现 AES-128。而 Gentry、Halevi、Smart^[21]又在 EUROCRYPT 2012 中使用 BGV 方案与 Smart 和 Vercauteren^[22]的批处理技术，给出了以 $\text{poly}(\log_2(\cdot))$ 开销成本实现的算术电路同态运算。他们还给出了部分场景下同态运算的加速方法，包括通过线性操作的成本同态实现 Frobenius 映射等。在 CRYPTO 2012 中，Brakerski 基于尺度不变性 (Scale Invariant) 技术提出了 BGV 方案的变种^[23]。这种方法替代了以往的模变换技术，将密文和噪声以密文模数 q 的因子缩减，将同态乘法产生的噪声增长从指数级降为线性级。

4.2 BFV 方案

2012 年，Fan 和 Vercauteren 对 Brakerski^[11]的方案进行了改进，并后续被命名为 Brakerski-Fan-Vercauteren (BFV) 方案或 FV 方案^[24]。Fan 等人通过使用模变换 (Modulus Switching) 简化了 BFV 方案自举的分析，并引入了两种优化重线性化过程的方法，使得方案具有更小的重线性化密钥和更好的同态计算效率。该方案基于 RLWE 问题设计，并在微软 SEAL 库中实现。BFV 方案流程如下：

参数初始化:

d 为 2 的幂， $R = \mathbb{Z}[x]/\langle x^d + 1 \rangle$ 。 q, p 为正整数。令 $\Delta = \lfloor q/p \rfloor$ ， $r_t(q) = q \bmod p$ 。对任意自然数 t ，记 $R_t = \mathbb{Z}_t[x]/\langle x^d + 1 \rangle$ 。 χ 为 R_q 上以 B 为上界的概率分布

密钥生成:

输入安全参数 λ ，输出小密钥 $sk = s \in \chi$ 。随机均匀生成 $a \in R_q$ ，计算 $-(a \cdot s + e) \bmod q$ (其中 $e \in \chi$ 是小的随机噪声)

输出 sk 和 $pk = (p_0, p_1) = (-(a \cdot s + e) \bmod q, a)$

加密:
输入公钥 $pk \in R_q^2$ 和消息 $m \in R_p$ 。随机选取 $u, e_1, e_2 \in \mathcal{X}$
输出密文 $c = (c_0, c_1)$ ，其中 $c_0 = (p_0 \cdot u + e_1 + \Delta m) \bmod q$ ， $c_1 = (p_1 \cdot u + e_2) \bmod q$
解密:
输入私钥 $s \in R_q$ 和密文 $c \in R_q^2$
该过程计算 $\lfloor p \cdot (c_0 + c_1 \cdot s \bmod q) / q \rfloor \bmod p = m$ 。密文 c 可以看作 s 的多项式，即 $c(s)$
同态加法:
即 $c_1(s) + c_2(s) \bmod q$
同态乘法:
两个密文的乘法会得到一个二次多项式，即 $c_1(s) \cdot c_2(s) = \alpha_0 + \alpha_1 \cdot s + \alpha_2 \cdot s^2$ 。以上结果可以通过重线性化（Relinearization）过程降维，从而转化为可解密的密文

Bajard 等人给出了应用于大系数密文 BFV 方案的 RNS 变种 BEHZ^[25]，在环维度 $2^{11} \sim 2^{15}$ 时，解密加速 5~20 倍，乘法加速 2~4 倍。该变种后续由 Halevi 等人^[26]在 CT-RSA 2019 中给出了改进版本 HPS 变种，提出了更简单、更高效的 CRT 基扩张和缩放过程，并验证了该方案可以适用于包括 YASHE 在内的任何基于原始 Brakerski 方案的具有尺度不变性的同态加密方案。Al Badawi 等人^[27]实现了文献[25-26]的方法，验证发现 HPS 变种相比 BEHZ 变种具有更优化的解密和同态乘法运行时间。然而，后续 Bajard 等人^[28]证明了 BEHZ 和 HPS 的实际噪声增长非常接近，而 HPS 的运行时间仅比 BEHZ 稍微占优。

在 EUROCRYPT 2018 中，Chen 和 Han^[29]通过给出低次的 lowest-digit-removal 多项式改进了同态位提取算法（Homomorphic Digit Extraction Algorithm），并应用该算法改进了大明文模数 BGV 方案和 BFV 方案的自举过程。同时为了在特定参数场景下节省自举带来的消耗，文献[29]中还提出了在 batching 同构意义下 R_t 的子空间 Z_t^k 为明文空间处理的“slim Bootstrapping”，仅需要进行一次位提取操作。基于此工作，Halevi 和 Shoup^[30]改进了 HElib 中的线性变换算法。而在 CT-RSA 2018 中，Chen 等人^[31]又为了解决明文多项式系数增长的问题，通过利用 Hoffstein 和 Silverman^[32]应用于 NTRU 中的方法，将明文空间从 R_t 转换到 $Z_{b^{n+1}}$ ，给出了相比原始 BFV 方案噪声增长更少且乘法深度更大的改进版本。

在 BGV 方案和 BFV 方案中，为了避免解密失败，噪声增长始终需要控制在一定范围内，这使得参数选取会受到制约。Costache 等人^[33]证明，当明文模数和乘法深度较小时，BGV 方案需要比 BFV 方案更大的参数集；反之，当明文模数较大时 BGV 方案比 BFV 方案表现更好。在 ASIACRYPT 2021 中，Kim 等人^[34]改进了 BFV 方案的加密和同态乘法，并给出了不需

要动态噪声估计的计算 BGV 方案密文模数的新方法。文献[34]的理论分析和实验结果还说明, 该 BFV 改进版本在不同明文模数下都具有比 BGV 方案更好的噪声增长。

5 CKKS 方案

在 ASIACRYPT 2017 中, Cheon 等人^[13]提出了一种适用于近似计算的层次同态加密方案 CKKS, 以及用于实现该方案的开源库 HEAAN。该方案流程如下:

参数初始化: $d = 2^M$, $R = \mathbb{Z}[x]/\langle x^d + 1 \rangle$。对于基 p, 模数 q_0, 自然数 L (预设层数), 令 $q_l = p^l \cdot q_0 (l = 1, \dots, L)$。注意到第 l 层的密文是 R_{q_l} 中的向量。对实数 σ, $DG(\sigma^2)$ 为 $\mathbb{Z}^d$ 上的离散高斯分布。对实数 ρ, 为 $\{-1, 0, 1\}^d$ 上, 以 $1-\rho$ 的概率选取 0, 以 $\rho/2$ 的概率选取 -1 或 1 的分布。X 是以 B 为上界的分布
密钥生成: 输入安全参数 λ, 整数 t。随机匀生成 $a \in R_{q_L}$, 计算 $b = -as + e \bmod q_L$ (其中 $e \in DG(\sigma^2), s \in \mathcal{X}$)。最后, 选取 $a' \in R_{q_L}$, $e' \in DG(\sigma^2)$, 计算 $b' = -as + e' + ts' \bmod t \cdot q_L$。 输出私钥 $sk = (1, s)$、公钥 $pk = (b, a)$、计算密钥 $evk = (b', a')$
加密: 输入公钥 $pk \in R_{q_L}^2$ 和消息 $m \in R$。随机选取值 $v \in ZO(1/2)$ 和 $e_0, e_1 \in DG(\sigma^2)$。 输出 $c = (\beta, \alpha) = vpk + (m + e_0, e_1) \bmod q_L$
解密: 输入私钥 $sk = (1, s)$ 和密文 $c \in R_{q_L}^2$。 输出 $m = \langle c, sk \rangle \bmod q_L = \beta + \alpha s \bmod q_L$
同态加法: 即 $c_1 + c_2$
同态乘法: 对于两个密文 $c_i (i = 1, 2)$ 的乘法为 $c_1 \cdot c_2 = (d_0, d_1) + \lfloor t^{-1} d_2 evk \rfloor \bmod q_L$, 其中 $(d_0, d_1, d_2) = (\beta_1 \beta_2, \alpha_1 \beta_2 + \alpha_2 \beta_1, \alpha_1 \alpha_2) \bmod q_L$。当两个密文 c 和 c' 处于不同层级 $l' < l$ 时, 需要先将层数高的密文层数下降到相同层级, 即 $l' < l$。该过程可通过重缩放 (Rescaling) 过程实现, 即输入第 l 层的密文 c, 输出第 l' 层的密文 $c' = \lfloor q_{l'} / q_l \rfloor c \bmod q_{l'}$

在 EUROCRYPT 2018 中, Cheon 等人^[35]又提出了 CKKS 自举算法, 进而将 CKKS 扩展为全同态方案。该方案通过线性变换实现系数-槽之间的转换, 并进行模约化 (Modular Reduction), 从而将最低层的密文更新为层数更高的能够继续进行计算的密文。Cheon 等人之后又引入了基于 CRT 的密文打包 (Packing) 技术, 提出了 CKKS 的变种^[36]。Boemer 等人^[37]优化了标量编码、密文-明文加法和乘法的运行时间。Kim 等人^[38]提出可以在乘法前对密文进

行重缩放,从而在计算过程中将误差最小化的新技术,提升了 CKKS 及其 RNS 类型方案的实用性。

CKKS 的自举算法也在不断改进。为了优化线性变换的时间复杂度,在 EUROCRYPT 2019 中,Chen 等人^[39]在原始 CKKS 自举算法^[13]的基础上,首次发现了自举过程中涉及的线性变换适用于快速傅里叶变换算法,并通过动态规划求解出最优的层级消耗,从而使得自举方案能够处理的槽数量比原始方案多 2~128 倍;Han 和 Ki^[40]则在文献[36]上做出了改进,提出了 Full-RNS 类型 CKKS 的一般化密钥交换 (Generalized Key-Switching) 技术。另外,由于同态方案仅支持加法和乘法,因此模约化阶段需要同态的计算模函数的近似多项式。为了优化逼近的精度和复杂度,Lee 等人在 EUROCRYPT 2021 中^[41]提出最小化近似误差的无穷范数的思想,通过多区间 Remez 算法求解出近似多项式;原始方案为了避免层级消耗,会使用正弦函数取代模函数作为逼近对象,Jutla 和 Manohar^[42]提出了不依赖于三角函数近似计算的方法改进 CKKS 自举效率,直接对模函数进行模拉格朗日插值 (Modular Lagrange Interpolation),从而避免了三角函数和模函数固有存在的基础误差 (Fundamental Error);而后,Jutla 和 Manohar^[43]又在 EUROCRYPT 2022 中提出了使用正弦级数逼近模函数的方法,极大程度提高了自举精度;在 ASIACRYPT 2022 中,Kim 等人^[44]又将近似模运算 (Evalmod) 替换为近似取整运算 (Evalround),使得自举算法可以选择更小的缩放因子 (Scaling Factor),节省了 1~2 层的层级消耗。

近些年,CKKS 方案的安全性评估问题也受到关注。在 EUROCRYPT 2021 中,Li 和 Micciancio^[45]提出了针对 CKKS 的密钥恢复攻击,这一攻击利用了 CKKS 的近似运算特性,并在本文提出了 IND-CPAD 模型。而后,在 CRYPTO 2022 中,Li 和 Micciancio 等人^[46]利用差分隐私技术精确分析了防御噪声大小对安全性的影响。

6 GSW 方案

2013 年,Gentry、Sahai、Waters 提出了 GSW 方案^[14]。该方案引入了近似特征值方法 (Approximate Eigenvector Method),通过矩阵运算实现密文同态操作。由于密文加法、乘法对应方阵的加法、乘法,密文维度始终维持在常值不变,因此不再需要使用密钥交换技术。另外,GSW 方案引入了压缩技术 (Flatten Technique,即 FL 操作),因此也不再需要使用模交换技术。这种新方法将同态乘法引入的噪声增长降到小多项式因子级别。

为了解决同态乘法中引入过大噪声导致无法解密的问题,该方案还引入了比特分解 (Bit Decomposition,即 BD 操作,相应的有 PowerofTwo 操作,记为 PT)。该方案流程如下:

密钥生成:

输入安全参数 λ , 乘法深度 L , 密文维数 $n = n(\lambda, L)$, 密文模数 $q = q(\lambda, L)$, 噪声分布 $\chi = \chi(\lambda, L)$ 。令 $m = O(n \log_2 q)$, $l = \lceil \log_2 q \rceil + 1$, $N = (n+1)l$, 参数集 $\text{params} = (n, q, \chi, m)$ 。选取 $t \leftarrow Z_q^n$, $B \leftarrow Z_q^{m \times n}$, $e \leftarrow \chi^m$, 令 $s = (1, -t) \in Z_q^{n+1}$,

$b = B \cdot t + e$, $A = [b \parallel B]$

输出私钥 $sk = v = PT(s)$, 公钥 $pk = A$

加密:

输入明文 $\mu \in Z_q$, 选取矩阵 $R \leftarrow \{0, 1\}^{N \times m}$, N 维单位矩阵 I_N

输出密文 $C = FL(\mu \cdot I_N + BD(R \cdot A)) \in Z_q^{N \times N}$

同态 NAND:

输入明文 (μ_1, μ_2) 对应的加解密文对 $(C_1, C_2) \in Z_q^{N \times N}$, 其中相应的密文噪声为 e_1 和 e_2

输出密文 $C_{\text{NAND}} = FL(I_N - C_1 C_2)$

作为同态 NAND 运算的结果, C_{NAND} 满足以下性质 $C_{\text{NAND}} \cdot v = (1 - \mu_1 \mu_2)v - \mu_2 e_1 - C_1 e_2$

然而, 较高的交互成本 (该方案的密文长度相比明文长度大许多) 和计算复杂度是 GSW 方案的主要缺点。为了降低计算开销, 后续出现了许多改进 GSW 方案自举过程的工作。Alperin-Sheriff 和 Peikert^[47-48]提出了一种新的自举算法, 不再将解密视为布尔电路, 而是看作算术函数。Hiromasa 等人^[49]优化了上述 AP 方案^[47], 基于 GSW 方案设计了支持对矩阵进行加密并同态进行矩阵运算操作的全同态方案。另外, Brakerski 和 Vaikunthanathan^[50]从 GSW 方案出发, 提出了第一个基于多项式近似因子 GapSVP 的全同态方案。Ducas 和 Micciancio 进一步提出了文献^[47]的 RLWE 版本, 通过自举过程的同时进行查表, 引入了同态计算两个密文 NAND 的新方法, 后续该工作被称为 FHEW 方案。

7 FHEW/TFHE 方案

在 EUROCRYPT 2015 中, Ducas 和 Micciancio 提出了 FHEW 方案^[51]。该方案基于 AP 自举框架^[47], 使用基于 Ring-GSW 密文设计的同态累加器实现 LWE 密文的自举。FHEW 的明文消息为 1bit, 每次自举仅计算一个 NAND 门, 首次将全同态加密方案的自举时间降到 1s 以内。

参数初始化: LWE 的维度和模数为 n, q , RLWE 的维度和模数为 N, Q , 其中 N 为 2 的幂, $q|2N$, $R_Q = Z_Q[X]/\langle X^N + 1 \rangle$ 。还需要选择一个分解基 $B \in Z$

密钥生成: 输入安全参数 λ , 输出 RLWE 密钥 $s \in R$, LWE 密钥 $s \in Z_q^n$

加密: 输入私钥 $s \in Z_q^n$, 消息 $m \in \{0,1\}$, 均匀选取 $a \in Z_q^n$ 和小噪声 $e \in Z_q$, 输出密文 $c = \left(a, \langle a, s \rangle + e + \lfloor \frac{q}{4} \rfloor m \right) \in Z_q^{n+1}$

解密: 输入密钥 $s \in Z_q^n$ 和密文 $c \in Z_q^{n+1}$, 计算 $\lfloor \frac{4}{q} \langle (-s \| 1), c \rangle \rfloor = \lfloor \frac{4}{q} \left(e + \lfloor \frac{q}{4} \rfloor m \right) \rfloor = m$

同态 NAND:

输入明文 $(\mu_1, \mu_2) \in \{0,1\}^2$ 对应的 LWE 密文对 $(C_1, C_2) \in Z_q^{n+1} \times Z_q^{n+1}$, 计算 $C = C_1 + C_2 = (a, b) \in Z_q^n \times Z_q$ 。可知 C 加密了 $\mu = \mu_1 + \mu_2 \in \{0,1,2\} \bmod 4$, 只要能够对 C 同态地计算映射 $f: Z_4 \rightarrow Z_4$, $f(0) = f(1) = 1$, $f(2) = 0$, 即可完成 NAND 的计算。

FHEW 通过同态地计算一个加密了 $\lfloor \frac{Q}{8} \rfloor (1 + X + X^2 + \dots + X^{N-1}) \cdot X^{\frac{2N}{q} \left(b - \langle a, s \rangle + \frac{q}{8} \right)}$ 的 RLWE 累加器 ACC, 并

提取该多项式的常数项来实现上述 f 的计算, 可以验证当 $\mu \in \{0,1\}$ 时, $\frac{2N}{q} \left(b - \langle a, s \rangle + \frac{q}{8} \right) \in [0, N-1]$, 因此上式结

果为 0; 当 $\mu \in \{2,3\}$ 时, $\frac{2N}{q} \left(b - \langle a, s \rangle + \frac{q}{8} \right) \in [N, 2N-1]$, 上式结果约为 $\lfloor \frac{Q}{4} \rfloor$, 即 $1 \in Z_4$ 在 Z_Q 中编码后的结果。最

后, 通过密钥转换和模数转换将其转化到 Z_q^{n+1} 上的 LWE 密文。

特别地, NAND 计算结果的噪声与输入噪声无关, 即该过程同时起到自举的作用。

ACC 的计算方法: 初始化 $ACC = \lfloor \frac{Q}{8} \rfloor (1 + X + X^2 + \dots + X^{N-1}) \cdot X^{\frac{2N}{q} \left(b + \frac{q}{8} \right)}$, 随后分 n 次同态地将 ACC 乘以 $X^{\frac{2N}{q} a_{i, S_i}}$ 。

这一步首先对 a_i 进行 B 进制分解, 即 $a_i = \sum_{j=0}^{l-1} a_{i,j} B^j$ ($l = \lceil \log_{2B} q \rceil$), 随后选择用 RGSW 加密的 $X^{\frac{2N}{q} a_{i,j} S_i}$ 与 ACC

相乘。RGSW 密文与 RLWE 密文的乘法称为内积 (External Product)。

之后, Chillotti 等人^[52]对 FHEW 进行了改进, 包括: ①采用 GINX 自举框架, 在将 ACC 乘以 $X^{\frac{2N}{q} a_{i, S_i}}$ 时, 不再对 a_i 进行分解, 而是对 s_i 进行分解, 同时使用 $\{0,1\}$ 分布的密钥, 减少所需外积次数和公钥大小; ②将 R_Q 上的运算转换到 $T[X] = R[X]/(X^N + 1, Z)$ 上, 从而用 FFT 代替 NTT。改进后的方案被称为 TFHE, 其在商用计算机上的单次自举仅需 10ms。不过, 由于二元密钥的安全性尚未得到充分研究, 因此一般推荐使用三元密钥^[53]。在使用三元密钥时, 自举的 NTT 次数与使用二元密钥时相同, 但是噪声增长和公钥大小翻倍^[54]。

在改进 FHEW/TFHE 自举效率方面,一方面 Lee 等人利用环同构操作同态地计算 $X^c \rightarrow X^{a_i c}$ ($a_i \in \mathbb{Z}_q^*$), 以此实现 ACC 的更新^[55]。这一方法克服了 FHEW/AP 效率较低的问题,也克服了 TFHE/GINX 中要求密钥分布较小(二元或三元分布,无法使用离散高斯分布)的安全性问题,同时大幅减小了公钥大小。另一方面 Bonte 等人^[56]和 Kluczniak^[57]同时提出了基于 NTRU 而非 RLWE 构造的 TFHE。由于 NTRU 密文仅由一个 R_Q 多项式构成,其所需的外积次数(和 NTT 次数)约为 RLWE 版本的 1/4,因此自举效率得到了提升。2023 年美密会上, Xiang 等人将 NTRU 和环同构这两条路径结合,获得了目前自举效率最高的 FHEW/TFHE 方案^[58]。

此外, Bonnoron 发现通过将 ACC 的初始值替换为其他多项式,可以在自举中计算一个查找表(Lookup Table, LUT)^[59], 而 NAND 门的计算可以看作是一个大小为 4 的查找表。这一利用自举计算 LUT 的方式也被称为函数式自举(Functional Bootstrap)或可编程自举(Programmable Bootstrap)。基于函数式自举, Bourse 等人^[60]和 Chillotti 等人^[61]设计了低延时的同态神经网络计算方法。此外, 全域自举、多值自举、高精度 LUT 查找等技术的出现,进一步丰富了函数式自举的功能与可用性^[62-65]。Ma 等人^[66]给出了 4 种全域自举算法和两种同态位分解算法,并将其方法通过同态方案转换方法应用在连续函数的函数自举场景下。而在 EUROCRYPT 2023 中, Liu 和 Wang^[67-68]提出了适用于三代同态方案的批处理自举方法,该方法在多项式模数下的均摊开销仅需要 $\tilde{O}(1)$ 次全同态乘法。

参考文献

- [1] RIVEST R L, ADLEMAN L M, DERTOUZOS M L. On Data Banks and Privacy Homomorphisms[J]. Foundations of Secure Computation, 1978.
- [2] GOLDWASSER S, MICALI S. Probabilistic Encryption and How to Play Mental Poker Keeping Secret All Partial Information[C]//Proceedings of the 14th Annual ACM Symposium on Theory of Computing, May 5-7, 1982, San Francisco, California, USA. ACM, 1982.
- [3] GAMAL T E. A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms[C]//Advances in Cryptology, Proceedings of CRYPTO'84, Santa Barbara, California, USA, August 19-22, 1984, Proceedings, 1984.
- [4] PAILLIER P. Public-key cryptosystems based on composite degree residuosity classes[J]. Proc. EUROCRYPT'99, Czech Republic, May, 1999.

[5] BONEH D, GOH E J, NISSIM K. Evaluating 2-DNF formulas on ciphertexts[C]//Theory of Cryptography: Second Theory of Cryptography Conference, TCC 2005, Cambridge, MA, USA, February 10-12, 2005. Proceedings 2. Springer Berlin Heidelberg, 2005: 325-341.

[6] GENTRY C. A fully homomorphic encryption scheme[D]. Stanford university, 2009.

[7] DUCAS L, MICCIANCIO D. FHEW: bootstrapping homomorphic encryption in less than a second. In Elisabeth Oswald and Marc Fischlin, editors, Advances in Cryptology-EUROCRYPT 2015-34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part I, volume 9056 of Lecture Notes in Computer Science, pages 617-640. Springer, 2015.

[8] CHILLOTTI I, GAMA N, GEORGIEVA M, et al. Faster fully homomorphic encryption: Bootstrapping in less than 0.1 seconds. In Jung Hee Cheon and Tsuyoshi Takagi, editors, Advances in Cryptology - ASIACRYPT 2016 - 22nd International Conference on the Theory and Application of Cryptology and Information Security, Hanoi, Vietnam, December 4-8, 2016, Proceedings, Part I, volume 10031 of Lecture Notes in Computer Science, pages 3-33, 2016.

[9] CHILLOTTI I, GAMA N, GEORGIEVA M, et al. Faster packed homomorphic operations and efficient circuit bootstrapping for TFHE. In Tsuyoshi Takagi and Thomas Peyrin, editors, Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part I, volume 10624 of Lecture Notes in Computer Science, pages 377-408. Springer, 2017.

[10] BRAKERSKI Z, VAIKUNTANATHAN V. Efficient fully homomorphic encryption from (standard) LWE. In Rafail Ostrovsky, editor, IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011, pages 97-106. IEEE Computer Society, 2011.

[11] BRAKERSKI Z, GENTRY C, VAIKUNTANATHAN V. (leveled) fully homomorphic encryption without bootstrapping. In Shafi Goldwasser, editor, Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012, pages 309-325. ACM, 2012.

[12] FAN J F, Vercauteren F. Somewhat practical fully homomorphic encryption. IACR Cryptol. ePrint Arch., 2012:144, 2012.

[13] CHEON J H, KIM A, KIM M, et al. Homomorphic encryption for arithmetic of approximate numbers. In Tsuyoshi Takagi and Thomas Peyrin, editors, Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part I, volume

10624 of Lecture Notes in Computer Science, pages 409-437. Springer, 2017.

[14] GENTRY C, SAHAI A, WATERS B. Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference*, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I, volume 8042 of Lecture Notes in Computer Science, pages 75-92. Springer, 2013.

[15] REGEV O. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM*, 56(6):34, 2009. Preliminary version in STOC'05.

[16] LYUBASHEVSKY V, PEIKERT C, REGEV O. On ideal lattices and learning with errors over rings[C]//*Advances in Cryptology-EUROCRYPT 2010: 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, French Riviera, May 30-June 3, 2010. Proceedings 29. Springer Berlin Heidelberg, 2010: 1-23.

[17] HOFFSTEIN J, PIPHER J, SILVERMAN J H. NTRU: A ring-based public key cryptosystem[C]//*International algorithmic number theory symposium*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998: 267-288.

[18] BRAKERSKI Z, VAIKUNTANATHAN V. Fully homomorphic encryption from ring-LWE and security for key dependent messages. in *Advances in Cryptology-CRYPTO 2011*, P. Rogaway, Ed. Berlin, Germany: Springer, 2011, pp. 505-524.

[19] HALEVI S, SHOUP V. HELib. Accessed: Sep.23,2022. [Online]. Available: <https://github.com/homenc/HElib>.

[20] GENTRY C, HALEVI S, SMART N P. Homomorphic evaluation of the AES circuit. in *Advances in Cryptology-CRYPTO 2012*, R. Safavi-Naini and R. Canetti, Eds. Berlin, Germany: Springer, 2012, pp. 850-867.

[21] GENTRY C, HALEVI S, SMART N P. Fully homomorphic encryption with polylog overhead. in *Advances in Cryptology-EUROCRYPT 2012*, D. Pointcheval and T. Johansson, Eds. Berlin, Germany: Springer, 2012, pp. 465-482.

[22] SMART N P, VERCAUTEREN F. Fully homomorphic encryption with relatively small key and ciphertext sizes. in *Proc. Int. Workshop Public Key Cryptogr.* Cham, Switzerland: Springer, 2010, pp. 420-443.

[23] BRAKERSKI Z. Fully homomorphic encryption without modulus switching from classical GapSVP. in *Advances in Cryptology-CRYPTO 2012*, R. Safavi-Naini and R. Canetti, Eds. Berlin, Germany: Springer, 2012, pp. 868-886.

[24] FAN J, VERCAUTEREN F. Somewhat practical fully homomorphic encryption. *Cryptol.*

ePrint Arch., Paper 2012/144, 2012. [Online]. Available: <https://eprint.iacr.org/2012/144>.

[25] BAJARD J C, EYNARD J, Hasan M A, et al. A full RNS variant of FV like somewhat homomorphic encryption schemes. in Proc. Int. Conf. Sel. Areas Cryptogr. Cham, Switzerland: Springer, 2016, pp. 423-442.

[26] HALEVI S, POLYAKOV Y, SHOUP V. An improved RNS variant of the BFV homomorphic encryption scheme. in Proc. Cryptographers' Track RSA Conf. Cham, Switzerland: Springer, 2019, pp. 83-105.

[27] AL BADAWI A, POLYAKOV Y, AUNG K M M, et al. Implementation and performance evaluation of RNS variants of the BFV homomorphic encryption scheme. IEEE Trans. Emerg. Topics Comput., vol. 9, no. 2, pp. 941-956, Apr, 2021.

[28] BAJARD J C, EYNARD J, MARTINS P, et al. Note on the noise growth of the RNS variants of the BFV scheme. Cryptol. ePrint Arch, 2019.

[29] CHEN H, HAN K. Homomorphic lower digits removal and improved FHE bootstrapping. in Advances in Cryptology-EUROCRYPT 2018, J. B. Nielsen and V. Rijmen, Eds. Cham, Switzerland: Springer, 2018, pp. 315-337.

[30] HALEVI S, SHOUP V. Faster homomorphic linear transformations in HELib. in Proc. Annu. Int. Cryptol. Conf. Cham, Switzerland: Springer, 2018, pp. 93-120.

[31] CHEN H, LAINE K, PLAYER R, et al. High-precision arithmetic in homomorphic encryption. in Proc. Cryptographers' Track RSA Conf. Cham, Switzerland: Springer, 2018, pp. 116-136.

[32] HOFFSTEIN J, SILVERMAN J. Optimizations for NTRU. in Proc. Public-Key Cryptogr. Comput. Number Theory, De Gruyter Math., 2000, pp. 77-88.

[33] COSTACHE A, LAINE K, PLAYER R. Evaluating the effectiveness of heuristic worst-case noise analysis in FHE. in Proc. Eur. Symp. Res. Comput. Secur. Cham, Switzerland: Springer, 2020, pp. 546-565.

[34] KIM A, POLYAKOV Y, ZUCCA V. Revisiting homomorphic encryption schemes for finite fields. in Advances in Cryptology-ASIACRYPT, vol. 13092, M. Tibouchi and H. Wang, Eds. Springer, 2021, pp. 608-639.

[35] CHEON J H, HAN K, KIM A, et al. Bootstrapping for approximate homomorphic encryption. in Advances in Cryptology-EUROCRYPT 2018, J. B. Nielsen and V. Rijmen, Eds. Cham, Switzerland: Springer, 2018, pp. 360-384.

[36] CHEON J H, HAN K, KIM A, et al. A full RNS variant of approximate homomorphic encryption. in Proc. Int. Conf. Sel. Areas Cryptogr. Cham, Switzerland: Springer, 2018, pp. 347-368.

[37] BOEMER F, COSTACHE A, CAMMAROTA R, et al. NGraph-HE2: A high-throughput framework for neural network inference on encrypted data. in Proc. 7th ACM Workshop Encrypted Comput. Appl. Homomorphic Cryptogr., New York, NY, USA, 2019, pp. 45-56.

[38] KIM A, PAPADIMITRIOU A, POLYAKOV Y. Approximate homomorphic encryption with reduced approximation error. in Proc. Cryptographers' Track RSA Conf. Cham, Switzerland: Springer, 2022, pp. 120-144.

[39] CHEN H, CHILLOTTI I, SONG Y. Improved bootstrapping for approximate homomorphic encryption. in Advances in Cryptology-EUROCRYPT 2019, Y. Ishai and V. Rijmen, Eds. Cham, Switzerland: Springer, 2019, pp. 34-54.

[40] HAN K, KI D. Better bootstrapping for approximate homomorphic encryption. in Topics in Cryptology-CT-RSA 2020, S. Jarecki, Ed. Cham, Switzerland: Springer, 2020, pp. 364-390.

[41] LEE J W, LEE E, LEE Y, et al. High-precision bootstrapping of RNS-CKKS homomorphic encryption using optimal minimax polynomial approximation and inverse sine function. in Advances in Cryptology-EUROCRYPT 2021. Cham, Switzerland: Springer, 2021, pp. 618-647.

[42] JUTLA C S, MANOHAR N. Modular Lagrange interpolation of the mod function for bootstrapping of approximate HE. Cryptol. ePrint Arch., 2020.

[43] JUTLA C S, MANOHAR N. Sine series approximation of the mod function for bootstrapping of approximate HE. in Advances in Cryptology-EUROCRYPT 2022. Cham, Switzerland: Springer, 2022, pp. 491-520.

[44] KIM S, PARK M, KIM J, et al. (2022, December). EvalRound Algorithm in CKKS Bootstrapping. In International Conference on the Theory and Application of Cryptology and Information Security (pp. 161-187). Cham: Springer Nature Switzerland.

[45] LI B, MICCIANCIO D. (2021). On the security of homomorphic encryption on approximate numbers. In Advances in Cryptology-EUROCRYPT 2021: 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, October 17-21, 2021, Proceedings, Part I 40 (pp. 647-677). Springer International Publishing.

[46] LI B, MICCIANCIO D, SCHULTZ M, et al. (2022, August). Securing approximate homomorphic encryption using differential privacy. In Annual International Cryptology Conference (pp. 560-589). Cham: Springer Nature Switzerland.

[47] ALPERIN-SHERIFF J, PEIKERT C. Faster bootstrapping with polynomial error. in Advances in Cryptology-CRYPTO 2014, J. A. Garay and R. Gennaro, Eds. Berlin, Germany: Springer, 2014, pp. 297-314.

[48] ALPERIN-SHERIFF J, PEIKERT C. Practical bootstrapping in quasilinear time. in Advances in Cryptology-CRYPTO 2013, R. Canetti and J. A. Garay, Eds. Berlin, Germany: Springer, 2013, pp. 1-20.

[49] HIROMASA R, ABE M, OKAMOTO T. Packing messages and optimizing bootstrapping in GSW-FHE. IEICE Trans. Fundamentals Electron., Commun. Comput. Sci., vol. E99.A, no. 1, pp. 73-82, 2016.

[50] BRAKERSKI Z, VAIKUNTANATHAN V. Lattice-based FHE as secure as PKE. in Proc. 5th Conf. Innov. Theor. Comput. Sci., Jan. 2014, pp. 1-12.

[51] DUCAS L, MICCIANCIO D. FHEW: Bootstrapping homomorphic encryption in less than a second. in Advances in Cryptology-EUROCRYPT 2015, E. Oswald and M. Fischlin, Eds. Berlin, Germany: Springer, 2015, pp. 617-640.

[52] CHILLOTTI I, GAMA N, GEORGIEVA M. TFHE: Fast fully homomorphic encryption over the torus. J. Cryptol., vol. 33, no. 1, pp. 1-58, 2019.

[53] MICCIANCIO D, POLYAKOV Y. Bootstrapping in FHEW-like cryptosystems. in Proc. 9th Workshop Encrypted Comput. Appl. Homomorphic Cryptogr., Nov. 2021, pp. 17-28.

[54] BONTE C, ILIASHENKO I, PARK J. et al. (2022, December). Final: Faster fhe instantiated with ntru and lwe. In International Conference on the Theory and Application of Cryptology and Information Security (pp. 188-215). Cham: Springer Nature Switzerland.

[55] LEE Y, et al. Efficient FHEW bootstrapping with small evaluation keys, and applications to threshold homomorphic encryption. Cryptol. ePrint Arch., 2022.

[56] BONTE C, ILIASHENKO I, PARK J, et al. (2022, December). Final: Faster fhe instantiated with ntru and lwe. In International Conference on the Theory and Application of Cryptology and Information Security (pp. 188-215). Cham: Springer Nature Switzerland.

[57] KLUCZNIAK K.(2022, November). NTRU-v-Um: Secure Fully Homomorphic Encryption from NTRU with Small Modulus. In Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (pp.1783-1797).

[58] XIANG B, ZHANG J, DENG Y, et al. (2023, August). Fast Blind Rotation for Bootstrapping FHEs. In Annual International Cryptology Conference (pp. 3-36). Cham: Springer Nature Switzerland.

[59] BONNORON G, DUCAS L, DILLINGER M. (2018). Large FHE Gates from Tensored Homomorphic Accumulator. In Progress in Cryptology-AFRICACRYPT 2018: 10th International Conference on Cryptology in Africa, Marrakesh, Morocco, May 7-9, 2018, Proceedings 10 (pp.217-251). Springer International Publishing.

[60] BOURSE F, MINELLI M, MINIHOLD M, et al. (2018). Fast Homomorphic Evaluation of Deep Discretized Neural Networks. In *Advances in Cryptology-CRYPTO 2018: 38th Annual International Cryptology Conference*, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part III 38 (pp. 483-512). Springer International Publishing.

[61] CHILLOTTI I, JOYE M, PAILLIER P. (2021). Programmable Bootstrapping Enables Efficient Homomorphic Inference of Deep Neural Networks. In *Cyber Security Cryptography and Machine Learning: 5th International Symposium, CSCML 2021, Be'er Sheva, Israel, July 8-9, 2021, Proceedings 5* (pp. 1-19). Springer International Publishing.

[62] GUIMARÃES A, BORIN E, ARANHA D F. Revisiting the functional bootstrap in TFHE. *IACR Trans. Cryptograph. Hardw. Embedded Syst.*, vol. 23, pp. 229-253, Feb. 2021.

[63] CARPOV S, IZABACHÈNE M, MOLLIMARD V. (2019). New Techniques for Multi-value Input Homomorphic Evaluation and Applications. In *Topics in Cryptology-CT-RSA 2019: The Cryptographers' Track at the RSA Conference 2019*, San Francisco, CA, USA, March 4-8, 2019, Proceedings (pp.106-126). Springer International Publishing.

[64] LIU Z, MICCIANCIO D, POLYAKOV Y. (2022, December). Large-Precision Homomorphic Sign Evaluation Using FHEW/TFHE Bootstrapping. In *International Conference on the Theory and Application of Cryptology and Information Security* (pp.130-160). Cham: Springer Nature Switzerland.

[65] LEE Y, et al. Efficient FHEW bootstrapping with small evaluation keys, and applications to threshold homomorphic encryption. *Cryptol. ePrint Arch.*, 2022.

[66] MA S, HUANG T, WANG A, et al. (2023). Fast and Accurate: Efficient Full-Domain Functional Bootstrap and Digit Decomposition for Homomorphic Computation. *Cryptology ePrint Archive*.

[67] LIU F H, WANG H. (2023, April). Batch bootstrapping I: a new framework for simd bootstrapping in polynomial modulus. In *Annual International Conference on the Theory and Applications of Cryptographics Techniques* (pp. 321-352). Cham: Springer Nature Switzerland.

[68] LIU F H, WANG H. (2023, April). Batch bootstrapping II: bootstrapping in polynomial modulus only requires $O(1)$ the multiplications in amortization. In *Annual International Conference on the Theory and Applications of Cryptographics Techniques* (pp. 321-352). Cham: Springer Nature Switzerland.

全同态加密算法实现技术综述^①

韦薇¹, 彭丛蓉¹, 徐润擎¹, 陈宇田¹, 曾陈铠¹, 彭聪¹, 何德彪¹

1. 武汉大学国家网络安全学院空天信息安全与可信计算教育部重点实验室, 武汉, 430072

通讯作者: 何德彪, E-mail: hedebiao@163.com

摘要: 全同态加密在理论上经历了重大的发展和不断的突破, 使其在外包计算和安全多方计算等各个领域得到广泛的应用, 保护了用户的数据隐私。然而, 由于计算开销大、存储成本高, 全同态加密成为制约实际系统性能的瓶颈。研究人员已经提出了多种实用的加速方案来解决这些问题。本文旨在提供一个全面的优化方案调研, 比较和分析全同态加密优化方案的优势与劣势。本文首先从同态加密的通用优化技巧入手, 深入阐述全同态加密的基本算子与通用优化技巧, 然后系统讨论不同平台下的最新优化方案, 在此基础上综合分析各类方案的运算速度, 为该领域的理论研究与实际应用提供支持。

关键词: 全同态加密; 优化实现; 隐私保护

A Survey on Implementation Techniques of Fully Homomorphic Encryption

WEI Wei¹, PENG Congrong¹, XU Runqing¹, CHEN Yutian¹, ZENG Chenkai¹, PENG Cong¹, HE Debiao¹

1. Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, 430072, China

Corresponding author: HE Debiao, E-mail: hedebiao@163.com

Abstract: Fully homomorphic encryption has experienced significant development and continuous breakthroughs in theory, which makes it widely used in various fields such as outsourcing

^① 基金项目: 山东省重点研发计划 (2021CXGC010107); 国家自然科学基金 (61932016, U21A20466)。

computing and secure multi-party computing to protect user privacy data. However, due to the high computational overhead and storage cost, fully homomorphic encryption becomes the bottleneck of the actual system performance. Researchers have proposed a variety of practical acceleration schemes to solve these problems. This paper aims to provide a comprehensive optimization scheme survey, in which the advantages and weaknesses of all homomorphic encryption optimization schemes are compared and analyzed. First of all, this paper starts with the general optimization techniques of homomorphic encryption and elaborates the basic operators and general optimization techniques of homomorphic encryption. Then this paper systematically discusses the latest optimization schemes under different platforms. On this basis, the performance of various schemes is comprehensively analyzed to provide support for theoretical research and practical application.

Keywords: Fully homomorphic encryption; Optimization implementation; Privacy protection

1 全同态加密高速实现概述

近年来,云计算技术及其衍生的云存储技术一直备受关注,各种提供云服务的平台如雨后春笋般涌现。通过云计算和云存储技术,海量数据的复杂计算和存储等任务可以外包给云服务提供商,实现资源共享并降低运算成本。然而,在确保数据安全和用户隐私的前提下进行安全计算是一个亟待解决的问题。同态加密技术允许人们在加密的数据上直接执行操作,如检索和比较,得出正确的结果,而无须在处理过程中解密数据。从根本上解决了将数据和操作委托给第三方时的保密性问题。

全同态加密(Fully Homomorphic Encryption, FHE)允许执行任意次的同态加法和同态乘法操作。许多密码学家将FHE的思想视为与公钥加密思想一样重要。FHE在多个领域具有广泛的潜在应用前景,包括基因组学、健康、国家安全和教育等。自2009年Gentry提出第一个FHE方案以来,全同态加密取得了长足的发展,在理论上经历了不断突破。FHE应用的主要限制是其性能瓶颈,即FHE的巨大成本不能满足实际应用的需求。目前主要有两种加速方式:第一种加速方式专注于算法优化,以加速FHE方案;另一种加速方式是研究在特定平台下加速全同态加密算法,如中央处理器(CPU)、图形处理器(GPU)、现场可编程门阵列(FPGA)等。并行计算策略也被广泛应用,通过将计算任务分解为多个子任务并行执行,从而提高了计算速度。

1. 基于CPU的高速实现

与GPU、FPGA等相比,CPU实现的加速效果并不明显。CPU本身并不是为执行繁重的数据计算任务而设计的。它可以通过多线程提供一定程度的并行计算能力,但这远不能满足

FHE 加速的需求。因此,大多数研究倾向于将其与 GPU 或 FPGA 相结合来实现 FHE 的加速。但是,CPU 的优势在于它是一个更通用的硬件平台,应用范围广泛。特别地,CPU 平台下的并行指令集能进一步提升同态加密的计算效率。

2. 基于 GPU 的高速实现

GPU 所能提供的定制灵活性介于 CPU 和 FPGA 之间。与 FPGA 和 ASIC 相比,GPU 上的通用计算在价格标准化时具有更高的效率。GPU 是一种功能强大但高度专业化的设备,需要仔细编码才能充分利用其提供的大量并行性。具体来说,编程模型和内存组织与 CPU 完全不同。与 FPGA 相比,它更容易开发程序。此外,还有一个计算统一设备架构(CUDA)工具包,以促进程序开发。

3. 基于硬件平台的高速实现

FPGA 是一种可重构电路的芯片。它是一个硬件可重构的体系结构,它可以提供更多的定制灵活性。CPU 和 GPU 都属于冯·诺依曼结构,而 FPGA 是一种无指令、无共享存储器的结构。这种结构使 FPGA 比 CPU 甚至 GPU 更节能。FPGA 中每个逻辑单元的功能在重新编程期间确定,而不需要指令、FPGA 具有流水线和数据并行性,而 GPU 仅具有数据并行性(流水线深度有限)。由于 FPGA 的灵活性,越来越多的研究者关注于利用 FPGA 来加速 FHE。ASIC 是一种专用芯片,不同于传统的通用芯片,ASIC 面向特定需求而专门定制的芯片。ASIC 芯片的计算能力和效率可以根据算法需求进行定制,因此与通用芯片相比,ASIC 具有以下几个显著优势:小型化、低功耗、高计算性能、卓越计算效率,以及大规模生产时成本更低。与 FPGA 相比,ASIC 能耗更低、计算效率更高,因此针对 ASIC 的同态加密算法定制化实现也是一项重要的研究领域。

2 全同态加密的通用优化技巧

全同态加密算法的优化技巧主要分为三大类,包括大整数运算、编码与多项式运算。通常,基于算法加速的 FHE 方案的加速效果是有限的。许多关于 FHE 加速的研究工作主要集中在 NTT 加速和自举。Barrett 模约减最常用于加速 NTT。算法优化主要有两种方式:一是用消耗高计算资源的操作替换需要高存储资源消耗的操作;二是提高算法的并行性,以便在此基础上设计硬件加速方案。一般来说,算法加速方案都无法达到突破性的加速效果,大多数作品只是应用已有的算法。因此,基于算法的加速方案需要理论上的突破,特别是对于自举,以便使 FHE 性能更接近实际应用需求。而基于算法优化的加速方案的优点是适应性更强,可以适应不同的平台。

2.1 大整数运算优化技巧

2.1.1 Karatsuba 乘法

1950 年, Kolmogorov 提出了一个猜想: 长度为 n 的整数乘法的复杂度最低为 n^2 的数量级。然而, 1960 年, Karatsuba 发现了一个复杂度 $O(n^{\log_2 3})$ 的算法^[1]。

在该算法中, 设 $R = b^n$, $d = 2n$, $u = (u_{d-1} \cdots u_0)_b$ 和 $v = (v_{d-1} \cdots v_0)_b$ 为两个 d 字节整数。将 u 和 v 分为高位和低位两个部分, 即 $u = U_1 R + U_0$ 和 $v = V_1 R + V_0$, 则有 $uv = U_1 V_1 R^2 + ((U_0 + U_1)(V_0 + V_1) - U_1 V_1 - U_0 V_0)R + U_0 V_0$, 该等式需要四次乘法来计算 uv 。然而, 由于 R 的乘法只是一个移位, 实际上仅进行了加法运算和三次乘法运算, 即 $U_1 V_1$ 、 $(U_0 + U_1)(V_0 + V_1)$ 及 $U_0 V_0$ 。因此实际应用中, 当整数的长度小于阈值时, 标准的乘法算法会比该递归实现速度更快, 该阈值 d_0 主要取决于所使用的处理器。完整的 Karatsuba 算法如算法 2-1 所示。

算法 2-1: Karatsuba 算法

Input:

n 字节整数 $u = (u_{n-1} \cdots u_0)_b$, m 字节整数 $v = (v_{m-1} \cdots v_0)_b$, $d = \max\{m, n\}$, d_0 为阈值

Output:

$(m+n)$ 字节整数 $w = (w_{m+n-1} \cdots w_0)_b$, 使得 $w = uv$

1: 若 $d \leq d_0$, 则返回 uv

2: $p \leftarrow \lfloor d/2 \rfloor$, $q \leftarrow \lceil d/2 \rceil$

3: $U_0 \leftarrow (u_{q-1} \cdots u_0)_b$, $V_0 \leftarrow (v_{q-1} \cdots v_0)_b$

4: $U_1 \leftarrow (u_{p+q-1} \cdots u_q)_b$, $V_1 \leftarrow (v_{p+q-1} \cdots v_q)_b$

5: $U_s \leftarrow U_0 + U_1$, $V_s \leftarrow V_0 + V_1$

6: 递归计算 $U_0 V_0$, $U_1 V_1$, $U_s V_s$

7: 返回 $U_1 V_1 b^{2q} + ((U_s V_s - U_1 V_1 - U_0 V_0))b^q + U_0 V_0$

2.1.2 Montgomery 模约减

Montgomery 引入了一种巧妙的方法来表示 Z/NZ 中的元素, 从而简化算术运算, 特别是乘法运算^[2]。该算法可以看作是 Hensel odd division 的一般化, 用于计算 2-adic 数的倒数。在实际使用中, 往往取 N 为素数。

定义 R 是大于 N 的整数, 并且与 N 互素, x 的 Montgomery 表示为 $[x] = (xR) \bmod N$, $x \in [0, N-1]$ 。 $u \in [0, RN-1]$, u 的蒙哥马利约减为 $\text{REDC}(u) = (uR^{-1}) \bmod N$ 。

令 $N' = (-N^{-1}) \bmod R$, k 为 $[0, N-1]$ 中使得 $k \equiv uN' \pmod R$ 成立的唯一整数。那么显然 $(u + kN)$ 是 R 的倍数。令 $t = (u + kN) / R$ 。由于 N 和 R 互素, 因此 $t \equiv uR^{-1} \pmod N$ 。根据假设 $0 \leq u < RN$, 容易证明 $0 \leq t < 2N$, 所以 t 或 $t - N$ 等于 $\text{REDC}(u)$ 。Montgomery 模约减算法如算法 2-2 所示。

算法 2-2: Montgomery 模约减算法

Input:

n 字节整数 $N = (N_{n-1} \cdots N_0)_b$, $\gcd(N, b) = 1$, $R = b^n$, $N' = (-N^{-1}) \bmod b$, 且有 $2n$ 字节整数 $u = (u_{2n-1} \cdots u_0)_b < RN$

Output:

n 字节整数 $t = (t_{n-1} \cdots t_0)_b$, 使得 $t = \text{REDC}(u) = (uR^{-1}) \bmod N$

1: $(t_{2n-1} \cdots t_0)_b \leftarrow (u_{2n-1} \cdots u_0)_b$

2: 对 $i = 0, 1, \dots, n-1$ 循环迭代

3: $k_i \leftarrow (t_i N') \bmod b$

4: $t \leftarrow t + k_i N b^i$

5: $t \leftarrow t / R$

6: 若 $t \geq N$, 则 $t \leftarrow t - N$

7: 返回 t

2.1.3 Barrett 模约减

Barrett 模约减于 1987 年被提出^[3]。对于 n 字节整数 N , 定义 $R(N) = \lfloor b^{2n} / N \rfloor$ 。若 u 为一个 $2n$ 字节整数, 则有

$$q = \left\lfloor \frac{u}{N} \right\rfloor = \left\lfloor \frac{\frac{u}{b^{n-1}} \frac{b^{2n}}{N}}{b^{n+1}} \right\rfloor$$

可近似为

$$\hat{q} = \left\lfloor \frac{\left\lfloor \frac{u}{b^{n-1}} \right\rfloor R}{b^{n+1}} \right\rfloor$$

有 $q-2 \leq \hat{q} \leq q$ ，并且在 90% 的情况下 $\hat{q} = q$ ，仅在 1% 的情况下 $\hat{q}$ 与 q 的差值为 2。在模约减运算中，若 $\hat{u} = u - \hat{q}N > N$ ，则最多需要对 $\hat{u}$ 做 2 次减法就可以得到模约值。Barrett 模约减预计算与 Barrett 模约减算法如算法 2-3 与算法 2-4 所示。

算法 2-3: Barrett 模约减预计算

Input:

n 字节整数 $N = (N_{n-1} \cdots N_0)_b$

Output:

$(n+2)$ 字节整数 $R = \lfloor b^{2n} / N \rfloor$

1: $R \leftarrow b^n$

2: 循环迭代

3: $s \leftarrow R$

4: $R \leftarrow 2R - \lfloor N \lfloor R^2 / b^n \rfloor / b^n \rfloor$

5: 直至 $R \leq s$

6: $t \leftarrow b^{2n} - NR$

7: 当 $t < 0$ 时，循环计算 $R \leftarrow R-1, t \leftarrow t+N$

8: 返回 R

算法 2-4: Barrett 模约减算法

Input:

$2n$ 字节整数 $u = (u_{2n-1} \cdots u_0)_b$ ， n 字节整数 $N = (N_{n-1} \cdots N_0)_b$ 且 $N_{n-1} \neq 0$ ，预计算值 $R = \lfloor b^{2n} / N \rfloor$

Output:

n 字节整数 $r = (r_{n-1} \cdots r_0)_b$ ， $u \equiv r \pmod{N}$

1: $\hat{q} \leftarrow \lfloor \lfloor (u / b^{n-1}) \rfloor R / b^{n+1} \rfloor \triangleright [q-2 \leq \hat{q} \leq q]$

2: $r_1 \leftarrow u \bmod b^{n+1}, r_2 \leftarrow (\hat{q}N) \bmod b^{n+1}$ and $r \leftarrow r_1 - r_2$

3: 若 $r < 0$ ，则 $r \leftarrow r + b^{n+1}$

4: $t \leftarrow t + k_i N b^i$

5: $t \leftarrow t / R$

6: 当 $r \leq N$ 时, 循环计算 $r \leftarrow r - N$

7: 返回 r

2.2 编码优化技巧

Smart 和 Vercauteren 在 BGV 方案中引入了 SIMD (Single Instruction Multiple Data) 操作^[4], 即单指令流多数据流, 通过在多项式环上使用中国剩余定理, 令密文呈现多槽 (Slots) 结构, 允许把明文数组插进槽中, 这样对两个密文进行同态操作时, 就可以对数组向量按分量进行同态操作, 提高了空间利用率和计算效率, 降低了密文转换率, 使得基于 RLWE 的同态加密方案相较于 LWE 方案呈现了巨大的优势。在引入 SIMD 操作之后, 可以将 RLWE 密文对应明文的 n 位全都有效使用, 因此后续以 BGV 为基础的全同态加密方案的设计大多基于 RLWE 问题。

令 ζ 是 $2N$ 次单位原根, 即 $\zeta^{2N} = 1 \bmod t, \zeta^N = -1 \bmod t$, $x^N + 1$ 可以因式分解为 $\prod_{i=0}^{N-1} (x - \zeta^{2i+1})$,

依据中国剩余定理 (Chinese Remainder Theorem, CRT) 有

$$R_t = \frac{\mathbb{Z}_t[x]}{x^N + 1} = \frac{\mathbb{Z}_t[x]}{\prod_{i=0}^{N-1} (x - \zeta^{2i+1})} \simeq \text{CRT} \prod_{i=0}^{N-1} \frac{\mathbb{Z}_t[x]}{x - \zeta^{2i+1}} = \prod_{i=0}^{N-1} \mathbb{Z}_t[\zeta^{2i+1}]$$

R_t 上任意多项式 $p(x) = \sum_i p_i x^i$ 都可以等价表示成它在 ζ^{2i+1} 上的估值; 将该运算记作 $\sigma(p(x))$, 用矩阵的形式表示为

$$\begin{bmatrix} m_0 \\ \vdots \\ m_{N-1} \end{bmatrix} = \begin{bmatrix} \zeta^{1 \cdot 0} & \cdots & \zeta^{1 \cdot (N-1)} \\ \vdots & & \vdots \\ \zeta^{(2N-1) \cdot 0} & \cdots & \zeta^{(2N-1) \cdot (N-1)} \end{bmatrix} \cdot \begin{bmatrix} p_0 \\ \vdots \\ p_{N-1} \end{bmatrix}$$

用向量 $[m_0, \dots, m_{N-1}]$ 表示多项式上的 N 个估值。这些估值在 SIMD 技术中就是需要进行编码的明文向量。也就是说, 在 BGV/BFV 中, 需要加密的信息是一个明文向量 $[m_0, \dots, m_{N-1}]$, 通过 SIMD 编码将其转换成 R_t 上的多项式 $p(x)$, 进而对 $p(x)$ 进行正常的 RLWE 加密和相应的同态操作, SIMD 编码实际上就是上面提到的多项式估值的逆运算, 记作 σ^{-1} , 用矩阵的形式表示为

$$\begin{bmatrix} p_0 \\ \vdots \\ p_{N-1} \end{bmatrix} = \begin{bmatrix} \zeta^{1 \cdot 0} & \cdots & \zeta^{1 \cdot (N-1)} \\ \vdots & & \vdots \\ \zeta^{(2N-1) \cdot 0} & \cdots & \zeta^{(2N-1) \cdot (N-1)} \end{bmatrix}^{-1} \cdot \begin{bmatrix} m_0 \\ \vdots \\ m_{N-1} \end{bmatrix}$$

Gentry 团队^[5]针对 SIMD 编码的特点优化了同态运算电路,对于层数较少的算数电路,可以实现对数多项式级别的开销;对于安全参数 λ 、任意 t 个门、深度 L 、平均深度为 $\Omega(\lambda)$ 的电路,可以在 $t \text{ polylog}(\lambda)$ 时间内完成同态计算。

GSW 类方案不同于 BGV/BFV 类加密方案,无法将一个密文向量打包加密为多条消息或多个密文,从而实现并行运算,Steven Myers 和 Adam Shull^[6]提出了一种打包密文的自举算法,将递归思想应用于 FHEW,计算复杂度降低为 $O(3^\rho \cdot \lambda^{1/\rho})$ 。Feng-Hao Liu 和 Han Wang^[7-8]在 MS18 方案打包算法的基础上,引入了适用于 FHEW 的 SIMD 算法,FHEW 的计算复杂度直接优化到 $\tilde{O}(1)$ 。

表 1 给出了 4 条不同技术路线全同态加密方案对比。BGV 类方案基于 LWE 和 RLWE 不同困难问题,其效率有所不同,在 RLWE 环结构下,单个密文可加密多条消息,从而提高了密文的利用率,同时减少公钥尺寸。一些最新成果给出了 GSW 类全同态加密方案的 SIMD 打包理论支撑研究,但因为自举效率低、技术存在局限性,所以在代码实现上仍不支持上述两种操作。

表 1 Gentry 类、BGV 类、GSW 类、CKKS 类全同态加密方案对比

方案类别	Gentry 类	BGV 类		GSW 类	CKKS 类
假设基础	理想格	LWE	RLWE	(R)LWE	RLWE
明文空间	$\mathbb{Z}_2$	$\mathbb{Z}_t$	R_t	$\mathbb{Z}_2$	—
密文空间	$\mathbb{Z}_d$	$\mathbb{Z}_q^n$	R_q^n	$\mathbb{Z}_q^{n \times n}$	R_q^n
单个密文存储明文数	1	1	n	1	n
是否支持 SIMD	—	是	是	暂不支持	是

2.3 多项式运算优化技巧

在基于格的全同态加密方案中,多项式运算往往是方案实现的主要性能瓶颈。而在多项式运算中,多项式乘法的运算最为复杂。因此,本文主要关注多项式乘法的加速算法。针对定义

在环 $\mathbb{R} = \mathbb{Z}_q[x]/(f(x))$ （其中 $f(x)$ 为首一多项式）上的多项式，存在两个模块：多项式乘法模块与模约减模块。其中，多项式乘法模块可以基于软件、硬件及 RSA/ECC 协处理器完成，主要优化算法包括 NTT/FFT、Toom-Cook、Karasuba 和传统的 SchoolBook 乘法算法；而模约减模块，常用的算法包括蒙哥马利（Montgomery）约减及 Barrett 约减，如图 1 所示。

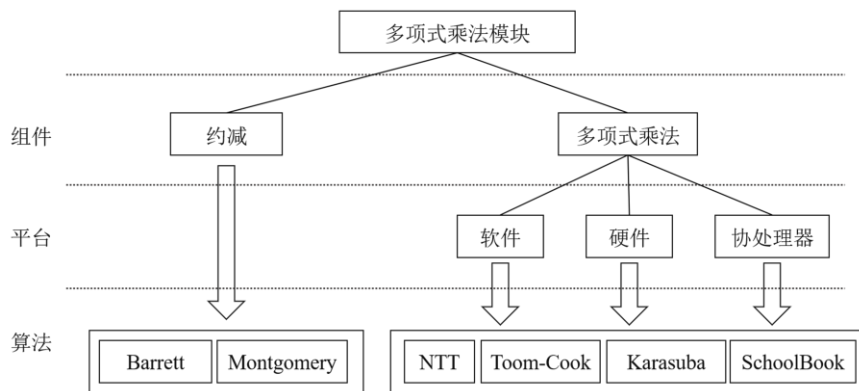


图 1 多项式乘法实现模块及平台

2.3.1 Karasuba 算法

Karasuba 算法是一种高效的多项式乘法算法。它依赖于分治的思想，将乘法的时间复杂度从传统的 $O(n^2)$ 降低至 $O(n^{\log_2 3})$ 。该算法的基本思想是将一个高次多项式分解为两个低次多项式：

$$A(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_2x^2 + a_1x + a_0$$

$$A_l(x) = a_{n/2-1}x^{n/2-1} + \cdots + a_1x + a_0$$

$$A_h(x) = a_{n-1}x^{n/2-1} + \cdots + a_{n/2+1}x + a_{n/2}$$

根据上述多项式的分解，计算多项式 $A(x)$ 及 $B(x)$ 的乘积，可以写作 $C(x) = A_h(x) \times B_h(x)x^n + ((A_h(x) + A_l(x)) \times (B_h(x) + B_l(x) - (A_h(x) \cdot B_h(x) + A_l(x) \cdot B_l(x))))x^{n/2} + A_l(x) \times B_l(x)$ 。这样，对于一个 $n-1$ 次多项式乘法，可以把它分成 3 个 $n/2-1$ 次多项式乘法和一些额外的加法操作。递归使用 Karasuba 算法，从而可以进一步将多项式分成更低的次数，降低多项式乘法运算的时间复杂度。

2.3.2 Toom-Cook 算法

Toom-Cook 算法可以看作 Karasuba 的一般形式，利用分治的思想，降低多项式的次数从

而提高乘法的计算效率。与 Karatsuba 算法将多项式分为高次与低次两部分不同, Toom-Cook 算法将多项式分成 K 等份。每次的划分, Toom-Cook 算法需要执行 3 步, 分别是 Evaluation、Multiplication 及 Interpolation。

Evaluation 阶段: 将 $n-1$ 次多项式 $A(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_2x^2 + a_1x + a_0$ 划分为 k 个 $n/k-1$ 次多项式, 如图 2 所示。

$$\begin{array}{c}
 A(x) = \underbrace{(a_{n-1}x^{n/k-1} + \cdots + a_{n-n/k})}_{\mathcal{A}_{k-1}} x^{n/k(k-1)} + \underbrace{(a_{n-n/k-1}x^{n/k-1} + \cdots + a_{n-2n/k})}_{\mathcal{A}_{k-2}} x^{n/k(k-2)} + \cdots + \\
 \underbrace{(a_{n/k-1}x^{n/k-1} + \cdots + a_0)}_{\mathcal{A}_0} \\
 \text{Let } y = x^{n/k}, \text{ written as: } A(x) = \mathcal{A}_{k-1}y^{k-1} + \mathcal{A}_{k-2}y^{k-2} + \cdots + \mathcal{A}_0
 \end{array}$$

图 2 Toom-Cook 算法 Evaluation 阶段

Multiplication 阶段: 在该阶段利用点值来表示计算乘积, 选择 $2k-1$ 个特殊点 $(y_0, \cdots, y_{2k-2})$ 代入计算点值, 从而得到 $A(y_0)B(y_0), \cdots, A(y_{2k-2})B(y_{2k-2})$ 。

Interpolation 阶段: 通过构造范德蒙矩阵, 并通过计算逆矩阵重构出多项式乘法的乘积。Interpolation 阶段计算矩阵乘法:

$$\begin{bmatrix} C_{2k-2} \\ C_{2k-3} \\ \vdots \\ C_0 \end{bmatrix} = \begin{bmatrix} (y_0)^{2k-2} & \cdots & 1 \\ (y_1)^{2k-2} & & 1 \\ \vdots & & \vdots \\ (y_{2k-2})^{2k-2} & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} A(y_0)B(y_0) \\ A(y_1)B(y_1) \\ \vdots \\ A(y_{2k-2})B(y_{2k-2}) \end{bmatrix}$$

从而得到 $C_{2k-2}, \cdots, C_0$ 的值, 多项式乘法的乘积可以通过以下公式重新构建:

$$C(y) = C_{2k-2}y^{2k-2} + C_{2k-3}y^{2k-3} + \cdots + C_0$$

得到 $C(y)$, 其中 $y = x^{n/k}$ 。

2.3.3 NTT/FTT 算法

NTT 是有限域上快速傅里叶变换 (FFT) 的一个特殊形式。对于多项式乘法, NTT 是将系数表示法转化为点值表示法的一种方法。通过使用 $2n$ 次方根或 n 次方根来降低系数表达向点值表达转换的复杂度, NTT 便成为计算多项式乘法的有效方式。

如果需要计算环 $\mathbb{R} = \mathbb{Z}_q[x]/(x^n + 1)$ 上的多项式 $A(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_2x^2 + a_1x + a_0$, 并且 n 是 2 次幂的形式。存在 n 次方根 ω , 满足 $\omega^n \equiv 1 \pmod q$, 那么将多项式由系数表达 $(a_{n-1}, a_{n-2}, \cdots, a_0)$ 向点值表达 $(A_{n-1}, A_{n-2}, \cdots, A_0)$ 转换的正向 NTT 可以表示为

$$A_i = \sum_{j=0}^{n-1} a_j \omega^{ij} \bmod q, \quad i = 0, 1, \dots, n-1$$

逆向 NTT 转换可以表示为

$$a_i = n^{-1} \sum_{j=0}^{n-1} A_j \omega^{-ij} \bmod q, \quad i = 0, 1, \dots, n-1$$

3 面向 CPU 平台的全同态高速实现

相比于 GPU、FPGA 等硬件平台，CPU 实现的加速效果并不明显。CPU 本身并不是为了执行繁重的数据计算任务而设计的。它可以通过多线程提供一定程度的并行计算能力，但这远不足以满足 FHE 加速的需求。因此，大多数研究更倾向于将其与 GPU 或 FPGA 相结合来实现同态加密算法的加速。然而，CPU 的优势在于它是一个比较通用的平台，有着广泛的应用。

3.1 国内外发展现状

2020 年，研究者们的工作表明 SIMD 可以有效地在 CPU 和 GPU 上加速 NTT 与 iNTT^[9-10]。Intel CPU 上的 SIMD 操作，如单一指令多数据流扩充（SSE）和 Advanced Vector eXtensions（AVX2），得到了广泛的应用。例如，NTL2 和 NTLlib^[11]等数学库都是使用 SSE 或 AVX2 实现的。由于同态加密库采用了 NTL 或 NTLlib，因此这些库间接受益于 SIMD 的加速，其中 PALISADE3、HElib4 和 HEAAN5 采用 NTL，FV-NTLlib6 采用 NTLlib。Mathemagix 库^[12]提供了使用 SIMD 编程模型的模块化整数运算的 Intel AVX 加速实现。NTLlib 使用 Intel SSE 和 Intel AVX2 指令提供了类似于环 $Z_q/(X^N+1)$ 的原语加速。然而，Mathemagix 和 NTLlib 都没有考虑使用 Intel AVX512-IFMA52 指令集的 Intel AVX512 实现。在过去的研究中，文献[13-14]使用 Intel AVX512-IFMA52 来加速大整数乘法而不需要模乘。Drucker 和 Gueron^[15]使用 Intel AVX512-IFMA52，通过 Montgomery 乘法来加速大整数模平方运算。

使用 AVX512 加速同态加密在 2021 年左右开始。2021 年，Jung 等人提出使用 AVX512 实现同态乘法^[16]。与 HEAAN5 相比，它们获得了 2.06 倍的加速比。2021 年，Boemer 等人提出通过以体系结构为中心的分析 and 优化加速全同态加密^[17]。该方案使用 AVX512-IFMA52（整数融合 Multiply Add）实现了向量模算术运算和 NTT/iNTT，并在 PALISADE v1.11.3 上实现了同态乘法。与原 PALISADE v1.11.3 相比，加速比提高了 2.59 倍。

在同态加密的运算中，迹函数的研究也是众多学者关注的方向之一。迹函数可以看作是一种特殊的线性变换。2014 年，Halevi 与 Shoup 首次提出了总和算法^[18]。它通过一种“重复-加倍”的方法来工作，这种方法递归地应用了旋转与求和。总和运算作用在明文槽上，而迹函数

作用在多项式系数上。它的设计对于任意长度的待处理向量（包括两个长度的非幂）都是通用的。迹函数的主要作用是隔离多项式的系数，当只需要提取一个特定的系数时，这个问题与LWE到RLWE的转换问题相同^[19-20]。

2017年，Chillotti等人提出了更快速的封装同态运算和高效的电路自启动TFHE^[20]。该方案在LWE密文上应用和环维数一样多的密钥交换过程。虽然这种方法具有很高的并行性，但由于需要 $O(N)$ 大小的计算密钥，它的计算效率很低，并且需要千兆字节或万亿字节的存储。随着研究的开展，在LWE到RLWE的转换中，这些限制已经通过递归自同构总和^[19]被消除。因此，存储和计算复杂度都降低了一个指数因子。

迹函数可以认为是线性变换。线性变换的功能是通过对旋转后的密文乘以一些常数进行同态求和来实现的。在迹函数求值中，乘法因子总是1，并且可以从线性变换的文献中得到其他的潜在解。2018年，Halevi和Shoup提出了一系列优化技术来降低同态线性变换的代价^[21]，方案引入了3种密钥交换策略的变体：最小密钥交换、BS/GS和完全密钥交换。这些策略在密文上执行单个自同构所需的密钥交换次数上存在差异。2020年，Bossuat等人提出了一套自同构的优化技术^[22]，其中关于一般线性变换的密钥转换的代价模型在BS和GS中是不对称的。代价模型的不对称性是由于在密文上应用自同构后存在乘法运算导致的，这就要求在每次迭代一个大步时对小步进行惰性降模操作。同时，他们的方案没有提供多线程模式下的计时结果。

2021年，Boemer等人基于Intel和Advanced Vector Extensions 512（AVX512）指令集，实现了多项式模乘和NTT的加速^[17]。他们提供了一个同态加密加速库，可以与SEAL库结合使用。该加速方案主要使用了两种优化方法，包括循环优化和数据并行计算优化，这两种优化方法都是基于指令集的。循环通过手动或使用预处理器指令进行展开，并通过手动调节展开因子。在手动展开的循环中，指令在可能的情况下被重新排序，以便进行最佳的流水线操作。基于该指令集的性质，可以同时处理8个64位整数，该加速方案通过将NTT和iNTT以元素的形式表示并一起处理，完成了对NTT的加速。对于以元素形式表示的向量与向量乘法和多项式与多项式乘法，同样基于同时处理8个64位整数的目的对输入数据进行对齐，实现性能优化。

2021年，Ishimaki等人设计的方案加速了CKKS的迹函数求值运算^[23]。同态迹函数求值是通过重复同态旋转后再进行加法（旋转求和）来实现的，是一种常用且耗时的子程序，能够同态求和向量的分量或同态提取多项式的系数。他们通过利用自同构的性质和使用多核环境，提出了一种使用循环展开的更有效的迹函数求值方法，以减少昂贵操作的数量。因此，该方案以略微增加所需存储量为代价，降低了计算成本。此外，它们在迹函数求值中依次展开连续的子循环，并对展开后的迭代次数进行参数化，进一步实现了对同态迹函数求值的加速。

2022年，Inoue等人采用了AVX512指令集对迹函数进行了加速^[24]。通过使用循环展开，他们实现了对迹函数的优化。他们的加速方案可以看作是AVX512指令集在加速FHE方面的另一种实现方案。

大多数研究^[23-24]都是基于 AVX512 指令集来加速 FHE, 因为主流的 CPU 指令集并不适合。这也说明了 CPU 对于 FHE 加速的局限性。

3.2 以架构为中心的全同态加密优化加速方案

在迄今为止提出的众多 HE 方案中, HE for Arithmetic of Approximate Numbers (HEAAN) 因其支持近似计算, 且不限限制适用于密文的算术运算次数的消息而在广泛的应用中迅速流行。这导致了大量的 HE 加速研究, 包括利用 FPGA 的研究。然而, 对同态乘法的计算复杂度和数据访问模式缺乏严格的分析, 且这些提案大多集中在参数规模较小的设计上, 使得在进行一系列复杂的算术运算时, 很难准确地估计 HE 加速器的性能。本方案基于对 HEAAN 的计算和内存访问特性的分析, 提取组成同态乘法的关键函数中的并行性, 并通过转置矩阵和将数据固定到线程等方法, 将并行性有效地映射到流行的并行处理平台上。下面给出方案的设计过程^[16]。

1. 循环反复排序以在 iCRT 中公开大规模并行

有限的 N 度并行可以通过在 iCRT 中对两个循环进行重新排序, 从而被扩展到 $N \cdot \text{PLimbs}$ 度并行。重新排序后, 标量和大整数之间的原始乘法序列变为矩阵-矩阵乘法。然后修改 iCRT 以便累积最内层循环中的部分和。通过循环重新排序, 得到的矩阵公开了 iCRT 中的 $N \cdot \text{PLimbs}$ 度并行, 为当代硬件平台提供了丰富的并行化机会。

2. 加速 CPU 上的同态乘法

利用同态乘法主要功能中固有的并行性, 首先将操作分发到多个线程, 然后分配到每个线程中的 AVX512 SIMD 通道, 同时最大限度地减少因数据访问模式不佳而导致的频繁缓存未命中。在此过程中, 输入矩阵的顺序 (是以行还是以列为主) 会影响性能。

(1) iCRT 中的矩阵转置。如果 SIMD 指令根据列方向访问以行为主顺序存储元素的矩阵 (反之亦然), 那么可能会导致利用率低。当访问间隔太大而硬件预取程序无法有效时, 由此产生的性能下降更为突出。iCRT 遇到此问题是因为矩阵-矩阵乘法访问临时矩阵, 该矩阵根据列方向以行主顺序存储元素, 使用 AVX512 中的散点指令隐式转置临时矩阵来解决此问题。

(2) 将 β 的大小减小到 2^{32} 。可以通过使用 2^{32} 而不是 2^{64} 的 β 来消除模拟 64bit 操作的开销, 因为 AVX512 支持 32bit mul 和 ADC。使用较小的 β 将 mul 和 ADC 的指令数更改为一个, 并将模块化 mul 的指令数减少到不到一半。然而, 该策略也存在不足之处, 即要 qLimbs 粗略地加倍才能表示大小达到 q 的数。 m 也必须加倍, 因为每个素数的上限都是 β 。较大的 qLimbs 和 m 意味着更大的预计算表和更多的 CRT、NTT、iNTT 和 iCRT 迭代。它还增加了除 mul、模块化 mul 和 ADC 之外的操作数量。初步实现显示, 当对 CPU 使用小型 β 时, 性能没有提高。

3. 加速 GPU 上的同态乘法

现代 GPU 拥有与单精度浮点运算 (FP32) 相同数量的整数 (INT32) 单元, 这使得 INT32 数的乘累加吞吐率与 FP32 数相同。如此巨大的整数运算吞吐量使得 GPU 成为加速 HE 运算的一个有吸引力的候选者。

(1) 并行化策略。基本的并行化策略是将每个独立可计算的输出元素分配给一个线程。特别地, 为 CRT 启动了 $N \cdot m$ 个线程, 每个线程计算一个输出元素 (一个余数)。NTT 和 iNTT 各启动 $N/2 \cdot m$ 个线程, 每个线程执行一个 Buttl 操作。通过上述的改变循环顺序, 将相应的核心操作转化为矩阵-矩阵的乘操作, 从而利用 $N \cdot \text{PLimbs}$ 度并行来最大化线程级并行度。

(2) 64bit 模拟与 32bit 字对比。与大多数原生支持 64bit 字的 CPU 不同, 现代 GPU 原生支持 32bit 字并模拟 64bit 整数运算。为了避免 64bit 模拟的开销, 先前的研究在 GPU 上加速 HE。使用了 32bit 的字符串和操作。在 GPU 上使用 32bit 字的另一个优点是使用转入库存和执行过程的操作可以在没有模拟的情况下执行。现代 NVIDIA GPU 支持 addc、subc 和 maddc 等指令的进位操作。这些指令的吞吐率与最近的 GPU 中没有进位的指令的吞吐率相同, 从而可以在没有模拟的情况下实现大整数的高效计算。

(3) CRT 优化。为了保证 accum 为两个字长, 使用 2^{57} 作为 AVX512 实现的素数下界, 而不是参考 HEAAN 的默认值 2^{59} 。然而, 在 β 为 32bit 的 GPU 上, 当素数小于 2^{30} 时, 只允许最多 4 次的累加, 以保证不会出现溢出问题, 这在使用代表性参数时几乎是不可能的, 因为 m 为 90 或更高。为了防止溢出, 可以在最内层循环中使用一个包含额外 ADC (ADDITION with Carry, 进位加法) 操作的三字 accum, 以避免昂贵的取模操作; 或在最内层循环中间歇地进行取模操作, 以确保 accum 仅跨两个字。该方案实现了两种策略, 并发现使用三字 accum 具有更好的性能, 因为它需要更少的模运算。

(4) iCRT 优化。为了减小内存的未命中损失, 本方案建议在 iCRT 中使用 N 并行时进行两种不同的优化: ①通过网格步幅循环方法简单地减少块和网格尺寸来使用更少的线程; ②通过在共享内存中分配数组来固定每个 accum 数组。

4. 可扩展性研究

本方案扩展了同态乘法的实现, 使其支持分布式内存多 GPU 系统, 并研究了相应的 CPU 和 GPU 可扩展性。对于采用多 Socket 处理器的 CPU 实现, 可以采用前面 CPU 上的加速同样的并行化方法来填充更多的核。然而, 对于使用多个 GPU 的实现, 采取了不同的方法, 因为它们采用分布式内存架构。

在单个 GPU 的内存中使用一个多项式, 让 GPU 对多项式进行操作。由于将每个多项式输入 GPU 中, 而一个同态乘法使用 4 个多项式作为输入, 因此在一个同态乘法中存在 4 个流, 每个流都是在 GPU 上执行的操作序列。当执行一元操作, 如 CRT、iCRT、NTT、iNTT 时, 一个流不需要向另一个流传输数据或同步数据。一个流只同步到另一个流进行二进制操

作, 等待它的一个操作数从另一个流中转移。通过将一个流分配给一个 GPU, 可以为单个同态乘法开发多达 4 个 GPU。如果使用的 GPU 少于 4 个, 那么可以将多个流分配到同一个 GPU 上。

3.3 同态迹函数加速方案

同态迹函数求值是通过重复同态旋转后再进行加法(旋转求和)来实现的。为了正确地旋转密文和未旋转密文相加, 需要对旋转密文进行特殊的密钥交换操作。由于密钥交换的计算代价昂贵, 轮换与求和本质上是低效的。本方案基于循环展开的迹函数求值方法, 提高了同态迹函数求值的效率^[23]。

本方案提出了下述算法, 作为现有迹函数的推广。

算法 3-1: 非滚动迹函数同态计算函数

Input:

- c : 密文
- h : 迭代次数
- $K = \{k_0, \dots, k_{h-1}\}$: 每个迭代的自同构指标的一组数组

Output:

密文 c'

```

1:  $c' \leftarrow c$ 
2: for  $i = 0$  to  $h-1$  do
3:    $c' \leftarrow c' \oplus \text{AKS}(k_i[0], c') \oplus \dots \oplus \text{AKS}(k_i[|k_i|-1], c')$ 
4: end for
5: return  $c'$ 
```

一般来说, 该算法中的每次迭代都需要在密文上进行不同的自同构, 然后进行密钥交换。上述算法可以利用许多优化来降低密钥交换的复杂性, 如下所述。

1. 优化技巧

该算法利用了许多众所周知的优化技术, 这些技术可以独立地应用于每个未滚动的迭代。在这里考虑一个由 n 个自同构和密钥交换过程组成的非滚动迭代。假设自同构应用于密文 $c = (c_0, c_1) \in R_{Q_t}^2$, 指数为 $t_0, t_1, \dots, t_{n-1} \in \mathbb{Z}_{2N}^*$ 。

(1) 惰性模数提升和预自同构。考虑到每次展开迭代的自同构都适用于同一个密文, 该方案利用惰性模数提升技术^[21-22]。具体来说, RNS 分解和模数提升都执行一次, 得到的 $R_{PQ_t}^\beta$

中的分解元素被重复用于不同的自同构，然后进行密钥交换。

进一步采用类似文献[10]中的惰性模数下降法，通过提升和计算同一密钥下的自同构，对 $c_0 \in R_{Q_0}$ 和 $R_{PQ_0}^2$ 上的内积结果施加 n 个不同的自同构。然后，在进行模数下降之前，将内积的结果求和。具体来说，方案的目标是计算

$$(\hat{c}_0, \hat{c}_1) = c_0 + \sum_{e=0}^{n-1} \psi_{t_e}(c_0) + \lfloor \frac{\sum_e \psi_{t_e}(\langle \text{swk}'_e[1], \hat{d} \rangle)}{P} \rfloor \quad c_1 + \lfloor \frac{\sum_e \psi_{t_e}(\langle \text{swk}'_e[1], \hat{d} \rangle)}{P} \rfloor$$

(2) 多种技术整合。通过结合上述技术，与算法 3-1 的初始形态相比，在内积前后的密钥交换中，所有操作的数量都减少了。操作复杂度的增加与内积复杂度的增加是不对称的，因此选择合适的 h 可以提高计算效率。算法 3-2 展示了每次展开迭代的过程。

算法 3-2: 非滚动迹函数同态评估 (Hoistsks)

Input:

- $c = (c_0, c_1)$: 密文
- $k = \{k_0, \dots, k_{|k|-1}\}$: 自同构数组
- $ek = \left\{ (\text{swk})_{s \rightarrow s \left(X^{(k[i])^{-1}} \right)} \right\}_{i \in [|k|]} \in (R_{PQ_0}^{\beta \times 2})^{|k|}$

Output:

密文 c'

- 1: $(c'_0, c'_1) \leftarrow c$
- 2: $\hat{d} \leftarrow \text{RNSDecompModUp}(c'_1)$
- 3: $v \leftarrow []$ // R_{Q_0} 中长度为 $|k|$ 的空数组
- 4: $\tilde{v}_0, \tilde{v}_1 \leftarrow []$ // R_{PQ_0} 中长度为 $|k|$ 的空数组
- 5: for $i = 0$ to $|k|-1$ do
- 6: $v[i] \leftarrow \psi_{k_i}(c'_0)$
- 7: $\tilde{v}_0[i] \leftarrow \psi_{k_i}(\langle \hat{d}, ek[i][0] \rangle)$
- 8: $\tilde{v}_1[i] \leftarrow \psi_{k_i}(\langle \tilde{d}, ek[i][1] \rangle)$
- 9: end for
- 10: $d_0 \leftarrow v[0]$
- 11: $\tilde{c}_0 \leftarrow \tilde{v}_0[0]$

```

12:  $\tilde{c}_1 \leftarrow \tilde{v}_1[1]$ 
13: for  $i = 0$  to  $|k|-1$  do
14:      $d_0 \leftarrow d_0 + v[i]$ 
15:      $\tilde{c}_0 \leftarrow \tilde{c}_0 + \tilde{v}_0[i]$ 
16:      $\tilde{c}_1 \leftarrow \tilde{c}_1 + \tilde{v}_1[i]$ 
17: end for
18:  $(c_0'', c_1'') \leftarrow (\text{RNSModDown}(\tilde{c}_0), \text{RNSModDown}(\tilde{c}_1))$ 
19:  $(c_0', c_1') \leftarrow (d_0, 0) \oplus (c_0'', c_1'')$ 
20: return  $(c_0', c_1')$ 
    
```

优化后的非滚动迹函数同态评估如算法 3-3 所示。

算法 3-3: 优化后的非滚动迹函数同态评估

Input:

- c : 密文
- h : 展开后迭代次数
- $K = \{k_0, \dots, k_{h-1}\}$: 每个迭代的自同构指标的一组数组
- $E = (ek'_0, \dots, ek'_{h-1})$: 每个自同构的评估密钥的向量阵列

Output:

密文 c'

```

1:  $c' \leftarrow c$ 
2: for  $i = 0$  to  $h-1$  do
3:      $c' \leftarrow c' \oplus \text{HoistKS}(c', k_i, ek'_i)$ 
4: end for
5: return  $c'$ 
    
```

2. 近似成本分析

对于算法 3-2 中的每个非滚动迭代，有如下代价。

- (1) RNS 分解和模提升（第 2 行）。
- (2) 内积自同构（第 5~9 行）。
- (3) 惰性 Modulus-Down（第 10~18 行）。
- (4) 加法和输出（第 19 行）。

文献[22,28]中通常考虑的模乘复杂度，忽略自同构、加法和实现因素（如内存访问和缓存未命中）的成本，可以得到一个近似最优的 h 。

实验证明，该方法相比之前的旋转求和方法，获得了 1.32~2.12 倍的运行时间。

3.4 基于 AVX512 指令集加速同态迹函数

基于 Intel AVX512, 文献[24]对迹函数进行了进一步的加速, 并与现有的循环展开优化迹函数进行了比较。结果表明, 实现的 AVX512 版本比非 AVX512 版本加速了 1.05~2.30 倍。下面给出方案的设计过程。

1. 矢量到矢量的融合乘加 (v2vFMA)

英特尔 HEXL 已经有了融合向量到尺度的融合乘加操作。然而, 向量之间的融合乘加操作尚未实现, 因此需要依次采用被定义为独立函数的模加和模乘。新实现的向量融合乘加运算依次应用模加和模乘运算, 如算法 3-4 所示。然后, 对该函数进行了实现, 使得在输出该函数结果时, 只需将输出模乘和模加结果时进行的两次模运算集成到一次运算中。

算法 3-4: EltwiseVectorVectorFMAMod (基于文献[17]的算法)

```
Input:
    q, X, Y, Z, barr_lo
Output:
    X'
1: prod_hi ← _mm512_hexl_mulhi_epi64(X,Y)
2: prod_lo ← _mm512_hexl_mullo_epi64(X,Y)
3: c1 ← _mm512_hexl_shrdi_epi64(L-1>(prod_lo, prod_hi)
4: c3 ← _mm512_hexl_mulhi_epi64(c1, barr_lo)
5: c4 ← _mm512_hexl_mullo_epi64(c3, q)
6: c4 ← _mm512_sub_epi64(prod_lo, c4)
7: c5 ← _mm512_add_epi64(c4, Z)
8: X' ← _mm512_hexl_small_mod_epu64<8>(c5, q * 2, q * 4)
9: return X'
```

通过与 MultAdd 的比较来评估 v2vFMA 实现。MultAdd 和 v2vFMA 都是用 AVX512 实现的。当输入数据量小于 220 字节时, v2vFMA 的执行时间比 MultAdd 长。但当输入数据量大于 220 字节时, v2vFMA 的执行时间比 MultAdd 短。当输入数据规模较小时, MultAdd 的缓存机制的有效性克服了 v2vFMA 的有效性, 即融合的乘加操作和减少的模函数数量。然而, 对于大的输入数据规模, MultAdd 单独实现乘法和加法会出现错误, MultAdd 必须将 $a_i \times b_i$ 的数据存储到 L3 缓存或内存中, 再加载。

2. 采用 AVX512 对迹函数进行展开

将 Intel HEXL 应用于同态加密库中使用的向量到向量的模算术运算中, 实现了非滚动踪

迹类型函数的加速。这里选择了 PALISADE v1.9.2 作为同态加密库，并对其进行一些小的修改，实现了模加、模减和模融合乘加。此外，该方案还对 HoistKS（算法 3-2）中每个迭代中的模块化操作进行了并行化。

实验表明，当 M 和 k 固定时，加速比随着 l 的增加而增加，因为密文的数据大小随着 l 的增加而增加。这种趋势是因为在输入数据量较大时，v2vFMA 实现比 Mult Add 具有更好的性能。

3.5 基于 AVX512-IFMA52 的同态加密库

基于英特尔®高级矢量扩展 512（Intel® AVX 512）指令集，Boemer 等人^[17]实现了多项式模乘和 NTT 加速。他们将其作为同态加密加速库，该库可与 SEAL 和 PALISADE 库结合使用。

加速方案主要采用了两种优化方法，包括循环优化和数据并行计算优化，这两种优化方法都是基于指令集的。循环可以手动展开，也可以使用预处理器指令展开，并具有手动调整的展开因子。在手动展开的循环中，指令在可能的情况下被重新排序以获得最佳流水线。该加速方案根据指令集可同时处理 8 个 64 位整数的特性，将 NTT 和 iNTT 以元素的形式表示并一起处理，完成了对 NTT 的加速。对于以元素形式表示的向量和向量乘法以及多项式和多项式乘法，也基于同时处理 8 个 64 位整数的目的对齐输入数据，以实现性能优化。

在多项式运算层中，将英特尔 HEXL 设计嵌入到 HE 库中，多项式为 RNS 形式。多项式层通常是不同同态加密方案所共有的，这使得英特尔 HEXL 能够以最小的 HE 库集成面来加速多个 HE 方案。图 3 显示了英特尔 HEXL 库集成到标准 HE 库的可视化框架。

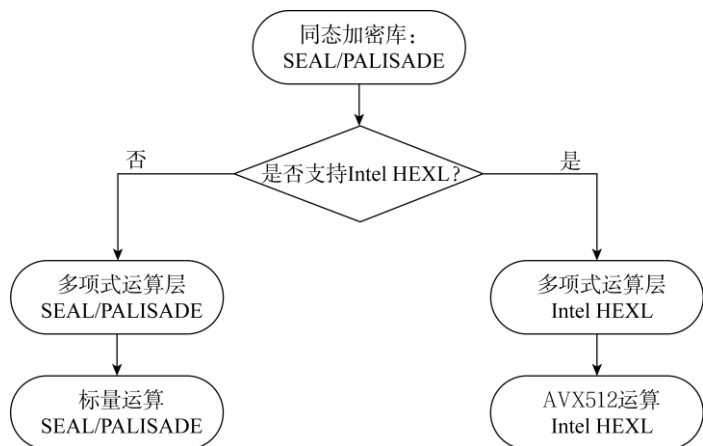


图 3 英特尔 HEXL 库集成到标准 HE 库的可视化框架

同态加密优化的算子主要有 NTT 运算与模乘运算。对于 RNS 形式下的整数模运算，HEXL 库设计者基于 AVX512 指令集设计了一套面向同态算子的整数运算指令，如图 4 所示。

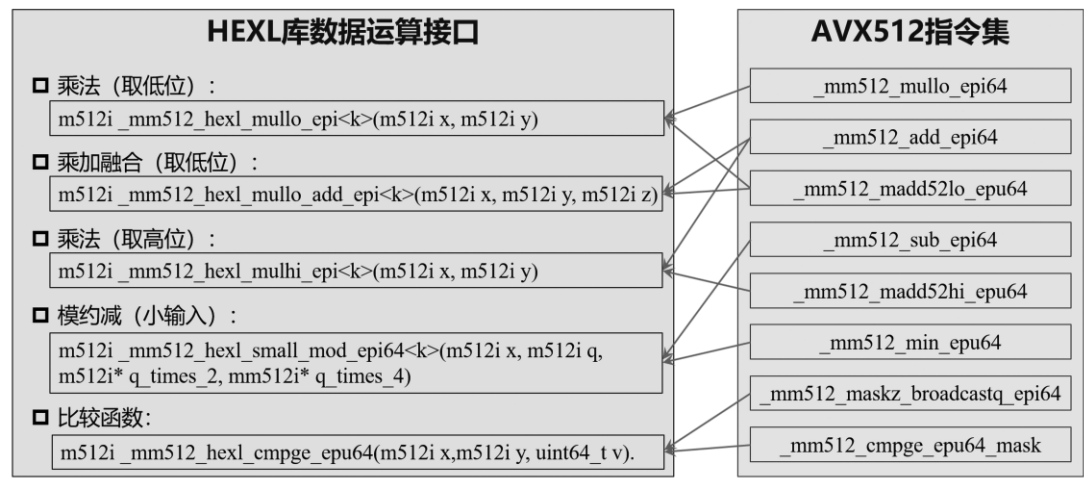


图 4 RNS 形式下的整数模运算

在多项式运算层面，对于 3 个核心运算部件：向量-向量模乘、正向 NTT 与逆向 NTT，基于 HEXL 库数据运算接口分别实现了并行版本的算子，向量-向量模乘算法如算法 3-5 所示。

算法 3-5：向量-向量模乘算法

Input:

q, X, Y

Output:

$X \cdot Y \bmod q$

```
1: X = _mm512_hexl_small_mod_epu64<InputModFactor>(X, q, twice_q);
2: Y = _mm512_hexl_small_mod_epu64<InputModFactor>(Y, q, twice_q);
3: __m512i prod_hi = _mm512_hexl_mulhi_epi<64>(X, Y);
4: __m512i prod_lo = _mm512_hexl_mullo_epi<64>(X, Y);
5: __m512i c1 = _mm512_hexl_shrdi_epi64<BitShift - 1>(prod_lo, prod_hi);
6: __m512i c3 = _mm512_hexl_mulhi_epi<64>(c1, barr_lo);
7: __m512i c4 = _mm512_hexl_mullo_epi<64>(c3, q);
8: c4 = _mm512_sub_epi64(prod_lo, c4);
9: __m512i result = _mm512_hexl_small_mod_epu64(c4, q);
10: return result;
```


遵循 Harvey NTT 蝴蝶算法，用 AVX512 实现正向 NTT，算法流程如算法 3-6 所示。遵循 Harvey iNTT 蝴蝶算法，用 AVX512 实现逆向 NTT，算法流程如算法 3-7 所示。

算法 3-6: 基于 AVX512 的正向 NTT 算法

Input:

$$q < \beta / 4; 0 < W < q; 0 \leq X, Y < 4q$$

Output:

$$X = X + WY \bmod q, Y = X - WY \bmod q$$

1: if !InputLessThanMod then

2: *X = _mm512_hexl_small_mod_epu64(*X, twice_modulus);

3: end if

4: __m512i Q = _mm512_hexl_mulhi_epi<BitShift>(W_precon, *Y);

5: __m512i W_Y = _mm512_hexl_mullo_epi<BitShift>(W_op, *Y);

6: __m512i T = _mm512_hexl_mullo_add_epi<BitShift>(W_Y, Q, neg_modulus);

7: if BitShift == 52 then

8: T = _mm512_and_epi64(T, _mm512_set1_epi64((1UL << 52) - 1));

9: end if

10: __m512i twice_mod_minus_T = _mm512_sub_epi64(twice_modulus, T);

11: *Y = _mm512_add_epi64(*X, twice_mod_minus_T);

12: *X = _mm512_add_epi64(*X, T);

算法 3-7: 基于 AVX512 的逆向 NTT 算法

Input:

$$q < \beta / 4; 0 < W < q; 0 \leq X, Y < 2q$$

Output:

$$X = X + Y \bmod q, Y = W(X - Y) \bmod q$$

1: __m512i Y_minus_2q = _mm512_sub_epi64(*Y, twice_modulus);

2: __m512i T = _mm512_sub_epi64(*X, Y_minus_2q);

3: if InputLessThanMod then

4: *X = _mm512_add_epi64(*X, *Y);

5: else

6: *X = _mm512_add_epi64(*X, Y_minus_2q);

7: __mmask8 sign_bits = _mm512_movepi64_mask(*X);

```
8:      *X = _mm512_mask_add_epi64(*X, sign_bits, *X, twice_modulus);
9: end if
10: __m512i Q = _mm512_hexl_mulhi_epi<BitShift>(W_precon, T);
11: __m512i Q_p = _mm512_hexl_mullo_epi<BitShift>(Q, neg_modulus);
12: *Y = _mm512_hexl_mullo_add_epi<BitShift>(Q_p, W_op, T);
13: if BitShift == 52 then
14:      T = _mm512_and_epi64(T, _mm512_set1_epi64((1UL << 52) - 1));
15: end if
```

3.6 方案对比

表 2 给出了上述几个基于 CPU 同态加密优化方案对比。

表 2 基于 CPU 同态加密优化方案对比

文献	优化方案	性能	主要优化算子
[16]	HEAAN	同态乘法在 CPU 上的性能比在 CPU 上运行 24 线程的 HEAAN 提高了 2.06 倍	CRT/iCRT、多项式运算
[23]	CKKS	对比之前的旋转求和方法 ^[18-19] 与 BS/GS 算法, 该方案获得了 1.32~2.12 倍的运行时间	旋转操作、同态加法
[24]	CKKS	AVX512 版本比非 AVX512 版本加速了 1.05 ~ 2.30 倍	旋转操作、密钥转换、同态加法
[17]	SEAL/PALISADE	相比 SEAL 库, AVX512 下 NTT 的加速比达到 4.7; 相比 PALISADE 库, NTT 的加速比达到 6.26	多项式运算、CRT 域下的整数运算

4 面向 GPU 的全同态高速实现

近年来, GPU 在全同态加速领域得到广泛应用。基于 GPU 的加速方案主要通过利用其提供的并行计算资源来实现。同时, 数据存储策略也将得到优化, 以减少访问数据所需的时间。然而, 由于 GPU 提供的有限定制灵活性, 基于 GPU 的加速方案的创新也受到限制。另外, 由于 GPU 是一个更通用的硬件平台, 其加速方案可以以加速库的形式呈现并投入使用。

4.1 CUDA 编程模型概述

CUDA 是由 NVIDIA 开发的并行计算平台和编程模型，用于利用 GPU 的强大计算能力来加速各种应用程序。CUDA 编程模型使开发人员能够在 GPU 上执行并行计算任务，从而实现更快的数据处理和计算。

(1) 主机和设备。在 CUDA 编程中，计算任务被划分为主机 (CPU) 和设备 (GPU) 两部分。主机负责协调工作和数据传输，而设备执行并行计算任务。

(2) 核函数 (Kernel Functions)。核函数是在 GPU 上并行执行的函数，它们是在设备上执行的代码片段。核函数的调用称为 “Kernel Launch”。核函数可以在每个线程中独立地运行，从而实现并行计算。

(3) 线程层次。CUDA 中的并行计算通过线程层次来管理。线程被组织成线程块 (Thread Block)，而线程块可以组成网格 (Grid)。每个线程块中的线程可以共享数据，并在共享内存中进行通信。

(4) 线程协调。在核函数中，开发人员可以使用内置的线程标识符来确定当前线程在线程块和网格中的位置。这些标识符可用于协调线程的行为和数据处理。

(5) 内存体系结构。CUDA 提供了不同类型的内存，包括全局内存、共享内存、常量内存和纹理内存。了解内存层次结构对于最大限度地提高性能至关重要。

(6) 数据传输。主机和设备之间的数据传输是一个重要的环节。数据可以从主机到设备传输，也可以从设备到主机传输。CUDA 提供了异步数据传输机制，允许在主机和设备之间同时进行计算与数据传输。

(7) 工具支持。NVIDIA 提供了一系列的开发工具，如 CUDA Toolkit，用于编写、编译和调试 CUDA 程序，其中包括编译器、性能分析工具和调试器。

4.2 国内外发展现状

2015 年，Wei Dai 等人提出了一个用于加速基于多项式环上同态方案计算的 CUDA GPU 库 cuHE^[29]。该库包含了多种优化技术，包括用于高效计算的代数技巧、内存最小化技术、内存和线程调度，以及针对 CUDA 的底层手动汇编优化，以充分利用 GPU 的大规模并行性和高内存带宽。论文中构建的算术函数被设计用于处理使用数论变换 (NTT) 和中国剩余定理 (CRT) 方法的大型多项式操作数，并扩展到用于实现由 López-Alt、Tromer 和 Vaikuntanathan 提出的分层同态加密方案的基本操作。

2019 年, Ahmad 提出了一种基于 CUDA 的 FV 实现方案^[30], 使用底层代数工具, 如中国剩余定理 (CRT)、剩余数系统 (RNS) 和离散伽罗瓦变换 (DGT), 来加速和简化在 GPU 上进行的 FV 计算。此外, 论文还展示了如何在 GPU 上执行整个 FV 计算, 而无须使用多精度算术。与 SEAL-FVRNS 和 NTLlib-FV 相比, 他们的 GPU 实现在密钥生成、加密、解密、同态加法和同态乘法方面的性能分别提高了 5.37 倍、7.37 倍、22.22 倍、5.11 倍和 13.18 倍 (或 2.03 倍、2.94 倍、27.86 倍、8.53 倍和 18.69 倍)。

2022 年, Özerk 等人针对大多数同态加密方案实现中的核心算术操作, 提出了一种基于 GPU 的 NTT/iNTT 和多项式乘法高速实现^[31]。论文消除了 NTT/iNTT 算法结构中尽可能多的依赖关系, 并根据多项式次数的大小自适应地选择共享内存或全局内存的工作模式, 这种混合方法相对于基准设计可实现多项式次数不同情况下的最高 9 倍速度提升。Özerk 对 Brakerski/Fan-Vercauteren (BFV) 同态加密方案的密钥生成、加密和解密操作进行了实验, (这些操作都严重依赖于基于 NTT 的多项式乘法), 与在 Intel i9-7900X CPU 上运行的高度优化的 SEAL 库相比, 该 GPU 实现在 Tesla v100 GPU 上分别提升性能 141.95 倍、105.17 倍和 90.13 倍。

2021 年, Wonkyung Jung 展示了首个基于 GPU 的 CKKS 自举实现 (CKKS 是支持近似数值算术的最有前景的 FHE 方案之一)^[32]。FHE 的一个主要缺点是其高计算成本。具体来说, 刷新密文中由连续 FHE 操作积累的噪声的自举步骤甚至可能需要几分钟的时间。这显著限制了 FHE 在许多实际应用中的使用。通过利用 FHE 中可用的大规模并行性, 该方案采用了以内存为中心的优化 (如内核融合和主要函数的重新排序), 与当时最优的 GPU 实现相比, 单个 CKKS 乘法的速度提高了 7.02 倍。

4.3 cuHE: 基于 GPU 的同态加密库

cuHE 是一个用于加速基于多项式环上同态方案计算的 CUDA GPU 库^[29], 通过代数技巧、内存最小化技术、内存和线程调度等多种优化方法, 以及底层的 CUDA 手动调优汇编优化来提高性能。特别地, cuHE 将参数选择过程调整为将同态加密评估密钥和预计算值最优地映射到正确的存储类型 (从最快和最常使用的到最慢且访问最少的)。此外, cuHE 还采用了 OpenMP 和 CUDA 混合编程, 实现在多个 GPU 上的同时计算。

cuHE 在 Prince 分组密码和同态排序等应用中取得了显著的加速效果。

(1) 完成了 Prince 同态分组密码的最快实现, 能够在 51ms (1 GPU) 内完成, 比先前的最优 GPU 实现快 25 倍。

(2) 与 CPU 实现相比, cuHE 能够以 12 ~ 41 倍的速度完成整数数组上的同态排序, 且适用于各种大小的数组。

基本的多项式算术运算，如乘法、加法和多项式模约减，cuHE 基于 CUDA GPU 的加速设计如下（针对 NVIDIA GeForce GTX 680 设备进行优化）。

①与 NTL 库的接口。大多数基于多项式的同态加密方案的实现都是基于 NTL 库构建的，cuHE 库提供了与 NTL 库数据类型相同的接口（特别是与多项式类 ZZx），以便在几乎不需要对程序进行修改的情况下，也能实现 GPU 加速。

②多项式表示。假设方案在电路的第 i 层工作，为了在 GPU 内存中存储 R_{qi} 下的一个 n 次多项式，cuHE 库使用一个由 $n \lceil (\log_2 q_i) / 32 \rceil$ 个 32 位无符号整数组成的数组，其中每 $\lceil (\log_2 q_i) / 32 \rceil$ 个整数表示一个多项式系数，将这种形式的多项式记为 RAW 域中的多项式，如表 3 所示。CRT 和 NTT 技术分别提供了多项式的 CRT 域与 NTT 域表示，每个域中的多项式结构、存储大小与操作例程如图 5 所示。由于数学属性，为了执行某个操作，多项式应该首先被转换为特定的域，除非它已经在所需的域中。

表 3 GPU 下的多项式表示

表示域	元素数据类型	元素格式
RAW	32bit 无符号整数	$n \lceil (\log_2 q_i) / 32 \rceil$
CRT	32bit 无符号整数	$t_i n$
NTT	64bit 无符号整数	$t_i n$

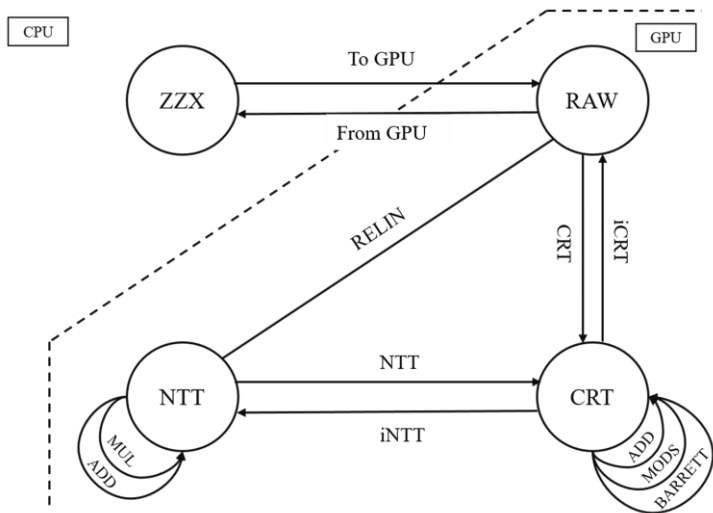


图 5 4 个域下的多项式运算与转换

③CRT/iCRT。一个简单的 CRT 内核设计是让每个线程处理一个系数 a_k ，然而，这会导致分散的全局内存访问。相反，cuHE 使用共享内存构建缓冲区，如图 6 所示。cuHE 不仅重新排列所有对全局内存的访问以保持一致性，还在读取或写入共享内存时避免了 Bank 冲突。iCRT 内核操作的设计与 CRT 类似。此外，在 iCRT 核中，cuHE 广泛使用汇编代码中的寄存器，以获得更好的性能。

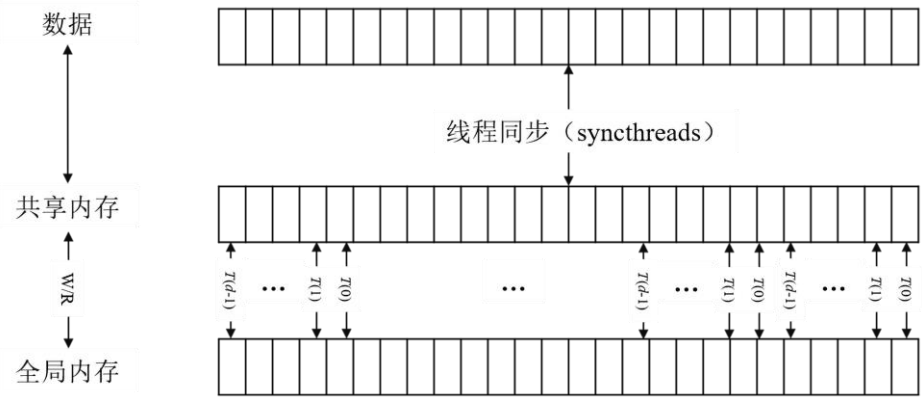


图 6 GPU 共享内存结构

④NTT/iNTT。cuHE 遵循 Dai 等人的方法在 GPU 上构建 NTT 方案。根据 FHE 方案的设置，仅支持 16384、32768 和 65536 个点的 NTT。cuHE 构建了 3 个 CUDA 内核函数，采用了四步 Cooley-Tukey 算法，通过几个较小尺寸的 NTT 来计算 N 点的 NTT。cuHE 库还在汇编代码中为 F_p 下的算术操作构建了内联设备函数。在内核函数中，使用共享内存来存储这些点，确保了一致的全局内存访问和快速的转置计算。预先计算 $2N$ 个旋转因子，并将它们绑定到纹理内存中。

⑤多项式乘法。多项式乘法需要首先将输入转换为 NTT 域，逐系数的乘法操作 DOTMUL 具有很高的并行性，非常适用于 GPU 计算。与 NTT 和 iNTT 相比，DOTMUL 在开销方面几乎可以忽略不计。

⑥多项式加法。多项式加法在同态电路中有两个重要的功能。一是在同态求解 XOR 门时使用，它可以简单地实现多项式加法。为此，加法操作在 CRT 域中执行。它为 GPU 提供足够的并行性，同时在不需要系数模约减的情况下，可以在环 R_q 中得到结果。二是在重线性化的累积部分。

⑦多项式 Barrett 约减。多项式计算在环 $R_{q_i} = \mathbb{Z}_{q_i} / m$ 上, 其中 $\deg m = n$ 。cuHE 预先计算了从模数生成的所有常数多项式, 并将它们存储在 GPU 内存中。Barrett 约减的目标是计算 $r = f(\bmod m)$ 。

为了计算一个分层电路, 除上面的底层操作之外, cuHE 还增加了其他过程来减少引入的噪声(如通过乘法)。AND 门后跟着一次重线性化操作。所有密文都经过模数切换处理, 以准备好进行下一层的计算。在 cuHE 的实现中, 密钥生成已被修改过, 以完成更快的重线性化。

4.4 基于 GPU 的快速 CKKS 自举

Wonkyung Jung 等人聚焦于全同态加密的自举步骤(刷新噪声), 以提升 FHE 方案的性能, 在 2021 年展示了首个基于 GPU 的 CKKS 自举实现(CKKS 是支持近似数值算术的最有前景的 FHE 方案之一), 并通过分析发现主要的性能瓶颈在于内存带宽^[32]。

该方案通过内存中心的优化(如内核融合和重新排序), 在 CKKS 乘法等操作中获得了显著的加速效果, 自举时间大幅降低。该实现与当时最优(SOTA)的 GPU 实现相比, 单个 CKKS 乘法的速度提高了 7.02 倍, 相当于单线程 CPU 的 257 倍加速。将其应用于逻辑回归模型训练, 与先前相同数据的 8 线程 CPU 实现相比, 实现了 40.0 倍的加速。

内核函数融合(Kernel Fusion)是将多个 GPU 内核函数合并成一个内核函数的技术。典型的 GPU 内核函数操作分为 3 个步骤: 首先它们将输入数据从 DRAM 加载到共享内存或寄存器中, 其次计算数据, 最后存储输出数据。内核函数融合将多个内核函数组合起来, 通过在寄存器或共享内存中重用数据, 减少了内核函数之间的全局内存访问次数。通过在内部和跨同态加密操作中合理地应用各种内核函数融合方法, 该方案显著减少了内存密集型同态加密操作所需的内存访问次数。

1. FHE 内运算融合

在单个同态加密操作内部, 该论文展示了 CKKS 中的主要同态加密操作(如乘法、旋转和重新缩放)有很多机会采用内核函数融合来减少内存访问次数。

下面以融合 ModUp 中的缩放内核函数与批处理 iNTT 为例进行详细讲解。

首先, 由于 iNTT 核函数的吞吐量因核函数调用开销减小而不能充分利用 GPU 中的线程级并行性, 该方案选择不再启动大量小型 iNTT 核函数, 而是将它们合并为一个大型 iNTT 核函数, 该核函数需要 $(\alpha\beta)N$ 个输入残差, 并且只需要启动一次。

其次, 由于缩放核函数的算术密度低, 变成了内存受限的情况, 将其与前面的批处理 iNTT 核函数融合在一起可以消除大部分所需的内存访问。具体来说, 由于该方案为批处理 iNTT 核

函数执行了每线程 8 次的 iNTT，每个线程在其寄存器中保存了 8 个 double-word 的输出元素，再将其存储到全局内存中。将这 8 个输出元素存储到全局内存之前，该方案将每个输出元素与其相应的 $\left[\hat{Q}_j^{''-1}\right]_{q_j}$ 或 $\left[\hat{Q}_j^{'''-1}\right]_{p_j}$ 相乘（缩放内核函数），这样总共可以消除 $2\alpha\beta N$ 次内存访问。

2. FHE 间运算融合

大部分自举时间用于计算线性变换。该论文发现 BSGS 算法的最内层求和在 GPU 上存在严重的内存带宽瓶颈，类似于内积操作，因此该方案将 CMULT 和 HADD 融合成单一的核函数 [批处理明文-密文的乘积累加（MAC）操作]，随之减少的内存访问次数能带来 11/3 倍的性能加速。

表 4 总结了单线程 CPU、该方案的 GPU 实现，以及 PrivFT^[34]（CKKS 在 GPU 上的最新实现）的同态加密操作的延迟情况。

表 4 单线程 CPU 与 GPU 下同态加密延迟对比

运算	执行时间/ms						加速比
	CPU	GPU					
	1 thread	base	fused	fused _L	fused _H	PrivFT	
N	2^{16}	2^{16}	2^{16}	2^{16}	2^{17}	2^{16}	
$\log_2 PQ$	2366	2366	2366	2364	3220	2360*	
$\log_2 Q$	2305	2305	2305	1693	2305	2300	
L	44	44	44	32	44	44	
d_{num}	45	45	45	3	3	45	
λ	98	98	98	100	128	98	
同态乘法	2644.8	33.51	17.4	2.96	7.96	55.884	7.02
旋转操作	2578.9	32.93	16.83	2.55	6.6	—	—
重线性化	145.4	0.846	0.635	0.49	1.20	1.632	1.36
同态加法	28.12	0.208	0.208	0.162	0.378	0.188	0.50
密文乘法	26.22	0.177	0.177	0.135	0.318	0.170	0.54

4.5 CARM_CUDA：逐字同态加密算法优化实现

2022 年,Shiyu Shen 提出了第一个针对异构 IoT 系统中同态乘法的 GPU 加速实现 CARM，适用于 BGV、BFV 和 CKKS 等多种同态加密方案的优化 GPU 实现^[33]。通过提供最小指令和

内存使用量的底层算术、性能和内存优先的配置，以及通用的设计，实现了高效的 RNS 同态乘法加速。在 Tesla V100S 上实现了 BGV、BFV 和 CKKS 的同态乘法速度提升分别达到 378.4 倍、234.5 倍和 287.2 倍，在 Jetson AGX Xavier 上分别实现了 8.8 倍、9.2 倍和 10.3 倍的提升。

CARM 库旨在加速中间层和云设备上的同态加密方案，以提供隐私数据的高效同态计算评估。CARM 的架构如图 7 所示，设计遵循主机-设备（HD）计算模型，其中 CPU 启动一个核函数，GPU 进行计算并返回结果。在功能上，CARM 可以分为两部分：一部分包含 GPU 内存的初始化和根据方案参数进行一些常数的预计算，如 NTT 表；另一部分包含 FHE 方案功能的实现，由低级算术层、多项式操作层和方案层 3 个层次组成。

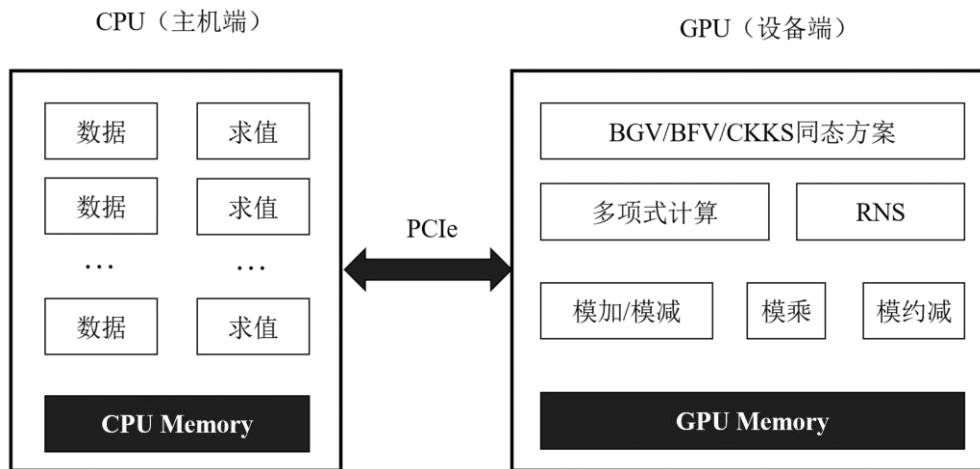


图 7 主机-设备模型与 CARM 库架构

基础层包含取模约减、加法和乘法操作。CARM 提供了在 CUDA 并行线程执行（PTX）指令中编写的数据时间实现，并通过最小化指令数量和寄存器使用进行了优化。中间层由多项式算术和 RNS 工具组成，用于加速方案函数的计算。在顶层，CARM 为 BGV、BFV 和 CKKS 提供了优化的同态加密评估核函数。此外，CARM 库支持 $|N| \in [11, 17]$ ，这涵盖了大多数应用程序的参数要求。

在 CARM 的具体设计上，专注于减少数据访问和同步操作，并提供了许多在计算成本和资源之间的权衡方案。

（1）内存中心优化。FHE 函数的计算本质上是内存受限的，因为它们具有低算术强度，但需要大量的数据访问。在这项工作中，CARM 利用内存中心优化来减少内存使用和 I/O 延迟。对于低级算术核函数，CARM 充分利用 RF 和 SMEM，并最小化临时寄存器的数量。基于此，CARM 将多个核函数融合成较大的核函数，用于高级同态加密操作，以减少数据传输，特别是 GMEM 访问。

(2) 同步。插入同步点是一种常用的适用于 GPU 的方法, 用于在存在数据交互时防止数据冲突。在整个工作中, CARM 适当安排读写数据的顺序, 以消除一些昂贵的同步。特别地, 在同一个 WARP 中的 32 个线程之间没有数据冲突, 故不需要同步。因此, CARM 通过对内部循环进行精细调整, 消除了 NTT/iNTT 中 5 个层中的线程同步。

(3) ILP 和 TLP。对于内存受限的计算, 重用数据并让一个线程执行更多操作可以增加指令级并行性 (ILP) 并减少 I/O 延迟, 从而提高整体性能。受限于 GPU 上固定的内置存储, 一个 SM 不能同时执行多个内存开销高的块, 导致一些工作组仍处于空闲状态, 从而降低了线程级并行性 (TLP)。然而, 低的 SM 占用率并不总是导致低性能, 因为将更多的数据存储在 SMEM 中可能会减少对 GMEM 的访问, 而 GMEM 的访问可能比算术操作消耗更多的周期。在 CARM 中, 选择了适合不同参数实现的 ILP 和 TLP, 从而实现了峰值核函数性能。

(4) 对于多项式运算模块。通过并行执行 NTT 中每层的蝶件从而达到快速实现 NTT 变换, 然而这种方式耗费的线程资源过多, 因此采用分层分级的方式, 将一次完整的 NTT 划分为两个 kernel, 减少对 GPU 内存数据的存取次数, 通过更快速的共享内存达到更高的效率。在每个 kernel 中, 使用 256 个线程进行实例化, 每 8 个线程组合为 1 个 block, 在 block 中可以通过与 RF 和 SMEM 的交互来提高访存效率。在同态乘法中, 一次完整的乘法往往包含多项式的 NTT 变换、标量点乘及 iNTT 变换, 通过与 GMEM 及 SMEM 的交互完成多项式的 NTT 变换, 将标量乘法融合到 RF 中进行运算, 从而降低点乘所需要的资源, NTT 与 iNTT 的 kernel 实现流程如图 8 所示。

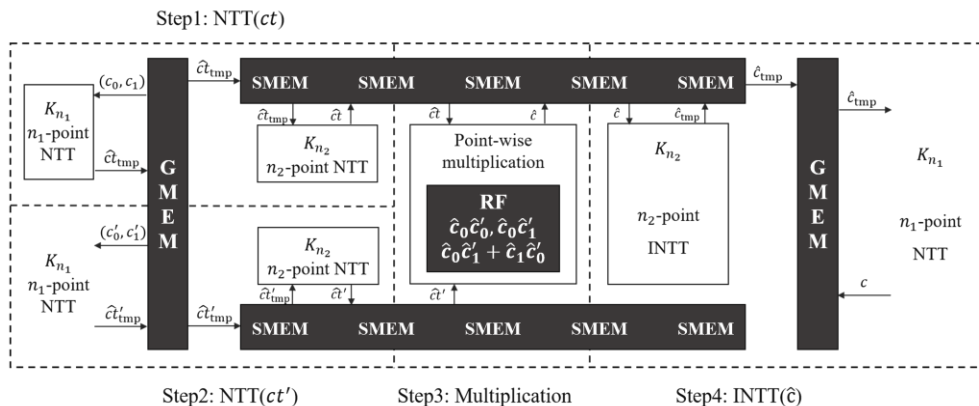


图 8 NTT/iNTT 的 kernel 实现流程

4.6 方案对比

表 5 给出了基于 GPU 同态加密优化方案对比。

表 5 基于 GPU 同态加密优化方案对比

GPU 实现	优化方案	性能	主要优化手段
方案[27]	Prince 分组密码；同态排序	Prince 同态分组密码在 51ms（1 GPU）内完成：25 倍 同态排序速度：12 ~ 41 倍	代数技巧、内存最小化技术、内存和线程调度等多种优化方法，以及底层的 CUDA 手动调优汇编优化
方案[28]	CKKS（Bootstrapping）	单个 CKKS 乘法的速度：7.02 倍	内存中心的优化（如内核融合和重新排序）
方案[29]	BGV/BFV/CKKS	（Tesla V100S 上的 RNS 同态乘法速度） BGV: 378.4 倍 BFV: 234.5 倍 CKKS: 287.2 倍	最小化指令数量和寄存器使用； 内核函数融合； 消除 NTT/iNTT 的线程同步

5 面向硬件平台的全同态高速实现

5.1 国内外发展现状

2019 年，Roy 等人利用硬件并行计算资源设计硬件加速架构并最终实现基于 Arm-FPGA 的联合架构^[25]，该实现引入了基于 NTT 并行计算的方法并提出改进，即使用片上存储器存储常数旋转因子以减少算法执行周期。该实现相比软件实现有 13 倍的效率提升。

2023 年，Agrawal 等人提出了 FAB 实现^[26]，该实现基于 FPGA 加速了全同态的自举操作，可用于 CKKS 算法。该实现平衡了存储代价和计算代价，充分利用了片上的存储空间，在仅使用 43 MB 片上 U/BRAM 的情况下，实现了高效的多项式系数读取操作。

2022 年, Su 等人提出了 MCNA 架构^[27], 该架构引入了无预处理的 NTT 算法实现与 Cyclic-Sharing 技术, 从而大幅减少了 NTT 运算的延迟与存储空间。且 MCNA 设计了可重构的 NTT 架构, 可以根据需要选择模乘算法, 进一步减少了硬件实现代价。

2022 年, Samardzic 等人所提出的 CraterLake 实现支持 CKKS、BGV 和 GSW 算法^[35], 该实现采用了 Boosted Keyswitching 技术与 Vector Chaining 技术, 并设计了转置网络技术加速 NTT 计算。相比较 CPU 上的软件实现, CraterLake 可达到 4600 倍的效率提升。

2023 年, Geelen 等人提出的 BASALISC 实现为 BGV 算法提供了加速支持^[36], BASALISC 采用了 NTT 友好的自举技术大幅加速了同态计算的效率, 并采用了三层 ISA 与四层存储结构的设计, 避免了访存冲突。BASALISC 在性能和资源消耗之间取得了很好的平衡, 相比较 HELib 库可达到 5000 倍的速度提升。

5.2 基于 FPGA 的同态加密高性能并行架构实现

Roy 等人提出的实现架构可为 BFV 算法提供加速^[25], 该实现为 ARM 与 FPGA 的混合架构, 该实现在 Xilinx Zynq UltraScale+ MPSoC ZCU102 开发套件上进行测试, FPGA 频率为 200MHz。为了减少时钟周期数, 该实现在块级和电路级分别使用不同的并行策略对任务调度进行了优化。

该实现采用了近似 CRT 技术^[37], 对于 $a \equiv \sum(a_i \cdot \tilde{q}_i \setminus b \bmod q_i) \cdot q_i^* - v' \cdot q$, $v' \equiv (\sum(a_i \cdot \tilde{q}_i \setminus b \bmod q_i) \cdot q_i^*) / q$, 可简化计算 $v' = (\sum(a_i \cdot \tilde{q}_i \setminus b \bmod q_i)) / q_i$, 由于 a_i 、 q_i 和 $\tilde{q}_i$ 都为 30bit 整数, 误差可控制在 2^{-53} bit 以内, 因此该近似计算对于同态计算的正确性没有影响。

NTT 计算核心部件中, 模约减电路为实现代价最为高昂的计算单元, 每个计算单元支持两个素数的模运算, 采用了 Barrett 模约减算法, 并采用了滑动窗口法对整数进行约减, 大幅提升了模块实现效率。

5.3 FAB: 平衡存储与计算代价的高效实现

2023 年, Agrawal 等人提出了 FAB 实现^[26], 该实现基于 FPGA 加速了全同态自举操作, 并提出了平衡存储和计算代价的硬件设计。FAB 可加速 CKKS 算法, 运行频率为 300MHz, 相较 CPU 实现拥有 456 倍的速度提升。

FAB 通过高效利用 FPGA 片上存储的 U/BRAM 块, 大幅提升了多项式系数的访存带宽。FAB 将多项式系数按比特映射至 U/BRAM 块, 在仅使用 43MB 片上存储容量情况下, 满足了 FHE 底层操作的带宽需求, 得到了更平衡的 FPGA 硬件设计。

同时, FAB 在 NTT 与 iNTT 算法中统一使用了 CT 蝶件, 从而可以重复利用其数据映射逻辑。此外, FAB 通过片上 URAM 存储读取旋转因子, 得到了更高效的 NTT/iNTT 实现。如表 6 所示, FAB 实现所需的模乘器数量、Register files 与片上存储比 BTS 实现与 F1 实现均大幅较低。

表 6 模乘器数量与存储空间对比

实现	模乘器/个	Register files/MB	片上存储/MB
F1	18432	8	64
BTS	8192	22	512
FAB	256	2	43

5.4 MCNA: 可重构的 NTT 架构

2022 年, Su 等人提出了 MCNA 架构^[27], 该架构基于 Xilinx Virtex-7 平台实现, 运行频率为 250 MHz, 该架构拥有极小的硬件实现面积与可重构的 NTT/iNTT 设计, 相比之前工作拥有 2.78~9.32 倍的效率提升。

MCNA 提出了一种无预处理和后处理的 NTT/iNTT 实现。Su 等人采用了 Constatn-Geometry NTT (CG NTT) 实现, 如算法 5-1 所示, CG NTT 为非原地实现的 NTT 算法, 可简化多核架构中的访存模式, 提升实现效率。并且 MCNA 将预处理与后处理和旋转因子读取结合, 进一步降低了计算复杂度。并且在 NTT 实现中, MCNA 采用了可重构的模乘器, 可根据参数的输入决定使用 Barrett 约减算法或 Shoup 算法^[39]。这一可重构特性使得 MCNA 可重用硬件资源, 降低了硬件实现的代价, 进一步减少了硬件面积。

算法 5-1: 无预处理的 CG NTT 算法

Input:

$A, N, q; \omega_{2N}^{-i}$, where $i = 1, 2, \dots, N-1$

Output:

$$a = \text{iNTT}_{\omega_N^{-1}}^N(A)$$

1: for ($s \leq \log_2 N, s = s + 1$) do

2: for ($j = 0; j < N / 2; j = j + 1$) do

```

3:       $k_1 = \lfloor j / 2^{s-1} \rfloor \cdot 2^{s-1}$ 
4:       $k_2 = 2^{s-1}$ 
5:       $a[2j] \leftarrow (A[j] + A[j + N/2]) / 2 \bmod q$ 
6:       $a[2j+1] \leftarrow ((A[j] - A[j + N/2]) \cdot \psi_{2N}^{-(2k_1+k_2)}) / 2 \bmod q$ 
7:  end for
8:  if  $s \neq \log_2 N$  then
9:       $A = a$ 
10: end if
11: end for
12:  $a \leftarrow \text{BitReverse}(a)$ 
13: return  $a$ 

```

MCNA 提出了 Cyclic-Sharing 技术, 该技术避免了 CG NTT 中的独立访存, 统一了 NTT 和 iNTT 的读写结构。同时, 该技术减少了 25% 的 MCNA 内存占用。如图 9 所示, Cyclic-Sharing 技术通过组织数据在 6 个 Bank 中的读取来减少访存冲突, 且该技术使得 MCNA 具有更好的可扩展性和更低的计算复杂度。

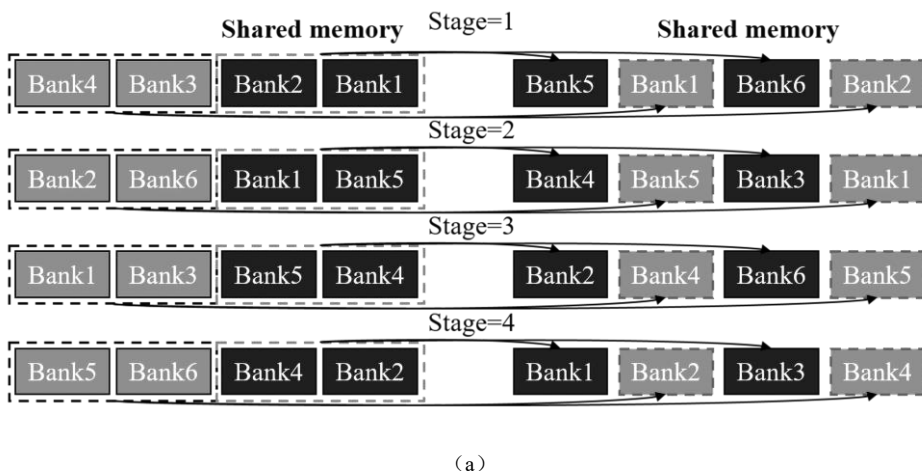
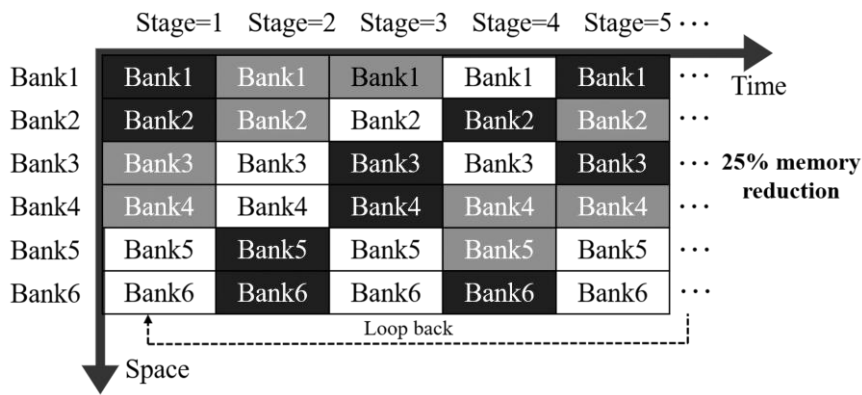


图 9 Cyclic-Sharing 访存模式



(b)

图 9 Cyclic-Sharing 访存模式（续）

5.5 CraterLake：高性能同态加密硬件加速器

Samardzic 等人所提出的 CraterLake 实现支持 CKKS、BGV 和 GSW 算法的硬件加速器。该实现采用 12/14 nm 工艺，芯片面积为 472.3 mm²，频率为 1 GHz。

CraterLake 采用了 Boosted Key Switching 技术，该技术的实现流程如算法 5-2 所示，该实现以密文多项式和 KeySwitching Hint (KSH)作为输入，输出两个密文多项式。CraterLake 通过 Boosted KeySwitching 技术减少了 10%的 NTT 运算数量，并减少了 50%的乘法与加法运算数量，并通过该技术与一种新的 changeRNSBase()函数实现结合，寄存器文件的使用量减少到原来的 1/60。

算法 5-2: Boosted 密钥转换算法

```
def boostedKeySwitch(p[0:L]):  
    pTmp[0:L] = iNTT(p[0:L])  
    pTmp[L:2L] = changeRNSBase(pTmp[0:L],[L:2L])  
    p[L:2L] = NTT(pTmp[L:2L])  
    for i=0,1:  
        prodi[0:2L]=p[0:2L] *KSHi[0:2L]  
        tmpi[0:2L] = iNTT(prodi [L:2L],[0:L])  
        mDTmpi[0:L] = changeRNSBase(tmpi[L:2L],[0:L])
```

```

modDowni[0:L] = NTT(mDTmpi[0:L])

ksi[0:L]=prodi[0:L] +modDowni[0:L]

return (kso[0:L],ksl[0:L])

def changeRNSBase(x[0:L], destModIdxList):

    for srcModIdx in [0:L]:

        for destModIdx in destModIdxList:

            C = constant[srcModIdx] [destModIdx]

            result[destModIdx] += x[srcModIdx] * C

    return result

```

CraterLake 通过引入新的转置网络技术加速 NTT 和同态计算,该网络无须任何控制单元。该转置网络采用了双层的转置架构,对于一个 $G \times G$ 的矩阵,第 i 行的数据通过通道内的本地转置实现,该转置过程为全流水线操作。而对于 $G \times G$ 的矩阵整体,转置过程需要跨通道进行,则通过一个固定的置换单元进行操作。相比较 F1 的实现,CraterLake 约减少了 1/16 的芯片面积与 5/12 的带宽。

CraterLake 通过引入向量链技术进一步提升了实现效率。如图 10 所示,该技术将多种函数单元串行执行,大幅减少了寄存器文件的使用。该技术结合 Boosted KeySwitching 算法可将操作数在生成后立即被使用,寄存器文件的使用量减少了约 3/4。

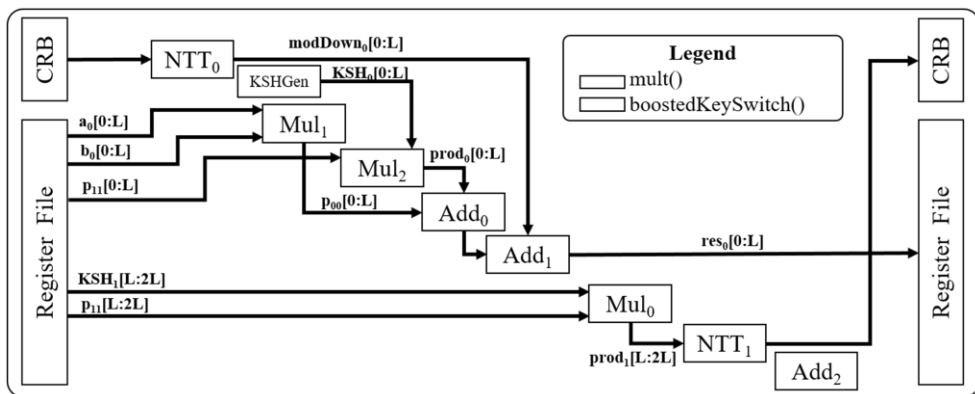


图 10 向量链流程

5.6 BASALISC：可编程的同态加密硬件加速器

BASALISC 为 BGV 算法的硬件加速器^[36]，该实现方案首次使用纯硬件实现了 fully-packed 的自举，相比较 HELib 有着 5000 倍速率的提升。该实现采用类 RISC 架构，以及三层指令集架构（Instruction Set Architecture, ISA）与四层存储结构。实现采用了 12 nm Global Foundries 工艺，芯片面积为 150mm²，频率为 1GHz。

该实现的三层指令集架构中 Macro-Instructions 层负责密文加法、乘法、模运算、密钥转换、自举、自同构等功能，Mid-level Instructions 层负责多项式点乘与加法、NTT、乘法累加、存储管理等功能，Micro-Instructions 层负责 PCIe 接口与机器控制指令。该指令集架构允许开发者在不同的抽象层次上进行调试。

为了解决全同态加密的访存性能瓶颈，BASALISC 采用了四层具有不同延迟的存储架构，其中片外远程存储器（DRAM）为 256GB，负责存储需要处理的数据与计算完成的数据，中间内存密文缓冲区（CTB）为 64MB，存储 32bit 长度的多项式系数，另外两层为局部寄存器文件（local Register File, RF）和累加寄存器（Accumulator register, ACC）。

BASALISC 采用了 NTT 友好的自举技术，该技术依赖于 Geelen 与 Vercauteren 所提出的简化解密引理^[38]：

引理（简化解密^[38]） p 为素数，令 $e > r \geq 1$ 并且 q 为足够大的参数使得 $\gcd(q, p) = 1$ 。如果 (c_0, c_1) 是 BGV 加密的密文，且明文 m 模 p^r 、密文模 q ，那么可以通过如下方式解密，其中 $q^{-1} \cdot q = 1 \pmod{p^e}$ ， $\lfloor \cdot \rfloor$ 表示系数取整。

$$c'_i \leftarrow \lfloor p^{e-rc_i} \rfloor_q, \quad w \leftarrow \lfloor q^{-1} \cdot (c'_0 + c'_1 \cdot s) \rfloor_{p^e}, \quad m \leftarrow \lfloor q \cdot \lfloor w / p^{e-r} \rfloor \rfloor_{p^r}$$

并且 BASALISC 引入了 FastBaseExt 与 SmallMont 两个原语对该引理进行实现，其中

$$\text{FastBaseExt}_{Q \rightarrow QP}(a) = \left(\left[\sum_{i=1}^k \left[a \cdot \left(\frac{Q}{q_i} \right)^{-1} \right] \cdot \frac{Q}{q_i} \right]_{p_i} \right)_{1 \leq j \leq l} \in \mathcal{R}_{QP}$$

令 $c = \text{SmallMont}_{QP \rightarrow Q}(b, m)$ ，对于 $b \in \mathcal{R}_{QP}$ ，且 $\|b\|_\infty \ll P \cdot m$ ，输出 $c = b \cdot P^{-1} \pmod{m}$ ， $c \in \mathcal{R}_Q$ 。

该算法的详细实现如算法 5-3 所示。

算法 5-3: NTT 友好自举算法

Input:

$ct \in \mathcal{R}_q^2$, 噪声 e

Output:

输出: $ct'' \in \mathcal{R}_q^2$, 噪声 e'' , 使得 $\text{Dec}(ct'') = \text{Dec}(ct), \|e''\|_\infty / q'' \ll \|e\|_\infty / q$

1: 对 $i = 0, 1$ 计算

2: $c_i'' \leftarrow [c_i \cdot p^{e-r} \cdot b]_q$

3: $c_i'' \leftarrow \text{FastBaseExt}_{q \rightarrow Qqb}(c_i'')$

4: $c_i' \leftarrow \text{SmallMont}_{Qqb \rightarrow Qq}(c_i'', q)$

5: $c_i'' \leftarrow \text{SmallMont}_{Qq \rightarrow Q}(c_i', p^e)$

6: $ct'' \leftarrow \lfloor ct'' / p^{e-r} \rfloor$

7: 返回 ct''

5.7 方案对比

表 7 所示为 CraterLake、F1 与 CPU 在执行不同神经网络或操作的性能对比, 测试采用 CKKS 方案, 并在 4 个深度测试方案与 4 个浅度方案上进行测试, 深度方案具有较深的乘法深度并需要自举操作, 而浅度测试乘法深度较低并无须自举操作。可以看到, 相较 F1 方案, Crater Lake 在深度方案上拥有约 11.2 倍的效率提升, 而在浅度方案上拥有大概 1.34 倍的效率提升。相较 CPU 上的实现, CraterLake 分别达到 4611 倍和 5220 倍的效率提升。

表 7 CraterLake、F1 与 CPU 的性能对比

操作	CraterLake	F1	CPU	vs. F1	vs. CPU
ResNet-20	249.45ms	2693ms	23min	10.8	5532
逻辑回归 (Logistic Regression)	119.52ms	639ms	356s	5.35	2979
长短期记忆网络 (LSTM)	138.00ms	2573ms	859s	18.6	6225
打包自举 (Packed Bootstrapping)	3.91ms	58.3ms	17.2s	14.9	4399
Deep Gmean Speed-up				11.2	4611

续表

操作	CraterLake	F1	CPU	vs. F1	vs. CPU
未打包自举 (Unpacked Bootstrapping)	0.10ms	0.21ms	877ms	2.1	8770
CIFAR Unencryp. Wghts	50.50ms	94.1ms	187s	1.86	3703
MNIST Unencryp. Wghts	0.14ms	0.13ms	561ms	0.93	4007
MNIST Encryp. Wghts	0.24ms	0.22ms	1369ms	0.92	5704
Shallow Gmean Speed-up				1.34	5220

表 8 所示为 BASALISC 与 HELib 性能对比,HELib 运行在 Intel Xeon E5-2630 v2 CPU @2.6 GHz 平台上, BASALISC 中 NTT 效率相比较 CPU 上的软件性能有 2500 倍的提升,而数据库查询操作有 8700 倍的效率提升,逻辑表达式的计算有 5400 倍的提升。

表 8 BASALISC 与 HELib 性能对比

操作	HELib	BASALISC	提升
NTT	27ms	11μs	2.5×10^3
Add/Sub	4ms	8μs	5.0×10^2
Plaintextmul	44ms	5μs	8.8×10^3
Mul (无 key switch)	58ms	20μs	2.9×10^3
Permutation (无 key switch)	12ms	11μs	1.1×10^3
Key Switching	580ms	292μs	2.0×10^3
Database Lookup	2325s	267ms	8.7×10^3
Thin Bootstrapping	160s	40ms	4.0×10^3
Logistic Regression	217000s	40500ms	5.4×10^3

表 9 所示为 BASALISC 与 F1、BTS^[40]、CraterLake 性能对比,相比较 BASALISC, BTS 实现无法计算 BGV 自举的所有自同构。BASALISC 相较于 F1 提供了无冲突的 CTB 与 NTT 实现。而对比 CraterLake 的实现, BASALISC 无须额外的单元或内存进行转置,更为高效。

表 9 BASALISC 与 F1、BTS、CraterLake 性能对比

实现	BASALISC	F1	BTS	CraterLake
算法	BGV	BGV/CKKS	CKKS	CKKS
电路面积	150mm ²	150mm ²	374mm ²	472mm ²

续表

实现	BASALISC	F1	BTS	CraterLake
制程	12nm	12/14nm	7mm	12/14nm
功耗	115W	180W	163W	320W
Bootstrapping 提升	4000	1830/1195	2237	4400
LR 提升	5400	-77200	1306	2978

表 10 所示为各方案的对比,其中 FAB 和 MCNA 为 FPGA 实现,而 CraterLake、BASALISC 为 ASIC 实现,其中性能对比仅展现各个方案所做出的性能提升,不作为不同方案之间性能的横向比较。相比 CPU 平台的实现,硬件实现方案均有较为显著的性能提升。

表 10 硬件实现方案对比

实现	优化方案	性能*	平台
方案[25]	BFV	相比 CPU 提升 13 倍	ARM-FPGA
FAB	Bootstrapping (CKKS)	相比 CPU 提升 456 倍	FPGA
MCNA	NTT	相比 CPU 提升 13 倍	FPGA
CraterLake	CKKS/BGV/GSW	相比 CPU 提升 4611~5220 倍	ASIC
BASALISC	BGV	相比 CPU 提升 5400~8700 倍	ASIC

注：*所对比的 CPU 平台性能、库与执行操作均不相同。

本文通过对全同态加密（FHE）的各种加速方法进行深入分析和比较，从多个角度全面探讨了同态加密加速的研究方向。本文的主要目的是为同态加密加速领域的研究人员提供一个清晰、全面、深入的视角，以利于更好地理解和应用同态加密技术，并为同态加密优化实现与实际应用研究提供有价值的指导和支持。

参考文献

[1] KARATSUBA A, OFMAN Y. Multiplication of multiplace numbers on automata[C]//Dokl. Akad. Nauk SSSR. 1962, 145(2): 293-294.

[2] MONTGOMERY P L. Modular multiplication without trial division[J]. Mathematics of computation, 1985, 44(170): 519-521.

[3] BARRETT P. Implementing the Rivest Shamir and Adleman public key encryption algorithm on a standard digital signal processor[C]//Conference on the Theory and Application of Cryptographic Techniques. Berlin, Heidelberg: Springer Berlin Heidelberg, 1986: 311-323.

[4] SMART N P, VERCAUTEREN F. Fully homomorphic SIMD operations[J]. Designs, codes and cryptography, 2014, 71: 57-81.

[5] GENTRY C, HALEVI S, SMART N P. Fully homomorphic encryption with polylog overhead[C]//Annual International Conference on the Theory and Applications of Cryptographic Techniques. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 465-482.

[6] MYERS S, SHULL A. Practical revocation and key rotation[C]//Topics in Cryptology-CT-RSA 2018: The Cryptographers' Track at the RSA Conference 2018, San Francisco, CA, USA, April 16-20, 2018, Proceedings. Springer International Publishing, 2018: 157-178.

[7] LIU F H, WANG H. Batch bootstrapping II: bootstrapping in polynomial modulus only requires $O(1)$ FHE Multiplications in amortization[C]//Annual International Conference on the Theory and Applications of Cryptographic Techniques. Cham: Springer Nature Switzerland, 2023: 353-384.

[8] LIU F H, WANG H. Batch bootstrapping I: a new framework for simd bootstrapping in polynomial modulus[C]//Annual International Conference on the Theory and Applications of Cryptographic Techniques. Cham: Springer Nature Switzerland, 2023: 321-352.

[9] KIM S, JUNG W, PARK J, et al. Accelerating number theoretic transformations for bootstrappable homomorphic encryption on GPUs, in Proc. IEEE Int. Symp. Workload Characterization (IISWC), Oct. 2020, pp. 264-275.

[10] NATARAJAN D, DRESLINSKI R. Performance characterization of latticebased cryptography workloads, in Proc. Int. Symp. Perform. Anal. Syst. Softw., Aug. 2020, pp. 220-222.

[11] AGUILAR-MELCHOR C, BARRIER J, GUELTON S, et al. NFLlib: NTT-based fast lattice library. In Cryptographers' Track at the RSA Conference. Springer, 2016: 341-356.

[12] VAN DER HOEVEN J, LECERF G, QUINTIN G. Modular SIMD arithmetic in Mathemagix. ACM Transactions on Mathematical Software (TOMS), 2016, 43(1): 1-37.

[13] EDAMATSU T, TAKAHASHI D. Accelerating Large Integer Multiplication Using Intel AVX-512IFMA. In International Conference on Algorithms and Architectures for Parallel Processing. Springer, 2019: 60-74.

[14] GUERON S, KRASNOV V. Accelerating big integer arithmetic using intel IFMA extensions. In 2016 IEEE 23rd Symposium on Computer Arithmetic (ARITH). IEEE, 2016: 32-38.

- [15] DRUCKER N, GUERON S. Fast modular squaring with AVX512IFMA. In 16th International Conference on Information Technology-New Generations (ITNG 2019). Springer, 2019:3-8.
- [16] JUNG W, LEE E, KIM S, et al. Accelerating fully homomorphic encryption through architecture-centric analysis and optimization. IEEE Access, 2021(9): 98772-98789.
- [17] BOEMER F, KIM S, Seifu G, et al. Intel HEXL: Accelerating Homomorphic Encryption with Intel AVX512IFMA52. In Proceedings of the 9th on Workshop on Encrypted Computing & Applied Homomorphic Cryptography (WAHC '21). Association for Computing Machinery, New York, NY, USA, 2021: 57-62. isbn: 9781450386562.
- [18] HALEVI S, SHOUP V. Algorithms in Helib, in Advances in Cryptology-CRYPTO (Lecture Notes in Computer Science), vol. 8616. Berlin, Germany: Springer 2014, pp. 554-571.
- [19] CHEN H, DAI W, KIM M, et al. Efficient homomorphic conversion between (ring) LWE ciphertexts, Cryptol. ePrint Arch., Tech. Rep. 2020/015, 2020.
- [20] CHILLOTTI I, GAMA N, GEORGIEVA M, et al. Faster packed homomorphic operations and efficient circuit bootstrapping for TFHE, in Advances in Cryptology-ASIACRYPT (Lecture Notes in Computer Science), vol. 10624. Hong Kong: Springer, 2017, pp. 377-408.
- [21] HALEVI S, Shoup V. Faster homomorphic linear transformations in Helib, in Advances in Cryptology-CRYPTO (Lecture Notes in Computer Science), vol. 10991. Santa Barbara, CA, USA: Springer, 2018, pp. 93-120.
- [22] BOSSUAT J-P, MOUCHET C, TRONCOSO-PASTORIZA J, et al. Secure and efficient bootstrapping for approximate homomorphic encryption, Cryptol. ePrint Arch., Tech. Rep. 2020/1203, 2020.
- [23] ISHIMAKI Y, YAMANA H. Faster Homomorphic Trace-Type Function Evaluation. IEEE Access, 2021(9): 53061-53077.
- [24] INOUE K, SUZUKI T, YAMANA H. Acceleration of Homomorphic Unrolled Trace-Type Function using AVX512 instructions. WAHC@CCS 2022: 47-52.
- [25] ROY S S, TURAN F, JARVINEN K, et al. FPGA-based high-performance parallel architecture for homomorphic computing on encrypted data, in Proc. IEEE Int. Symp. High Perform. Comput. Archit. (HPCA), Feb. 2019, pp. 387-398.
- [26] AGRAWAL R, DE CASTRO L, YANG G, et al. FAB: An FPGA-based accelerator for bootstrappable fully homomorphic encryption[C]//2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA). IEEE, 2023: 882-895.
- [27] SU Y, YANG B L, YANG C, et al. A highly unified reconfigurable multicore

architecture to speed up NTT/INTT for homomorphic polynomial multiplication[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2022, 30(8): 993-1006.

[28] HAN K, KI D. Better bootstrapping for approximate homomorphic encryption, in Topics in Cryptology-CT-RSA (Lecture Notes in Computer Science), vol. 12006. Cham, Switzerland: Springer, 2020, pp. 364-390.

[29] DAI W, SUNAR B. cuHE: A homomorphic encryption accelerator library[C]// Cryptography and Information Security in the Balkans: Second International Conference, BalkanCryptSec 2015, Koper, Slovenia, September 3-4, 2015, Revised Selected Papers 2. Springer International Publishing, 2016: 169-186.

[30] AL BADAWI A, POLYAKOV Y, AUNG K M M, et al. Implementation and performance evaluation of RNS variants of the BFV homomorphic encryption scheme[J]. IEEE Transactions on Emerging Topics in Computing, 2019, 9(2): 941-956.

[31] ÖZERK Ö, ELGEZEN C, MERT A C, et al. Efficient number theoretic transform implementation on GPU for homomorphic encryption[J]. The Journal of Supercomputing, 2022, 78(2): 2840-2872.

[32] 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: 114-148.

[33] SHEN S, YANG H, LIU Y, et al. CARM: CUDA-Accelerated RNS Multiplication in Word-Wise Homomorphic Encryption Schemes for Internet of Things[J]. IEEE Transactions on Computers, 2022.

[34] AL BADAWI A, HOANG L, MUN C F, et al. Privft: Private and fast text classification with homomorphic encryption[J]. IEEE Access, 2020, 8: 226544-226556.

[35] 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. 2022: 173-187.

[36] GEELEN R, VAN BEIRENDONCK M, PEREIRA H V L, et al. Basalisc: Programmable asynchronous hardware accelerator for bgv fully homomorphic encryption[J]. Cryptology ePrint Archive, 2022.

[37] HALEVI S, POLYAKOV Y, SHOUP V. An improved RNS variant of the BFV homomorphic encryption scheme[C]//Topics in Cryptology-CT-RSA 2019: The Cryptographers' Track at the RSA Conference 2019, San Francisco, CA, USA, March 4-8, 2019, Proceedings. Springer International Publishing, 2019: 83-105.

[38] GEELEN R, VERCAUTEREN F. Bootstrapping for BGV and BFV Revisited[J]. Journal of Cryptology, 2023, 36(2): 12.

[39] SHOUP V. NTL: A library for doing number theory[EB/OL]. 2001. <http://www.shoup.net/ntl/>.

[40] 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. 2022: 711-725.