软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2021,32(4):1165−1185 [doi: 10.13328/j.cnki.jos.006165]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

后量子密码算法的侧信道攻击与防御综述*

吴伟彬, 刘 哲, 杨 昊, 张吉鹏

(南京航空航天大学 计算机科学与技术学院,江苏 南京 211106)

通讯作者: 刘哲, E-mail: zhe.liu@nuaa.edu.cn

摘 要: 为了解决量子计算对公钥密码安全的威胁,后量子密码成为密码领域的前沿焦点研究问题.后量子密码通过数学理论保证了算法的安全性,但在具体实现和应用中易受侧信道攻击,这严重威胁到后量子密码的安全性.基于美国 NIST 第 2 轮候选算法和中国 CACR 公钥密码竞赛第 2 轮的候选算法,针对基于格、基于编码、基于哈希、基于多变量等多种后量子密码算法进行分类调研,分析其抗侧信道攻击的安全性现状和现有防护策略.为了深入分析后量子密码的侧信道攻击方法,按照算法核心算子和攻击类型进行分类,总结了针对各类后量子密码常用的攻击手段、攻击点及攻击评价指标.进一步地,根据攻击类型和攻击点,梳理了现有防护策略及相应的开销代价.最后,根据攻击方法、防护手段和防护代价提出了一些安全建议,并且还分析了未来潜在的侧信道攻击手段与防御方案.

关键词: 后量子密码;侧信道攻击;故障攻击;能量分析攻击;时间攻击

中图法分类号: TP309

中文引用格式: 吴伟彬,刘哲,杨昊,张吉鹏.后量子密码算法的侧信道攻击与防御综述.软件学报,2021,32(4):1165–1185. http://www.jos.org.cn/1000-9825/6165.htm

英文引用格式: Wu WB, Liu Z, Yang Hao, Zhang JP. Survey of side-channel attacks and countermeasures on post-quantum cryptography. Ruan Jian Xue Bao/Journal of Software, 2021,32(4):1165−1185 (in Chinese). http://www.jos.org.cn/1000-9825/6165.htm

Survey of Side-channel Attacks and Countermeasures on Post-quantum Cryptography

WU Wei-Bin, LIU Zhe, YANG Hao, ZHANG Ji-Peng

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: To solve the threat of quantum computing to the security of public-key cryptography, post-quantum cryptography has become a frontier focus in the field of cryptography. Post-quantum cryptography guarantees the security of the algorithm through mathematical theories, but it is vulnerable to side-channel attacks in specific implementation and applications, which will seriously threaten the security of post-quantum cryptography. This study is based on the round 2 candidates in the NIST post-quantum cryptography standardization process and the round 2 candidates in the CACR public key cryptography competition in China. First, classification investigations of various post-quantum cryptographic algorithms are conducted, including lattice-based, code-based, hash-based, and multivariate-based cryptographic algorithms. Then, their security status against side-channel attacks and existing protection strategies are analyzed. To analyze the methods of side-channel attack against post-quantum cryptography, it is summarized that the commonly used post-quantum cryptography side-channel attack methods, attack targets, and attack evaluation indexes for various post-quantum cryptography according to the classification of core operators and attack types. Furthermore, following the attack types and attack targets, the existing countermeasures for attack and the costs of defense strategies are sorted out. Finally, in the conclusion part, some security suggestions are put forward according to the attack method, protection means, and protection cost, and also the potential side-channel attack methods and

* 基金项目: 国家自然科学基金(61802180); 江苏省自然科学基金(BK20180421); 国家密码发展基金(MMJJ20180105); 中央高校基础研究基金(NE2018106)

Foundation item: National Natural Science Foundation of China (61802180); Natural Science Foundation of Jiangsu Province of China (BK20180421); National Cryptography Development Fund (MMJJ20180105); Fundamental Research Funds for the Central Universities (NE2018106)

收稿时间: 2020-05-23; 修改时间: 2020-07-02, 2020-08-14; 采用时间: 2020-10-19; jos 在线出版时间: 2021-01-22

defense strategies in the future are analyzed.

Key words: post-quantum cryptography; side-channel attack; fault attack; power analysis attack; timing attack

密码学是网络安全的基石,目前业界主流的传统公钥密码算法主要基于大整数分解问题和离散对数问题,但这两类数学难题都已被 Shor 算法在多项式时间内所攻破[1].为了应对量子计算对公钥密码算法的威胁,全球的密码学者都在积极开展后量子密码算法(post-quantum cryptography,简称 PQC)的研究.后量子密码在设计时,设计者必须从理论上证明该算法的安全性,但密码算法的安全性除了需要证明理论安全以外,还需要考虑算法具体实现上的安全,即物理安全.侧信道攻击能够利用密码算法在密码芯片上运行时泄露出来的侧信息,分析并恢复出密钥或加密信息,是密码算法物理安全的主要威胁手段.

近些年,已有部分学者针对后量子密码算法的侧信道攻击与防御开展研究现状的调研工作,如:2016 年,Bindel 和 Buchmann 等人[2]分析了基于格的签名方案面对多种不同类型(随机化、归零和跳过)的一阶故障攻击,并根据对故障攻击的分析提出了 6 种不同类型攻击相应的防御对策,但该工作的调研只限于故障攻击;2018 年,Khalid 和 Rafferty 等人[3]针对格基密码方案中的重要组件(错误采样器)进行调查研究,根据各种错误采样器存在的侧信道威胁,提出了错误采样器的安全建议,但是,该工作只限于采样器;2018 年,文献[4]调查了基于格的侧信道攻击(SCA)方面的相关工作,包括入侵性攻击和被动攻击以及已提出的防御对策,但是该工作没有针对攻击点和防御代价进行分析;文献[5]调查了多种后量子密码方案在不同微控制器的实现效果,包括对公私钥大小、签名验签时间和加密解密时间等信息的分析,但没有针对可能存在的侧信道攻击进行调研;又如,2018 年,文献[6]分析了 R-LWE 加密算法抵御故障攻击的能力,并且根据攻击点和手段的分析,提出了加强 RLWE 的防御对策;同年,文献[7]就基于编码的后量子密码方案进行了调研,从理论安全到物理安全都进行了分析;2019 年,文献[8]调查了基于格的密码方案的发展情况,分析和总结了在软件和硬件中实现格密码的挑战与建议;这些调研只针对一类或者其中一个组件进行分析,没有针对不同类型密码和不同侧信道攻击手段的安全性进行对比分析,因此本文针对不同类型的后量子密码系统进行了调研和深入的分析.

本文针对 NIST 最新公布的二轮候选后量子密码算法和中国 CACR 公钥密码竞赛第 2 轮入选算法进行了调研,分析了最新的侧信道安全研究进展和防御策略,主要工作分为 3 点.

(1) 将 NIST 第 2 轮后量子密码的不同类型作为研究对象,同时参照中国密码学会 CACR 竞赛入选的后量子算法类型,整理与归纳这些密码现有的侧信道安全性水平和可能的防御能力.

(2) 针对现有后量子侧信道攻击方法,汇总不同信道与不同方式的攻击方法,并按照算法核心算子进行分类,给出攻击结果的评价指标,从攻击方法、攻击点、攻击评价等多个角度进行侧信道攻击的深入分析.

(3) 针对现有后量子侧信道防护方法,整理不同对抗策略的应用方式,调研最易遭受攻击的核心算子的防护策略以及现有防御策略的代价瓶颈,从安全性和防护代价两个维度开展分析.

由于提交到 NIST 中的基于超奇异同源的后量子密码算法仅 SIKE 一种算法,其研究工作目前甚少,且因为该密码方案的性能效率较低,目前的主要研究在于加速优化实现方面,因此本文没有编写关于基于超奇异同源的详细调研.针对超奇异同源密码系统的第一个错误攻击是在 2017 年由 Ti[9]提出来的,它的攻击方法是通过故障注入将基点变为随机点.在 Ti 工作的基础上,Gelin 和 Wesolowski 提出了一种环中止故障攻击,它利用了同源计算的迭代结构[10];Koziel 提出了 3 种改进的精致能量分析(refined power analysis,简称 RPA)攻击[11],它们是基于二次可拓域的零值表示和同源算法,攻击目标主要在于 3 点蒙哥马利梯度算法.

本文第 1 节简述后量子密码的研究背景、安全挑战.第 2 节~第 5 节分别分析基于格、基于编码、基于哈希、基于多变量的后量子密码算法的侧信道攻击方法与防御策略.第 6 节进行梳理和总括不同类型的后量子密码算法的特点、攻击手段,以及不同攻击手段的防护策略及其防护策略的代价.最后第 7 节总结全文.

1 背 景

本节主要包含两部分:第 1 部分讲述后量子密码算法的发展状况和研究现状;第 2 部分介绍后量子密码算

法目前主要面临的侧信道攻击手段.

1.1 后量子密码算法

为了防止量子计算机部署后对目前的密码系统造成不可逆转的伤害,目前各国都在积极地推动后量子密码算法 PQC 标准的制定,最有代表性的是美国 NIST 在 2015 年开启后量子密码算法标准征集项目,2016 年正式开始征集算法,如今已进入第 2 轮评选筛选;中国在后量子密码标准的征集方面也做了不少工作,2019 年中国密码管理局委托中国密码学会(CACR)举办了后量子密码算法竞赛的征集,这场竞赛是中国在后量子密码算法标准制定的预赛,意味着中国也开始了后量子密码标准的制定征程.下面分别介绍 NIST 和 CACR 的具体情况.

(1) 美国 NIST 的 PQC 征集[12]情况

NIST(National Institute of Standards and Technology)面向全球所有密码学者征集抗量子密码算法,以制定下一代公钥密码标准,在规模上,NIST 的 PQC 密码竞赛是目前全球影响力最大的密码标准征集竞赛.截止 2017 年 12 月 21 日,NIST 公布共接收 69 份算法,后来有 5 种算法退出评选,因此在一轮评估中实际有效数只有 64 份,表 1 为美国 NIST 后量子密码征集情况:主要包含了 NIST 一轮、二轮共接收算法数量、算法应用类型及困难类型(左边第 1 列)等;算法应用类型中 PKE、KEM、SIG 分别代表公钥加密、密钥封装和签名方案.根据 NIST 在 2020 年 7 月 22 日关于第 3 轮候选算法的最新公布,最终 PKE/KEM 候选算法 4 个,候选 PKE/KEM 算法 5 个,最终 SIG 算法 3 个,候选 SIG 算法 3 个.

Table 1 The NIST post-quantum cryptography standardization process

表 1 美国 NIST 后量子密码征集情况

NIST	PKE/KEM			SIG			总数		
	1 轮	2 轮	3 轮	1 轮	2 轮	3 轮	1 轮	2 轮	3 轮
格	21	9	5	5	3	2	26	12	7
编码	16	7	3	2	0	0	18	7	3
多变量	3	0	0	8	4	2	11	4	2
哈希	1	0	0	2	1	1	3	1	1
其他	4	1	1	3	1	1	7	2	2
总数	45	17	9	20	9	6	64	26	15

注:由于基于多变量的 DME 算法既实现了 KEM 又实现了 SIG,因此 NIST 一轮 KEM+SIG 的总数合计为 65

(2) 中国 CACR 的 PQC 征集[13]情况

表 2 给出中国 CACR 后量子密码的征集情况.

Table 2 The CACR post-quantum cryptography design competition

表 2 中国 CACR 后量子密码征集情况

CACR	PKE/KEM		SIG		AKE		总数	
	1 轮	2 轮	1 轮	2 轮	1 轮	2 轮	1 轮	2 轮
格	12	6	6	3	8	2	26	11
编码	4	1	0	0	0	0	4	1
超奇异	1	0	0	0	1	1	2	1
其他	2	0	2	1	1	0	5	1
总数	20	7	8	4	10	3	38	14

中国目前尚未向全球征集密码算法,由 CACR 举办的密码竞赛只面向国内的密码学者,下面是 CACR 在 2019 年举办的密码竞赛情况(只关注于公钥密码算法,不关注对称密码).CACR 共收到 38 份,表 2 为 CACR 一轮、二轮共接收的算法数量、算法应用类型及困难类型(左边第 1 列)等.算法应用类型中 PKE、KEM、SIG、AKE 分别代表公钥加密、密钥封装、签名和认证密钥协商方案.

由上文 NIST 和 CACR 的情况可知,后量子密码从数学难题上可分为基于格、基于编码、基于哈希、基于多变量和基于超奇异同源等几种主要的密码类型.

1.2 侧信道攻击

侧信道攻击从目标密码系统在平台运行中获取侧信息,这与其他形式的密码分析形成了对比,在其他形式的密码分析中,一般都是攻击算法及其底层的计算问题.所有的电子设备都会以多种方式泄露信息,侧信道攻击通过来自目标设备泄露的这些信息来查找与密钥相关的信息,这些可能是设备内部操作的时间或功率轨迹,或者是设备产生的错误输出.侧信道攻击可分为以下几类:时间攻击、能量分析攻击、故障注入攻击等.

(1) 时间攻击(timing attack,简称 TA)[14].密码设备的运行时间可以构成一个信息通道,为攻击者提供所涉及的密钥参数的宝贵信息.在时间攻击中,攻击者可以根据密码设备运算的时间推断出所执行的运算操作,而由这些操作即可推算出所涉及的密钥的信息.

(2) 能量分析攻击:密码设备的功耗可以提供有关发生的操作和相关参数的大量信息,通过这些功耗信息可以获取与功耗相关的操作和数据信息.

- 简单能量分析(simple power analysis,简称 SPA)[15].简单能量分析是侧信道能量分析攻击中最简单的一种攻击,它记录密码系统设备的功率轨迹(也称为能量迹,能量迹是指在加密操作时的一组功耗测量值),并对其进行检查,以识别可能用于破解密码系统和检索密钥的可见特征.
- 差分能量分析(differential power analysis,简称 DPA)[15].比较流行和强大的能量分析攻击是差分能量分析攻击.DPA 不需要对加密硬件进行任何形式的物理入侵,它可以由任何对其内部工作方式(即密码体制)有足够了解的攻击者实施,比如:仅了解一些密码系统的实现过程而不用掌握执行密码算法的密码芯片内部结构.DPA 攻击试图使用密码系统消耗的功率与输入数据之间的统计相关性来提取密钥信息.
- 相关能量分析(correlation power analysis,简称 CPA)[16].CPA 是基于电路的实际功耗和功耗模型之间的关系,如汉明重量模型与功耗呈线性关系,来进行分析,因为正确密钥对应操作的功耗与中间值的汉明重量之间的相关系数会达到最大.
- 模板攻击(template attack,简称 TA)[17].模板攻击是一种简单能量分析攻击的变体,从理论上讲,这是最强的侧信道攻击形式.这种攻击要求敌手拥有一个完全可以控制的相同设备,以建立各种操作指令的模板.

(3) 故障攻击(fault attack,简称 FA)[18].错误的计算有时是发现正确密钥的最简单的方法.故障攻击(有时也称为故障注入攻击)是一种更强大的密码分析技术,其原理是通过篡改设备,在加密操作期间注入一些影响加密操作的物理行为(如注入电磁、时钟频率、电压等),使其执行一些错误操作,利用错误行为的结果泄漏出涉及密钥的信息.

2 基于格的 PQC

2.1 基于格的密码系统

(1) 格及格的困难问题

格是一种代数结构,n 维满秩格 Λ 是 R^n 上的离散加法子群,具有性质 $spam(\Lambda)=R^n$,它由一组基 $B=\{b_0,b_1,...,b_n\}\in R^{n\times n}$ (称为格基)生成,n 维满秩格 Λ 可以表示为 $\Lambda=L(B)=\{Bx.x\in Z^n\}$.

- 最短向量问题(SVP)的定义:给定格基 $B=\{b_0,b_1,...,b_n\}$, 目的是在这组基构成的格中找到一个非零的最短向量,即找到一个向量 $v\in L(B)$, 使得 $\|v\|=\lambda_1(L(B))$. 基于 SVP 问题演变出近似版本 SVP_γ,对于 $\gamma>1$,给定格基 $B=\{b_0,b_1,...,b_n\}\in R^{n\times n}$, 找到一个非零向量 $v\in L(B)$, 使得 $\|v\|\leqslant\gamma\circ\cdot\lambda_1(L(B))$. 对于小的因子 γ,近似的 $\gamma>1$ 更困难,而对于增大的因子 γ,近似的 SVP_γ 更容易.
- 最近向量问题(CVP)的定义:给定格基 $B=\{b_0,b_1,...,b_n\}$ 和目标向量 t,目的是在这组基构成的格中找到一个最接近 t 的向量 $v\in L(B)$, 即找到一个向量 $v\in L(B)$, 使得 $\|v-t\|$ 最小.基于 CVP 问题演变出近似版本 CVP_γ,对于 $\gamma>1$,给定格基 $B=\{b_0,b_1,...,b_n\}\in R^{n\times n}$, 找到一个非零向量 $v\in L(B)$, 使得 $\|v-t\|\leqslant\gamma\circ\cdot dist(t,L(B))$, 其中,$dist(t,L(B))$为 t 到格 $L(B)$的距离.

(2) 带错误的学习问题

目前基于格的公钥密码算法和密钥交换协议绝大部分都是基于 2005 年由 Regev 提出的带错误的学习问题(learning with error,简称 LWE)[19]及其变形问题来构造的.它与其他古典的格困难问题(例如 SVP 和 CVP)相比,LWE 问题已被证明功能更加全面.

LWE 分布的定义为 n、q 是正整数上的元素,χ是在 $\mathbb{Z}$ 上的一个分布,给定 $s\in\mathbb{Z}_q^n$, 通过均匀随机选择 $a\in\mathbb{Z}_q^n$, 以及从分布χ中选取整数错误 $e\in\mathbb{Z}$,LWE 分布 $\mathcal{A}_{s,\chi}$ 输出 $(a,\langle a,s\rangle+e \bmod q)\in\mathbb{Z}_q^n\times\mathbb{Z}$.一般来说,LWE 问题分两种.

- 搜索型 LWE(search LWE,简称 sLWE)的定义:令 n、m、q 是正整数的元素,χ_s、χ_e 是$\mathbb{Z}$上的分布,给定 $(A,b=As+e)$, 求秘密向量 s.其中,矩阵 $A\in\mathbb{Z}_q^{m\times n}$, 秘密向量 $s\in\chi_s^n$, 错误向量 $e\in\chi_e^n$.
- 判定型 LWE(decisional LWE,简称 dLWE)的定义:令 n、m、q 为正整数,χ_s、χ_e 是$\mathbb{Z}$上的分布,判定下面两个分布: $D_0:(A,b)$ 和 $D_1:(A,u)$, 其中, $b=As+e, A\in\mathbb{Z}_q^{m\times n}, s\in\chi_s^n, e\in\chi_e^n, u\in\mathbb{Z}_q^m$.

这两种 LWE 在一定程度上是等价的,能求解 sLWE 问题的方法也能求解 dLWE.对于 LWE 问题中的错误分布 χ,一般都采用离散高斯分布方式.

(3) 格密码体制分类

所有格基公钥密码体制的算法可以分成 3 类:标准 LWE(learning with error,简称 LWE)、模 LWE(module learning with error,简称 MLWE)和环 LWE(ring learning with error,简称 RLWE).关于 LWE 问题,上面已有描述;RLWE 是在 2010 年由 Lyubashevsky 等人[20]提出的,是基于环上带错误的学习问题,其困难性等价于理想格的 SVP 困难问题的最糟糕情形;MLWE 是在 2015 年由 Langlois 等人[21]提出的基于模上带错误的学习问题,它是 LWE 和 RLWE 的推广.总体而言,MLWE 比 RLWE 具有更好的安全性能权衡.它们的区别可简述为:LWE 的分布样本分布在$\mathbb{Z}$上;而 RLWE 的样本分布在环上,常用的环为整数环 $R_q=\mathbb{Z}_q[x]/(x^n+1)$; 对于 MLWE 的样本分布在特殊环上,常用的有幂次环(是环 R_q 的特殊情形)和安全素数环.

2.2 针对格基PQC的侧信道攻击

(1) 能量分析攻击

能量分析一直是侧信道的主要攻击手段,近几年关于格密码的能量分析攻击有许多工作,如:2016 年,Pessl 针对文献[22]提出的随机洗牌策略进行安全分析,然后针对随机洗牌防护策略的高斯采样提出了一种新的 SPA 攻击[23];2017 年,Huang 等人[24]针对受掩码保护的格密码中的核心算子 NTT 实施了单能量迹攻击,是第一个针对格密码的单能量攻击,该攻击适用于大部分格密码算法;2018 年,Kim 和 Hong 针对 FrodoKEM 方案中的高斯抽样提出单能量迹攻击方案[25],这是首个针对高斯采样的单能量攻击;2019 年,文献[26]的作者 Pessl 和 Primas 等人也针对 KYBER 中的 NTT 提出了单能量迹攻击方案,并且还提出了 3 种提高攻击效率和成功率的方法;2020 年,Huang 和 Chen 等人[27]针对 NTRU Prime 算法中的多项式乘法提出了 4 种能量分析方法,主要针对不同版本的 NTRU Prime 算法的实现(如:参考实现、优化实现和保护实现);2020 年,Ravi 等人[28]针对 Round5、LAC、Kyber、Frodo、NewHope 等多种格密码算法进行了选择密文攻击,其主要攻击目标是纠错码和 FO 变换.

针对格基 PQC 的能量分析攻击大多基于密钥系数与功耗之间的关系.下面简单描述针对格基密码的侧信道攻击原理及过程.以文献[28]针对 Round5 算法的攻击方法为例:该方法的攻击目标是针对 XEf 纠错码,它们找到 XEf 纠错码的泄漏点:在 XEf 解码过程中的恢复消息步骤涉及到翻转比特位置的决策操作,在决策操作中包含了一个多位逻辑运算的计算操作,该操作也是解码操作决策是否纠错的最后一步,记该操作为 M.M 的输入是寄存器集 r 的修改形式,标记为 re',经过大量实验得出:若码字 c 合法,则 re'=0;若码字 c 带有一个或多个错误位,则 $re'\neq 0$.通过收集的电磁波可以很明显地检测出 c=0 和 c=1 两种情况下内部通用寄存器值的差异(计算时同一时刻的比较).利用上述泄漏点进行攻击的核心思路是:攻击者通过精心选择解密过程的密文,使其产生的码字可以唯一地识别某个目标密钥系数的值.攻击的大概过程是:由这些选定的密文触发解封装操作,随后使用 Welch's t-test 聚类技术对电磁波进行分析,揭示了密钥 s 中某个系数的值.重复这一过程以实现全密钥的恢复.

格基 PQC 的能量分析攻击成功率大部分都超过 90%,且分析时间也较短.如在上述文献[24]中,无论针对无掩码的 NTT,还是带掩码防护的 NTT,其攻击结果都显示:当噪声小于 0.4 时,两者的成功率基本上都能维持在 80%~95%之间.文献[26]提出了一种利用 Belief Propagation 方法改善的单能量攻击,针对恒定时间实现但无掩码的 NTT 实现,改良后的攻击方案可以在噪声的汉明重量为 1.5 的条件下,其成功率维持在 90%以上(对于基础版的攻击方案,当最高噪声的汉明重量大于 0.9 时,其成功率骤降);而针对带掩码实现的 NTT,单独使用 belief propagation 方法的攻击成功率依然很低(当噪声汉明重量为 0.9 以上时,成功率小于 40%),不过,针对掩码实现的 NTT 应用了一种其他方法,使得在汉明重量小于 0.3 时,成功率依然接近 1,针对采用 Lazy Reduction 实现的 NTT,它们的攻击在噪声汉明重量小于 1.3 时,成功率保持在 90%以上;文献[28]针对 Round5 的攻击大概需要 978 条能量迹,成功率为 99%,一次完整的攻击迭代需要 95s(包括 10s 的预处理时间),3 轮总共 270s,即 4.5 分钟完成密钥恢复;针对 LAC128 的选择密文攻击,需要 2×512=1024 条能量迹,平均成功率为 98%,一次迭代用时约 525s (8.75 分钟),完整密钥恢复用时为 1 490s(3 次重复用时≈25 分钟);针对 Kebey512 的 FO 变换的攻击大概需要 5×256×2=2560 个能量迹(n=256,k=2)就能检索出完整密钥,在约 3 次攻击的重复操作情况下,恢复出完整密钥的平均成功率约为 99%,一次完整的重复攻击(包括能量迹获取)大约需要 230s,因此,密钥的完全恢复可以在 650s 内完成(在 10.83 分钟内进行 3 次迭代).

(2) 故障攻击

故障攻击是一种强有力的攻击手段,也一直备受关注,针对格密码的故障攻击一直处于热潮.如 2018 年 Bruinderink 和 Pessl 等人[29]针对 qTESLA、Dilithium 两种密码算法提出了差分故障攻击;2019 年,Ravi 和 Roy 等人[30]针对 NewHope、KYBER、Frodo 等多个不同密码方案中高斯分布/中心二项式的 nonce 随机种子提出了故障攻击方案;McCarthy 和 Howe 等人[31]针对 FALCON 密码算法提出了故障攻击方案;Valencia 等人在文献[32]中提出针对后量子密码的故障敏感度分析,攻击点在于乘法器、加法器,其原理是利用了设备处理的数据与设备对故障的敏感性之间的相关性.

针对格密码的故障攻击一般在密码算法运行时注入故障诱导随机种子 nonce 复用,从而促使算法计算出错误结果.下面简述一下其中的故障攻击原理[30],以 NewHope KEM 方案中密钥生成过程为例,构成公钥的主要元素 b 是由密钥 s 和错误 e 经过多项式乘法得到(在 NTT 域上的运算),生成 s 和 e 采样的唯一区别是 nonce 的不同(即:noiseseed 都一样,s 的 nonce 为 0,e 的 nonce 为 1).假设加密方案中 Ring-LWE 的实例为 $b=a\times s+e\in R_q$, 上面的方程是环上的一个由 n 个方程($2n$ 个未知数)所构成的线性方程组(每个多项式都有 n 个系数,s、e 为未知量,因此有 $2n$ 个未知量).因此,当攻击者注入故障(利用电磁故障注入让 nonce 跳过更新,保持原来的值)使得 s 和 e 都使用同一个种子,也就是 nonce 一样,那么上面方程就变成 $b=a\times s+s\in R_q$, 这个有缺陷的 LWE 实例是一个由 n 个方程(n 个未知数)所构成的线性方程组(每个多项式都有 n 个系数),因此这可以使用高斯消元法很简单地求解出私钥 s.

故障攻击的攻击效果主要取决于注入的故障数和注入成功率.如在文献[29]中,在 Keccak-f 最后一次(下面记为 1P)和倒数第 2 次(下面记为 2P)调用的平均注射故障数为 39 和 93;而在矩阵 A 的生成、随机种子采样(1P/2P)、多项式乘法和哈希等不同地方注入故障的成功率分别为 54.4%、24.8%/23.9%、25%~90%、91%.文献[30]实现了精确地注入故障,即百分百成功,因此其复杂度可以转换为注入故障数,在 FRODO 和 NEWHOPE 两种方案中,恢复密钥和消息分别只需注入 1 个故障;而对于 MLWE 类型的密码方案(如 Kyber 和 Dilithium)所需要的故障的数量为矩阵 A 的维度.

(3) 其他攻击

能量分析攻击和故障攻击是针对格的侧信道攻击的两种主要攻击手段,但也存在其他类型的侧信道攻击手段,如时间攻击、冷启动攻击等.2018 年,Albrecht 和 Deo 等人[33]针对 Kyber 和 Newhope 中的 NTT 进行了冷启动攻击(cold boot attacks);2019 年,D'Anvers 和 Tiepelt 等人[34]针对 LAC、Ramstake 两种算法的纠错码提出了时间攻击方案;Espitau 和 Fouque 等人[35]针对 BLISS 方案中的拒绝采样、高斯采样、多项式乘法等多个核心算子进行攻击.

针对格密码的时间攻击的效率较高,而其他类型攻击的效果与攻击方法和攻击点相关.如文献[34]针对LAC 中的纠错码进行时间攻击需要大约 2 400 个能量迹,在 2 分钟内可恢复全部密钥(在 Intel(R) Core(TM) i5-6500 CPU,3.20GHz 平台下).文献[35]的攻击效果虽然针对拒绝采样的攻击效率较低,n=256 和 n=512 的BLISS 算法中拒绝采样的攻击时间分别为 17 个小时(2^{47} 个时钟周期)和 38 天(2^{53} 个时钟周期),但对于高斯采样,恢复出完整密钥的成功率为 95%,针对多项式乘法,当噪声标准差在 3.0 及以下时,单解的平均时间在 10ms 以下.

综上所述,格密码中涉及密钥相关的核心操作主要有多项式乘法、NTT(用于加速多项式乘法操作的一种方法)、高斯分布采样、中心二项式采样、纠错码以及易受故障攻击的随机种子产生过程,在这些核心操作计算过程中,均会涉及到私钥或消息等关键信息.在不同格基密码方案中,多项式乘法操作和 NTT 操作不一定是共存的,如 LAC 只有稀疏多项式乘法;在基于格的 qTESLA 签名方案中,稀疏多项式乘法和 NTT 都存在.而高斯分布采样和中心二项式采样用于密钥多项式、噪声多项式的产生,高斯分布采样相对于中心二项式采样精度更高,但较耗时,因此,除格基签名方案以外,其他密码方案基本都采用了中心二项式采样来替代高斯分布;纠错码用于降低格基密码的解密错误失败率.图 1 所示为格基后量子密码的近 3 年攻击分析图,从图中可以清楚地看到,针对后量子密码的主要攻击手段为故障攻击和能量分析攻击,从攻击点来说,主要分布在格密码的各个核心算子,包括 NTT、多项式乘法、纠错码、随机种子等.

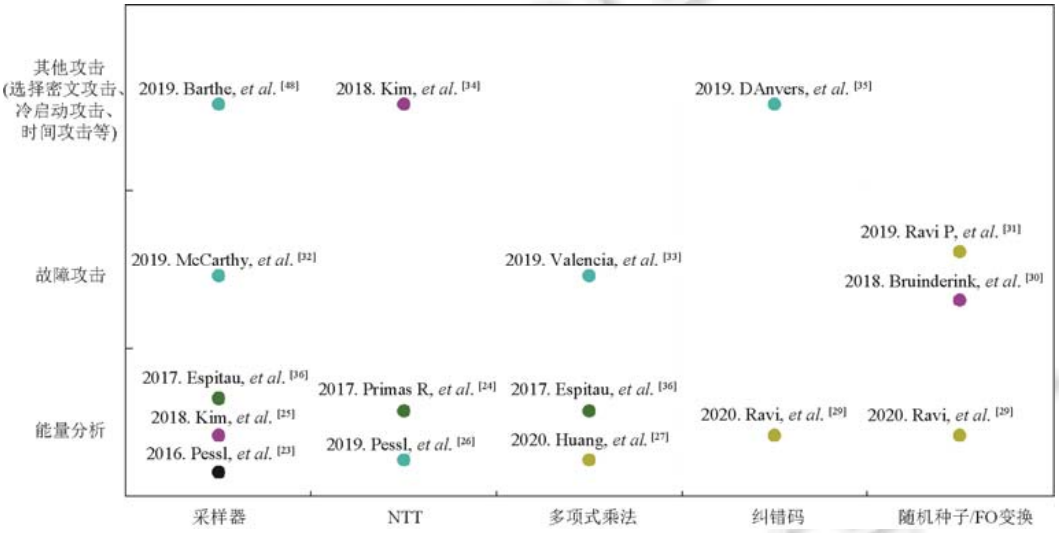

Fig.1 Side-channel attack analysis of lattice-based PQC

图 1 格基 PQC 的侧信道攻击分析图

2.3 侧信道防御策略

隐藏、掩码和检测技术是预防侧信道攻击的有效防御策略,下面简述近几年的防御策略研究工作,分别对能量分析、故障攻击和时间攻击的防御工作进行介绍.

(1) 能量分析的防御工作

针对能量分析的防护手段主要是随机化和掩码.如在 2014 年,Roy 等人[36]针对高斯采样提出了随机洗牌的防御策略;在 2018 年,Oder 等人[37]提出一种掩码实现的二项式采样器,该采样器可用于随机错误变量 e 的生成;在 2019 年,Pessl 和 Primas 等人[26]针对 NTT 进行能量分析攻击,然后推荐使用随机洗牌的方法作为防护对策;而在掩码防御策略上,2015 年,Reparaz 等人[38]提出布尔掩码的 RLWE 解密实现;在 2016 年,他们又使用加性掩蔽来确保 RLWE 实现[39]的安全性,同时指出该掩蔽方案可通用于其他加法同态的加密方案;又如 2018 年,Barthe 等人[40]利用算术掩码与布尔掩码的转换技术提出了一种可证明的任意阶安全采样器,并且实现了带掩码的 GLP 方案.

不同防护手段的防护代价和防护效果差距较大,掩码的安全性更高,但代价高于其他防护策略.在文献[37]中,通过加隐藏策略来保护错误多项式的采样,根据实验结果分析,CCA-2 安全的解密过程一共增加了 226 476 个时钟周期(无掩码版本),其中没有使用隐藏策略时需要消耗 4 416 918 个时钟周期数;而在掩码实现的采样器中,增加隐藏策略的保护采样器在解密过程中一共增加了 305 887 个时钟周期,其中没有使用隐藏策略时所耗时钟周期数为 25 334 493(隐藏技术增加了 1.21%的代价).文献[38]采用并行掩码译码器只需要 $1/2\times n\times N$ 个周期,如在 $n=256, N=16$ 的情况下,掩码解码器需要 2 048 个周期,而整个掩码解密操作总共需要 7 478 个周期,无保护的解密操作大概需要 2 800 个周期,掩码加密的时钟周期同比增长率为 167%.2016 年,Reparaz 等人[39]与他们在 2015 年的相关工作[38]进行了对比,在对前两步的等式可以采用离线计算方式以及在实现的便捷性(如:采用掩码表实现)上得到了改进.文献[40]所提出的掩码方案与无防护的实现方案相比,在 d-probing 模型中,当 d=2,3,4,5,6 时,耗时分别为 8.15、16.4、393.5、62.1、111(s),分别同比增长 15、30、73、115、206 倍(其中,无保护实现的耗时为 0.540s).

(2) 故障攻击的防御工作

对于故障攻击的防护手段主要是随机化和增加故障监测机制.如在 2016 年,Bindel 等人[2]讨论了各种故障攻击方案及对策,然后提出了应对随机化故障的对策.包括增加输入密钥的大小和计算多项式的逆;Espitau 等人[41]提出了几种可能的故障攻击,并讨论了减轻这些攻击的可能对策;2018 年,Bruinderink 等人[29]提出了 3 种针对故障攻击的通用对策:a) 签名的重新计算;b) 签名后再进行验证签名;c) 在签名时增加随机性;2019 年,Howe 等人[42]针对故障攻击也提出了 3 种不同的对策:a) 低代价对策:计算寄存器相同值的重复次数;b) 检验输出分布的均值与方差;c) 进行卡方检验.

不同防护机制的有效防护点和所需代价有一定的差别.如在文献[5]提出的 3 种方案中,签名的重新计算方案无法保护生成矩阵的种子产生过程;而签名后进行验证签名方案不能保护随机种子采样;但在签名时增加随机性这种方案可以做到他文中所有攻击点的保护.文献[7]提出的 3 种防御策略分别对应低代价、标准、高代价这 3 个版本,低代价的防御策略需要 36 个 Slices(额外增加了 8%的代价,在无保护的高斯采样中需要 33 个 Slices),标准防御策略需要 55 个 Slices(额外增加了 44%),而高代价的防御策略虽然安全性上较高但非常影响性能,该策略额外增加了 126 个 Slices(额外增加了 3.8 倍).

(3) 时间攻击的防御工作

针对时间攻击的防御对策主要是算法的恒定时间实现,从而减少运行时间维度泄露的信息.针对时间攻击常利用的攻击点,主要有以下研究工作:2016 年,Khalid 和 Howe 等人[43]提出了基于 FPGA 的适用于大范围离散高斯采样器的恒定时间硬件实现;2017 年,Micciancio 和 Michael 也提出了高斯采样的恒定时间实现方案[44],他们的思想是将标准差较小的样本与标准差较大的样本相结合;2018 年,Karmakar 等人[45]也提出了新的高斯采样恒定时间实现方案,它们将采样器的输出值表示为输入位的布尔函数,从而实现了完全恒定时间的采样算法;但是高斯采样在性能上表现不佳,这一缺点一直是高斯采样的短板,为了弥补这一缺陷,2019 年,Karmakar 等人[46]利用优化了的位切片,实现了加速采样;除了高斯采样外,在其他核心算子的防御策略上也有部分研究工作,如 2019 年,Barthe 等人[47]针对 BLISS 签名算法提出了抗时间攻击的防御方案;2019 年,Walters 和 Roy 实现了恒定时间的 BCH 纠错码[48],并应用于 LAC.

针对格密码核心算子的恒定时间实现的代价都小于 0.5 倍这一特性,有些恒定实现的性能还比非恒定时间实现的版本要高.文献[47]利用 SUPERCOP 测试工具得出:无论在低四分位数、中位数和高四分位数的开销均在 220 个时钟周期左右(基本与非恒定时间实现的原 BLISS 算法的性能一致),而对于 Dilithium 算法中恒定时间实现的高斯采样参考实现版本的性能最高比他们的实现慢 5.8 倍(需要 1 526 个时钟周期),Dilithium 算法的恒定时间高斯采样的 AVX2 实现除低四分位采样会比他们的实现快之外,中位数和高四分位数性能比他们的实现都要慢 0.5 倍~1 倍.文献[45]提出来的恒定时间高斯采样算法的效率远高于恒定时间的 CDT 高斯采样算法,如:在不包含伪随机数发生器时,要产生 64 个高斯样本,CDT 算法大概需要 28 231 个时钟周期(精度参数 λ=128),而他们的高斯采样算法只需要 11 814 个时钟周期,同比减少了 58%,并且,他们还实现了 SIMD 版本的采

样器,产生 256 个样本只需要 19 605 个周期(精度参数λ=128).文献[46]利用位切片来提升高斯采样的效率,与最快的非恒定时间实现的 CDT 算法相比,他们的算法在最差情况下性能下降 33%(最快非恒定时间的字节扫描 CDT 算法每秒可以签名 10 327 次,他们的算法在同安全级别下每秒可签名 7 025 次);与普通的非恒定时间 CDT 采样算法相比,最差情况下,性能下降 13%;但与恒定时间的线性搜索 CDT 算法相比,他们的性能可以提升 15%以上;与上述工作[45]相比,他们在标准差为 6.155 43 时,性能提升了 11%;但在标准差为 2 时,该算法的性能最高可以提升 37%.

根据上述不同侧信道攻击手段的防御策略和攻击点,梳理出一幅分析图,如图 2 所示,从近几年格密码的防护分析图中可以清楚地看到,单独针对 NTT 和纠错码的防护工作较少;针对高斯采样的防护工作最为丰富,从时间攻击到故障攻击,再到能量分析攻击的防护策略均有较多的研究工作;同时也有许多针对整个密码方案(如加解密或者签名)以实现防护策略的研究.

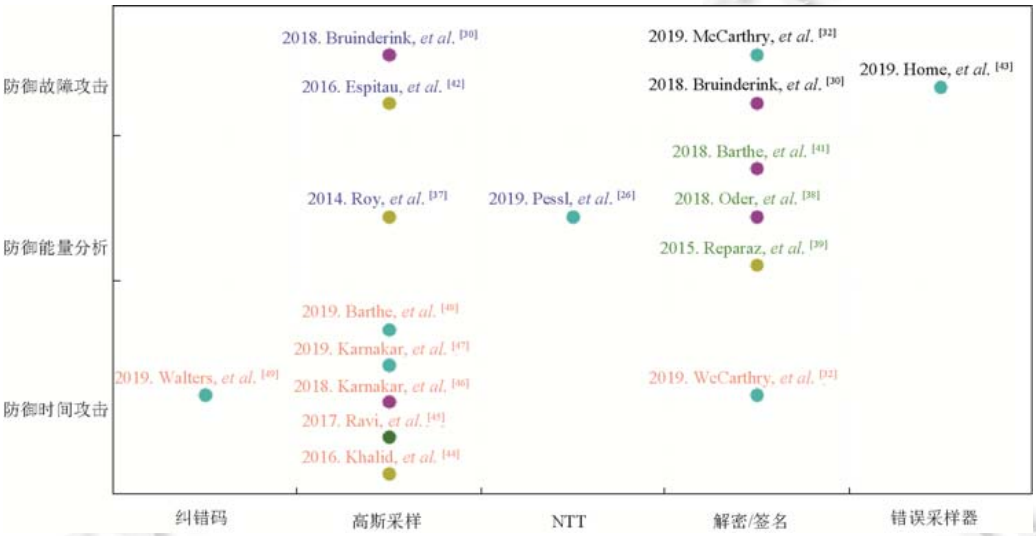

Fig.2 Side-channel defense analysis of lattice-based PQC

图 2 格基 PQC 的侧信道防御分析图

3 基于编码的 PQC

3.1 基于编码的密码系统

McEliece 于 1978 年推出了第一个基于编码的密码系统[49].该系统不是基于数论原语,而是来自编码理论的难题.它的安全性依赖于两个问题:(a) 校验子解码问题的难度;(b) 区分二进制 Goppa 码和随机线性码的难度.与其他 PKC 相比,McEliece 的方案具有这样一个优势:加密和解密的复杂度和对称密码一样,具有非常高效的性能[50].此外,解决校验解码问题的最佳攻击是码长指数型,所以 McEliece 方案具有较高的潜在优势.

(1) 编码的基本原理

本节主要简述如何利用编码理论为 PKC 提供有效的解决方案,即简述编码的基本原理.若需要了解更多编码理论,可参考文献[51].

(a) 循环矩阵(circulant matrix):令向量 $v=(v_0,v_1,...,v_{n-1})\in F_2^n$,则向量 v 生成的循环矩阵为 $rot(v)=\begin{bmatrix} v_0 & v_{n-1} & \cdots & v_1 \\ v_1 & v_0 & \cdots & v_2 \\ \vdots & \vdots & \ddots & \vdots \\ v_{n-1} & v_{n-2} & \vdots & v_0 \end{bmatrix}\in F_2^{n\times n}$,所以两个向量 $u,v\in R$ 的乘积 $rot(\cdot)$ 可以表示为 $u\cdot v=u\cdot rot(v)^T=(rot(u)\times v^T)^T=v\times$

$rot(u)^T = v \cdot u$.

(b) 线性编码(linear code):[n,k]-线性码 C 表示一个长度为 n、维数为 k 的线性码,则[n,k]-线性码 C 是 k 维 R 上的一个子空间,我们将 C 的元素称为码字.线性码的目的是检测和纠正错误.

(c) 对偶码(dual code):码字 C 的对偶 $C^{\perp}$ 是指 $F_{q^m}^n$ 的向量子空间的正交补空间.

(d) 矩阵生成器(generator matrix):如果 $G \in F_2^{k \times n}$ 满足码字 $c = \{mG, m \in F_2^k\}$, 则称 G 是[n,k]-线性码 C 的矩阵生成器.

(e) 奇偶校验矩阵(parity-check matrix):给定一个[n,k]-线性码 n,若 $H \in F_2^{(n-k) \times n}$ 满足 $C = \{v \in F_2^n, Hv^T = 0\}$ 或 $C^{\perp} = \{uH, u \in F_2^{n-k}\}$. 因此,$C$ 的一个校验矩阵 H 就是 $C^{\perp}$ 的一个 $(n-k) \times n$ 的生成矩阵.

(f) 校验子(syndrome):令 $H \in F_2^{(n-k) \times n}$ 是[n,k]-线性码的奇偶校验矩阵, $v \in F_2^n$ 是一个字(word),那么 v 的校验子为 Hv^T,并且 $v \in C iff Hv^T = 0$.

(g) 最短距离(minimum distance):令 C 是 R 上的[n,k]-线性码,ω是 R 的范数.那么 C 的最短距离为 d= $\min_{u,v \in C, u \neq v} \omega(u-v)$ (目前基于编码的后量子密码算法使用的距离为汉明距离或秩距离).

(2) 困难问题

基于编码的密码系统的困难问题大多都基于解码问题的变体,它包括寻找与给定向量最接近的码字.而当处理线性码时,接收向量的校验子与接收向量这两者的解码问题一样,因此下面只简述校验子解码(syndrome decoding,简称 SD).

(a) 校验子分布(SD distribution):对于正整数 n、k、w 的 $SD(n,k,w)$分布为:随机挑选 $H \xleftarrow{\$} F_2^{(n-k) \times n}$ 和满足 $\omega(x) = w$ 的 $x \xleftarrow{\$} F_2^n$,然后输出 $(H, \sigma(x) = Hx^T)$.

(b) 校验子解码问题(syndrome decoding problem,简称 SDP):已知矩阵 $H \in M_{n-k,n}(F_{q^m})$ 和一个整数 w>0,求一个距离 $\omega(x)$小于 w 的向量 $x \in F_{q^m}^n$, 使得 $Hx^T = s$. 注: $M_{n-k,n}(F_{q^m})$ 表示所有在 F_{q^m} 上的 $(n-k) \times n$ 矩阵.

下面是校验子解码问题的两个变种:搜索型 SDP 和决策性 SDP.

(c) 搜索型 SDP(search SD problem):给定 $(H, y^T) \in F_2^{(n-k) \times n} \times F_2^{n-k}$, 求向量 $x \in F_{q^m}^n$ 使得 $Hx^T = y^T$ 且 $\omega(x) = w$.

(d) 决策型 SDP(decision SD problem):给定 $(H, y^T) \in F_2^{(n-k) \times n} \times F_2^{n-k}$, 判断 (H, y^T) 是否满足 $SD(n,k,w)$ 分布,或者均匀分布于 $F_2^{(n-k) \times n} \times F_2^{n-k}$.

3.2 针对编码PQC的侧信道攻击

(1) 能量分析攻击

本节简述近几年能量分析对编码 PQC 的影响.Von 等人[52]在 2014 年提出了基于 QC-MDPC McEliece 密码系统的简单能量攻击,实现了消息恢复攻击和私钥恢复攻击.他们根据在密钥生成时是否执行条件转移指令,可以观察到不同的功耗模式,依此得出密钥的相关信息.然后,他们还为此提出了一个使用 ARM Thumb-2 汇编语言的恒定时间实现的防御对策,更具体地说,他们采用了掩码方案,该掩码值要么是 0,要么所有的位都是 1,并采用逻辑 AND 指令来选择要使用的数据,最终实现抵御攻击.Chen 等人[53]在 2015 年提出了针对 QC-MDPC McEliece 密码系统的水平差分能量攻击,它在解密时通过选择密文进行 DPA 攻击,从而实现密钥恢复.同时还提出了一个基于布尔掩码的阈值实现的防御对策[54].在 2016 年,Chou 提出了一种基于 QC-MDPC 编码的恒定时间实现[55],以抵御定时攻击.随后,在 2017 年,Rossi 等人[56]在 CHES 2017 上表示这种对策容易受到差分功率分析(DPA)的攻击,但是他们所提出的 DPA 仍然不能完全恢复密钥,需要进一步求解线性方程才能获得完整的密钥信息.因此,在 2019 年,Sim 等人[57]提出了一种能够完全恢复密钥的多能量迹攻击,与 Rossi 等人的攻击相比,该攻击使用多个能量迹来恢复整个密钥,解决了需要额外求解线性方程的问题.

针对基于编码的后量子密码算法的能量分析可通过校验子 H 的相关计算作为攻击点.如文献[56]的攻击思

路.Rossi 等人提出的 DPA 攻击主要攻击点为比特翻转函数.它的泄露点在于 $S=H_{priv}\cdot v$, 其中, $H_{priv}=H_0\,|\,H_1$ 是私钥.由解密算法可知, $v=(c\,|\,0)$, 因此,该结果可以表示为 $S=H_{priv}\cdot\begin{pmatrix}c^T\\0\end{pmatrix}=(H_0,H_1)\cdot\begin{pmatrix}c^T\\0\end{pmatrix}=H_0\cdot c^T$, 因为该公式中的乘法可以通过计算旋转(即 XOR 操作)后的密文 $r_{x_i}(c)$ 来实现,并且在循环中,每个旋转后的向量 $r_{x_i}(c)$ 在计算时存储到临时内存位置,本轮 XOR 结果为 $S_i=S_{i-1}\oplus r_{x_i}(c)$, 即前一次迭代的 XOR 结果作为本次迭代 XOR 操作的一部分.因此通过侧通道分析模型假设设备的功耗取决于存储在内存中的每个旋转向量 $r_{x_i}(c)$ 的最左位(位位置 0)是 0 还是 1.注:c 的第 x_i 位被 r_{x_i} 旋转到第 0 位,并被 $r_{x_{i-1}}$ 旋转到第 1 位,然后先利用中间值 S_i 把能量迹 T 分成两个集合(对应于 0 和 1),然后计算这两个集合的均值并得到差异曲线,若在不同的地方有大的尖峰,表明有信息泄露,即穷搜 64 种可能的 x_i.到此只能恢复出 H_0,完整私钥为 H=$H_0|H_1$.下面简述如何利用 H_0 恢复出完整密钥,因为公钥 $P=(H_1^{-1}\cdot H_0)^T$, 设 $Q=P^{-1}$, 那么 Q 利用 H 可表示为 $Q\cdot H_0^T=H_1^T$, 矩阵 H_0 和 H_1 可以用矩阵的第 1 行 h_0 和 h_1 分别表示,因此可表示为 $Q\cdot h_0^T=h_1^T$, 其中,Q 可计算得到,且 h_0 已知,因此利用线性方程可以解出 h_1.

较短时间和少量的能量迹即可满足针对基于编码的能量分析.文献[56]提出的 DPA 攻击,可以实现 100%的攻击成功率,并且在性能上恢复完整密钥最多不超过 10 000 次迭代.在时间上,当搜索空间长度为 8 时,128bit 安全大概需要 2s;当搜索空间长度为 16 时,攻击 128bit 安全大概需要 4 分钟,迭代长度意味着 DPA 分析的精度.文献[57]提出的多能量迹攻击结果表明,该攻击方法在 32bit 处理器上大约只需 50 条能量迹即可满足攻击 80bit 安全的校验子计算.

(2) 时间攻击及其他攻击

2008 年,Strenzke 等人[58]针对使用 Goppa 编码的 McEliece 解密过程提出了时间攻击,在 2018 年,Eaton 和 Lequesne 等人在文献[59]中提出针对后量子密码算法 QC-MDPC 方案进行时间攻击,引入了稀疏向量的距离谱的概念,并证明了距离谱足以恢复出向量,他们的思路是通过观察许多错误的明文,恢复出 QC-MDPC 密钥的距离谱;Paiva 和 Terada 在文献[60]中针对非恒定时间实现的 HQC 方案也提出了时间攻击方法,攻击的原理是非恒定时间实现的 HQC 的解密时间取决于 BCH 纠错码中的错误数量;根据非恒定时间实现的 BCH 纠错码来实现时间攻击的还有 2019 年由 Wafo-Tapa 等人[61]针对 HQC 提出的选择密文时间攻击方案,该方案利用了解码错误的权重与 BCH 码解码算法的运行时间之间的相关性;2020 年 3 月,Danner 和 Kreuzer 也提出了针对使用二进制不可约的 Goppa 码的故障攻击[62],该攻击可在 25 分钟内攻破最高级别的安全参数.

时间攻击最主要的原理是非恒定时间实现的算法会因操作不同的数据而泄露出关键信息.下面举例说明针对非恒定时间的 HQC 攻击方法[60].整个攻击方案分为两步:频谱恢复和密钥重构.频谱恢复是使用时间信息的部分.为了方便描述其过程,这里假设 Alice 是持有密钥的目标设备.下面是频谱恢复的大概过程.

1) 攻击者发送许多合法密文给 Alice,然后记录 Alice 对每条密文解密的时间信息.

2) 因为所有密文都是攻击者产生的,因此攻击者知道 r_1 和 r_2.攻击者利用频谱恢复算法迭代地构建两个数组 T_x 和 T_y,使得 $T_x[d](T_y[d])$是 d 在 $r_2(r_1)$频谱内的解密时间的平均值.

3) 获得最短距离 k:$T_y[k]$为数组中最小值.

4) 利用 Paiva 等人提出的 BuildD 算法,从 T_y 中分离出不在 x 频谱内的距离集合 D.

密钥重构是利用上面获得的信息来计算私钥中的 y:由上述步骤得到集合 D 和在频谱内的距离 k,然后再利用密钥重构算法恢复出 y,对于这一算法此处不再详述,因为该算法是 Guo 等人[28]提出的密钥重构算法的一个简单扩展版.

针对编码的后量子密码算法的时间攻击所需解密次数较多,而故障攻击的注入成功率远低于一些针对格密码的故障攻击的成功率.文献[60]指出,针对非恒定时间 BCH 纠错码的时间攻击大约需要 4 亿次解密才可能实现密钥重建,若解密次数在 6 亿次以上,则几乎光谱之外的所有距离都能被正确识别.文献[61]共发动了 1 000 次攻击,结果显示,大概有 88%的成功率,在复杂度上,针对 128bit、192bit、256bit 这 3 个不同的安全级别 HQC 的攻击,所需要的解码数分别为 5 441、5 852、6 631.文献[62]针对 Goppa 码的故障攻击,注入故障的成功率在

50%以下,并且,随着算法安全级别的提高,他们的注入故障成功率也会明显下降;在攻击所需消耗时间方面,对于 128bit 安全级别的攻击大概需要 170s;针对 192bit 安全级别的攻击大概需要 566s,对于最高安全级别的攻击也只需 1 451s(约 25 分钟).

3.3 侧信道防御策略

近几年,由于编码 PQC 的侧信道防护工作相对格密码较少,因此这里不再细分攻击类型的防护策略.在 2008 年发表的文献[58]中,从密文排列角度出发,提出了一种对高速缓存的时间攻击,之后,Strenzke 等人又提出了一种基于地址掩蔽的对策,但随后在 2016 年,该对策遭到 Petrvalsky 等人[63]提出的 DPA 的攻击.而第一个基于准循环低密度校验码(QC-MDPC)的能量分析是在 2015 年,由 Chen 等人[53]提出,该攻击针对的是在硬件实现上为 QC-MDPC 码校验子计算(利用了选择单密文攻击).之后,Chen 等人为此提出了一个对策[54],使用与校验矩阵大小相同的布尔掩蔽;同年,他们继续针对原来的对策方案进行了扩展[64],主要包括在密钥和校验子上应用掩蔽的对策,Chen 等人通过在校验子的计算和解码部分使用阈值来避免一阶侧通道攻击.2016 年,Chou 等人提出了一种基于准循环低密度校验(QC-MDPC)的基于密码的固定时间实现[55],以抵御时间攻击,但于 2017 年,该方案受到了 Rossi 等人提出的差分功率分析的攻击[56],该攻击方案利用了校验子的计算不是以向量矩阵积的形式进行而是以排他或旋转之间的形式来描述校验矩阵这一特征,因此相应的对策是在进行校验子计算之前用随机码字对密文进行掩码,这是之前在 2016 年由 Petrvalsky 等人提出的保护对策[63],该方法利用线性码的特性,不需要大量的随机比特,有利于低成本的嵌入式设备.2019 年,Wafo-Tapa 等人[61]针对采用非恒定时间实现的 BCH 纠错码的 HQC 提出的选择密文时间攻击方案,提出了恒定时间实现的 BCH 纠错码;2020 年,Danner 和 Kreuzer[62]针对使用二进制不可约的 Goppa 码提出了故障攻击,之后又提出了两种防御策略,其一是通过检查解码后输出值的权重来发现故障注入;其二是给出检测故障注入的另一种方法——重新加密输出.

针对基于编码的掩码实现效率较低,所需额外开销较大这一问题,文献[54]对校验矩阵实现了掩码防护,该掩码防护对策与无保护方案相比,总体上增加了 8 倍的资源;在 Flip-Flops(FFs)、Look-Up Tables(LUTs)、Slices 等方面分别为 3 045、4 672、1 549,而无保护方案分别为 412、568、148,同比增长了 7.4 倍、8.2 倍、10.5 倍.

4 基于哈希的 PQC

4.1 基于哈希的密码系统

基于哈希(散列)的密码方案是后量子密码学中重要的一类,它以仅使用密码哈希函数创建数字签名而闻名.进入 NIST 二轮的哈希密码方案只有 SPHINCS+签名方案.这种方案的主要优点是其安全性仅依赖于泛型哈希函数的某些加密属性.因此,如果所选的哈希函数在将来被攻破,则可用新的哈希函数替换被攻破的哈希函数.下面对如何基于哈希构造数字签名作一简单介绍.

最早利用哈希函数构建的数字签名算法是 Lamport 在 1979 年提出来的[65],但是,该算法容易被使用两个已正确的签名伪造出另一个伪签名,且该算法的签名和密钥都太大,因此,受 Lamport 启发,Merkle 根据 Winternitz 的一个猜想提出了一个一次性签名 WOTS(Winternitz-one-time signature)[66],它由 3 个值参数化.

* ω:WOTS 使用的字的大小;
* ℓ_1:待签名消息的字(大小为 ω)的字数量;
* ℓ_2:签名算法中校验值(大小为 ω)的字数量.

给定一个安全级别参数 n、哈希函数 $F:\{0,1\}^n \to \{0,1\}^n$ 和一个随机位掩码集合 $r=\{r_1,r_2,...,r_{\omega-1}\} \in \{0,1\}^{(\omega-1)\times n}$,则链接函数(chaining function)的 $c^i(x,r)$定义为

$$\begin{cases} c^0(x,r)=x \\ c^0(x,r)=F(c^{(i-1)}(x,r)\oplus r_i),\ i<\omega \end{cases}.$$

签名后的长度为 $\ell=\ell_1+\ell_2$, ω和 W 的关系是ω=2^W.其中,

$$\ell_1=\frac{n}{W},\ell_2=\frac{\log(\ell_1(\omega-1))}{W}+1,\ell=\ell_1+\ell_2.$$

在 Lamport 提出签名方案 40 年后,出现了第一个后量子签名方案——XMSS 方案[67],但是由于 XMSS 签名方案是一种有状态的签名方案,也恰恰是因为这一缺陷,使得它不满足 NIST 对签名方案的标准定义;而后,Daniel 等人提出了无状态的基于哈希的签名方案——SPHINCS,SPHINCS 方案里用了许多 XMSS 的组件,但为了消除状态,使用了较大的密钥和签名;进入 NIST 二轮候选的 SPHINCS+签名方案与 SPHINCS 类似,是 Daniel 等人基于 SPHINCS 进行了修改和改善的方案,提高了安全性和鲁棒性,但都使用上述的 WOTS 基础概念,而 WOTS 与密钥长度和安全级别息息相关,如 SPHINCS-256 的参数为:安全级别 n=256,字大小 ω=2^4=16bit;消息字数量 $\ell_1=64$, 校验码字数量 $\ell_2=3$; SPHINCS+签名方案对上述 WOTS 参数进行轻微的修改和定制,命名为 WOTS+.在 WOTS+中,主要定义了两个参数 n 和 ω.在 WOTS+中, $\omega=\{4,16,256\}$, 是安全参数,影响着消息、私钥、公钥的长度,签名后的长度 $\ell=\ell_1+\ell_2$ 也有所变化,长度如下: $\ell_1=\frac{8n}{W},\ell_2=\frac{\log(\ell_1(\omega-1))}{W}+1,\omega=2^W$. 链接函数的作用是利用 WOTS+和公钥来计算每轮迭代,即计算哈希链.

4.2 针对哈希PQC的侧信道攻击

近年来,基于哈希的后量子密码算法的侧信道攻击研究工作较少,因此这里不再细分攻击类型.2017 年,Genêt 在其硕士论文[68]中提出了针对 SPHINCS 算法的简单能量攻击、差分能量攻击和故障攻击方法;2018 年,Castelnovi 等人[69]首次提出了针对 SPHINCS、GRAVITY-SPHINCS 和 SPHINCS+等算法的底层框架的故障攻击,它允许创建适用于所有 NIST 一轮候选算法(XMSS、LMS、SPHINCS+和 Gravity-SPHINCS)的通用签名伪造.同年,Genêt 等人[70]针对 SPHINCS 密码算法也提出了故障攻击方案,并且在文献[71]中针对 SHA-2 和 BLAKE-256 两种哈希算法提出了差分能量攻击方案,且将该攻击应用于 XMSS、BLAKE、SPHINCS 等方案.

针对基于哈希的后量子密码的故障攻击主要目的是迫使底层 OTS 被重用.下面简述文献[70]中的攻击方法.首先,攻击者必须能够对消息 M 进行签名以获得有效的签名.如果重新签名相同的消息 M,该方案将生成相同的签名以及完全相同的超树路径.但是,如果在构造任何子树(非树根) $0\leqslant i<d-1$ 的过程中发生错误,算法将输出不同的签名.对已知子树的签名进行伪造.对 M 的签名进行剪修伪造,只需修改某个子树,其他部分均为原来正确的子树.如图 3 所示(图 3 来源于文献[70]):左边的图片显示了一个消息 M 正在与一个 SPHINCS 超树签名,而高亮显示的子树正在受到攻击.在右侧,超树在这个子树处被切割,并与一个伪造子树的分支进行嫁接,这一步是使用注入电压故障的方式来让底层的 FTS 得到重用,从而实现对任意消息 M' 进行签名.最后,当故障攻击迫使底层的 OTS(FTS)被重用时,利用 Genêt 等人提出的处理算法来识别错误签名中的 WOTS 密钥值(使用 WOTS 实例的公钥,可以通过猜测损坏的子树根的所有块来识别密钥值).

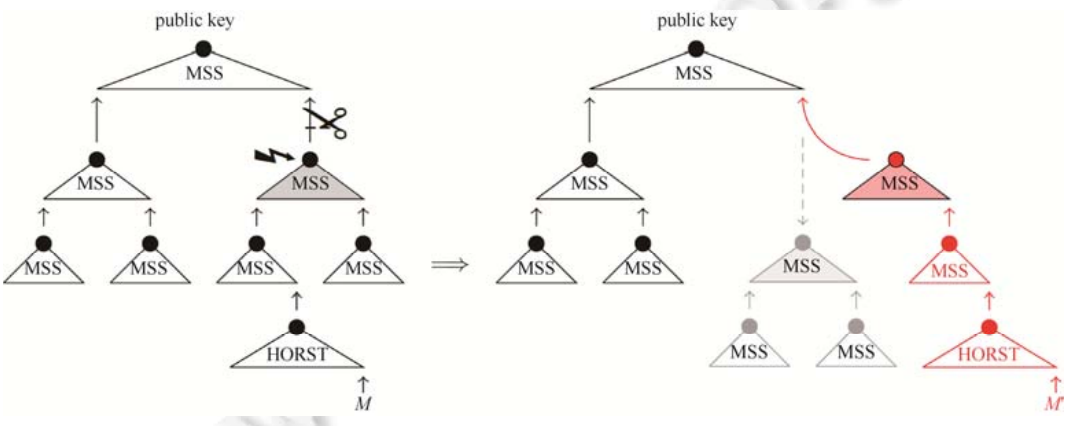

Fig.3 An illustration of the tree grafting against SPHINCS

图 3 攻击分析图

基于哈希的后量子密码算法的侧信道攻击可通过少量错误签名即可完成签名伪造.文献[69]的实验结果表明,仅需要收集 3 条错误签名就可以实现 GRAVITY-SPHINCS 的伪造签名,而伪造代价是大约 2^{20} 个哈希计算.同时,如果拥有更多的错误签名数量,那么所消耗的计算代价将会更低.如当拥有 10 个时,所需代价仅为 $2^{5.5}$ 个哈希计算,而当拥有 20 个时,仅需 4 个哈希计算.文献[70]指出,若给定 8 个错误签名,攻击者就可以在 64 次尝试内拥有超过 30%的伪造成功率;如果给定 15 个错误签名,那么攻击者可以在 35 次尝试内即可拥有超过 90%的伪造成功率.文献[71]所做的差分能量分析共需要收集 10 000 条能量迹,猜测计算值的空间大概为 2^{16}.

4.3 侧信道防御策略

基于哈希的后量子密码的侧信道防御策略研究工作主要如下:2018 年,Kannwischer 等人在文献[71]中通过对基于状态哈希的 XMSS、SPHINCS 等签名方案进行分析,提出了一种基于 SHA2 的伪随机数发生器的新型差分能量分析方法,并建议使用隐藏技术来作为抵御策略,利用对混合过程的隐藏来减少相关性,使得它们的执行顺序随机排列,迫使攻击者对收集到的能量迹进行同步对齐,从而增加了 DPA 的复杂性.而在文献[69]中,Castelnovi 等人针对 SPHINCS 提出了故障攻击,并且也指出可使签名计算冗余,使攻击复杂化,但这种方式会带来巨大的开销(时间和空间上).他们还推荐了一种有效的对策:以某种方式将超树的不同层连接起来,如此,计算树的错误将导致一个无效的签名,即一个与公钥不同的根值.除此之外,文献[72]也提出了一种特殊的重新计算方法,旨在避免 Merkle 树中的错误,称为交换节点(RESN)的重新计算,可以实现故障检测.

故障检测技术是基于哈希的后量子密码算法应对故障攻击的高效防御策略.文献[72]提出了一种故障检测策略以抵御故障攻击,目的是检测出错误计算从而避免签名伪造,根据其作者的实现结果可知,针对 BLAKE 方案,所付出的代价在面积开销和吞吐量上分别下降为 33.1%和 14.5%;针对 SPONGENT 的几个变种方案,在面积开销方面都额外增加了 22%~24%不等,吞吐量最少时增加了 8.3%.

5 基于多变量的 PQC

5.1 基于多变量的密码系统

基于多变量的密码系统的数学难题是求解一个有限域上的非线性多变量方程式组.通常,多变量密码系统的性能较优,不需要过多的计算资源,这使其对低成本设备中的应用程序具有吸引力.在多变量密码系统中,HFE(hidden field equations cryptosystem)[73]和 UOV(unbalanced Oil-and-Vinegar)[74]签名方案是较早的、研究较多的以及最为广泛的两种密码系统.最早的多变量公钥密码系统是由 Matsumoto 与 Imai 在 1988 年提出来的 MI(Matsumoto-Imai)加密算法[75],不过后来 Patarin[76]攻破了 MI 加密算法.之后,研究学者们继续对 MI 密码进行研究改进,目前主要使用的多变量签名方案是 HFE 和 UOV 的变种,如 NIST 二轮中的 LUOV 方案是 UOV 的变种,而 NIST 二轮中 GeMSS 签名方案是 HFE 的变种.自从 1997 年 Patarin[77]提出“UOV 方案”后,UOV 已成功地经受住了近 20 年密码分析的检验.该方案简单,签名小,速度快,因此,目前许多基于多变量的密码系统均采用该方案或基于该方案进行改善和变种,但 UOV 也存在一定的缺点,主要缺点是其公钥非常大.下面简述 UOV 的基本原理.

UOV 签名方案需要使用单向函数(即不可逆映射)作为限门函数.由 n 个变量 m 个方程组成的多变量二次方程式 $P=(p^{(1)},...,p^{(m)})$ 可表示如下:

$$P^{(k)}(x_1,...,x_n)=\sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij}^{(k)}x_ix_j+\sum_{j=1}^{n}p_i^{(k)}x_i+p_0^{(k)},\ \text{其中},\ k=1,...,m,$$

而 p_{ij}^{k}、p_i^{k}、p_0^{k} 都是 E_q 上随机选择.

上面的限门函数 P 可以分解为 $P=F\circ T$, 其中,T 为可逆线性映射函数,F 是可逆二次映射函数,如下:设 n 变量的 m 可逆二次映射为 $F:F_q^n\to F_q^m$,也称为中心映射.令 $V=\{1,...,v\},O=\{v+1,...,n\}$, 其中,$|V|=v,|O|=o$, 则 n 个变量的多元二项式方程 $F=(f^{(1)},...,f^{(m)})$ 定义为

$$F^{(k)}(x)=\sum_{i\in O,j\in V}\alpha_{ij}^{(k)}x_ix_j+\sum_{i,j\in V,i\leqslant j}\beta_{ij}^{(k)}x_ix_j+\sum_{i\in V\cup O}\gamma_i^{(k)}x_i+\eta^{(k)},$$

其中, $x=(x_1,...,x_n), k=1,...,o, n=v+o, m=0$, o 变量称为油变量,v 称为醋变量.

下面简单描述如何利用 F、T 来计算恢复 P.即给定自变量 x,目标求因变量 y,使得 $P(y)=x$.首先,F 是可逆二次函数,通过计算 $F(y')=x$, 求出 y',然后通过线性映射 T,计算 $y=T^{-1}(y')$, 从而可以得出 $P(y)=x$.为了减少密钥大小,一般地,在设计方案时,只保存私钥种子和公钥种子,而不存储整个 F 和 T,这些种子的字节数少,占用空间小,在需要使用公私钥时,再重新调用相应的公私钥对生成函数即可.

5.2 针对多变量PQC的侧信道攻击

近几年来,基于多变量 PQC 的侧信道攻击工作较少,主要为能量分析和故障攻击.在 2018 年,Park 和 Shim 等人[78]针对基于多变量的 Rainbow 方案提出了相关能量攻击方案,利用等效密钥可以完全恢复出完整密钥;2019 年,Krämer 等人[79]针对 UOV 和 Rainbow 算法提出了故障攻击方案;2020 年,Shim 等人[80]又针对 Rainbow 算法提出了新的故障攻击方案,目的是诱导签名中使用的随机醋变量发生故障,其故障模型根据醋值的泄漏类型分为 3 种情况:重复使用、泄露和置零,并且证明了 UOV 的等价密钥在多项式时间内可完全恢复.

下面简述针对基于多变量的后量子密码的相关能量分析原理.以文献[78]为例,该方案使用了等价密钥的概念,这里简单描述等价密钥的概念:设 $GL_m(F_q)$在 F_q 上的 m 维的线性群找到两个线性可逆映射, $\Sigma\in GL_m(F_q)$, $\Omega\in GL_n(F_q)$, 满足 $P=S\circ F\circ T=S\circ\Sigma^{-1}\circ(\Sigma\circ F\circ\Omega)\circ\Omega^{-1}\circ T$. 令 $S'=S\circ\Sigma^{-1}, F'=F, T'=\Omega^{-1}\circ T$, 则称 (S',F',T') 为等价密钥.Park 等人提出的攻击点是以矩阵向量积运算的位置为目标,最后恢复出密钥仿射映射 S、T、F.攻击原理和步骤大致如下.

首先利用签名中的矩阵向量乘积:即计算 $\alpha=\tilde{S}(h)$, 恢复出 S.假设矩阵为 A,向量为 X,则矩阵向量的乘积为 X',可以表示为 $x'=\begin{bmatrix}x'_1\\x'_1\\\vdots\\x'_n\end{bmatrix}=Ax^T=\begin{bmatrix}a_{11}&\cdots&a_{1n}\\\vdots&\ddots&\vdots\\a_{n1}&\cdots&a_{nn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$, 其中, $x'_i=\sum_n^{j=1}a_{ij}\cdot x_j=a_{i1}\cdot x_1+a_{i2}\cdot x_2+...+a_{in}\cdot x_n, 1\leqslant i\leqslant n.$

控制向量 X 的所有元素来恢复仿射映射 S,就可以使用中间结果来获得矩阵 A 第 i 行的所有元素.获得 S 后需要继续恢复 T',这里可以利用签名中最后一步的矩阵向量乘积:计算 $\sigma=\tilde{T}(\beta)$, 除此之外,还需要用到等价密钥的概念.假设等价密钥的矩阵向量乘积如图 4 所示(图 4 来源于文献[78]),可以清晰地看到:

$$x'_{v1+o_1+1}=x_{v1+o_1+1}, x'_{v1+o_1+2}=x_{v1+o_1+2},...,x'_n=x_n,$$

因此可以利用矩阵 A 的第 v_1+o_1+1 行到 n 行恢复出 A_2,利用下面的中间值来恢复 $Guess\cdot x_k, v1+o_1+1\leqslant k\leqslant n.$

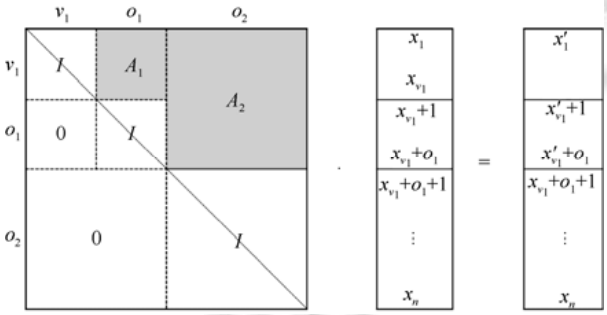

Fig.4 Matrix-vector product using the equivalent key

图 4 矩阵 A 的等价密钥图

得到 A_2 后,再利用等式 $x'_{v_1+t}=1\cdot x_{v_1+t}+\sum_{k=1}^{o_2}A^{(2)}_{v_1+t,k}\cdot x_{(v_1+o_1+k)}, 1\leqslant t\leqslant o_1$, 其中, $A^2_{i,j}(1\leqslant i\leqslant v_1+o_1, 1\leqslant j\leqslant o_2)$ 表示子矩阵 A_2 的第(i,j)个元素.这里,利用 $x_{v1+1}, x_{v_1+2},..., x_{v_1+o_1}$ 与 A_1 相乘,然后利用 $Guess\cdot x_k, v+1\leqslant k\leqslant v_1+o_1$ 中间值得到 A_1 的所有元素,再根据 S,T'即可轻松利用线性多项式求解 T 和 F.

基于多变量 PAC 的侧信道攻击效果较好,具有较高的成功率.文献[78]中,只用 30 个能量迹就可恢复出正确

密钥;文献[79]中,针对Rainbow方案的不同安全级别的参数,其成功率均在89%~93%;在特殊情况下,如在矩阵A是可逆的情况下,在固定醋变量时,对于有限域F_{16}、F_{31}、F_{256},其成功率分别达到93.3%、96.6%、99.6%.

5.3 侧信道防御策略

抵御能量分析最常用的对策是在算法增加隐藏或掩码技术.针对基于哈希的密码算法的防护近几年的工作较少,有一部分原因在于哈希签名方案的安全性,在很大程度上依赖于所选取的哈希函数,因此,单独针对签名方案的攻击与防御较少.在文献[78]中,作者除了提出等价密钥的CPA攻击外,还推荐在实现中使用随机仿射映射以抵御CPA攻击,如虚拟操作的随机插入和操作的变换是一种常用的隐藏技术,并且,作者还推荐了另一种方法——对随机数使用逻辑屏蔽的方法.在防御代价上,文献[65]使用消息随机化得到的矩阵向量与一般矩阵向量乘积相比,需要额外使用$2n$个乘法和1个求逆(n为矩阵维度),无防护算法需要n^2个乘法和加法,因此从总体来看,这种防护策略所带来的额外开销较少.

6 讨 论

本节我们对不同类型的后量子密码方案的特点以及它们的攻击点和防护策略进行总结,讨论未来可能的威胁和防御策略,并对未来的前景加以展望.

6.1 后量子密码的特性

后量子密码系统主要基于以下几种不同的数学难题:基于格、基于编码、基于多变量、基于哈希、基于超奇异同源等难题,其中,基于格的后量子密码系统在性能上拥有很好的优势,部分原因得益于其使用了NTT,降低了计算复杂度;基于编码的后量子密码系统也是一类比较有竞争力的密码系统,其在性能上拥有很好的表现,而且基于编码的密码系统因底层基于纠错码,因此本身拥有很好的纠错能力;基于超奇异同源的密码系统具有公钥尺寸小的特点,这有利于应用在嵌入式和物联网等资源受限的芯片中,但目前在性能上远不如其他类型的公钥密码系统;基于多变量密码系统主要用于签名方案(NIST二轮4个,CACR无),因为这类签名方案要求的硬件资源非常少,签名速度快,所以很适合用于低功耗设备,比如智能卡、RFID芯片等;基于哈希的密码系统与其他密码系统相比,这种方案的主要优点是其安全性仅依赖于泛型哈希函数的某些加密属性,因此,如果所选的哈希函数将来被破坏掉,则可以很容易地用新的哈希结构替换它们.另外,近年来总体研究方向更偏向于基于格和基于编码,根据本文调研近几年的数据来看,格基密码系统是目前研究数量最多、成果最多(主要是指在性能优化以及安全实现方面),编码密码系统次之,这两类密码系统均主要用于公钥加密/密钥封装.

6.2 针对不同类型后量子密码的侧信道攻击与防御

(1) 侧信道攻击对后量子密码算法的安全性具有较大影响.在基于格方面,侧信道攻击主要从采样器(离散、中心二项式、拒绝等)、NTT、多项式乘法、纠错码等几个核心部位进行攻击,值得注意的是,多项式乘法、NTT和采样器是被攻击比较多的部位;在基于编码方面,侧信道攻击主要从纠错码、校验子计算、稀疏矩阵的计算以及解密过程等部位进行攻击,其中,校验子计算和纠错码是被核心攻击的部位,而且主要攻击手段是时间攻击;在基于哈希方面,侧信道攻击主要是以底层的HORST和伪随机发生器作为攻击点;基于多变量的主要攻击点在于矩阵向量乘积;最后在基于超奇异同源方面,本文没有细述,但根据我们查阅的数据来看,主要攻击点有蒙哥马利梯度算法.对攻击点情况的总结见表3.

Table 3 The attack point of different cryptosystems

表3 不同类型密码系统的攻击点

类型	攻击点
格	采样器(离散、中心二项式、拒绝等)、NTT、多项式乘法、纠错码
编码	纠错码、校验子计算、稀疏矩阵的计算以及解密过程
多变量	矩阵向量乘积
哈希	HORST和伪随机发生器
超奇异	蒙哥马利梯度算法

(2) 在防护方面,针对能量分析攻击主要存在的攻击点应加以防护,而防护主要有两种手段:隐藏(随机化)和掩码.根据防护策略的代价数据,掩码的代价相对隐藏技术会高一些,如针对校验矩阵的掩码对策的代价与无保护方案相比,大概增加了 8 倍的资源,而使用消息随机化得到的矩阵向量与一般矩阵向量乘积相比,需要额外使用 2n 个乘法和 1 个求逆,所需代价不到 1 倍;而时间攻击的最好防护手段是采用恒定时间实现;故障攻击的防护策略是故障检测.根据本文对所调研的文献工作进行统计,汇总成表 4,其中,2x 代表增加 2 倍的代价(这里,代价以性能为主).通常,密码系统的防护可以分成 3 类:应对能量分析的防护、时间攻击的防护和故障攻击的防护.时间攻击的防护主要采用恒定时间实现,即将算法进行恒定时间实现或接近恒定时间实现(多次执行同一算法的时间几乎一样);而能量分析最常用的防护策略有两种:随机化和掩码方案,其中,随机化在理论上只是增加攻击难度(这种难度很高,但非完全不能攻破),掩码方案可有效抵御(n 阶掩码抵御 n 阶攻击);对于故障攻击的防护方法为故障检测和随机化.

Table 4 Counter measures and its costs of different attacks

表 4 不同攻击类型的防护手段及代价

攻击类型	防护手段	防护代价
能量分析攻击	隐藏/掩码	<1x/2x~15x
时间攻击	恒定时间实现	<1x
故障攻击	随机化/故障检测	<1x

6.3 后量子密码未来可能的威胁和防御对策

(1) 未来可能潜在的侧信道攻击:未来可能的侧信道攻击可能来自两个方面,一方面来自目前已存在的侧信道攻击的新型混合攻击,上文已列出主要的经典侧信道攻击手段,可考虑混合多种上述攻击方法,结合多种手段、侧信息和分析方法可能发掘出新的混合攻击方式,再利用人工智能、AI 等手段,缩短分析时间,以提高攻击效率;另一方面来自于挖掘新的侧信息资源,比如在一些新的硬件平台上(Intel PT/TSX/MPX/CAT 等和 ARM v8 架构的 ARM Cotex A77/A78 等平台)会有新的硬件特性,这些新特性可能存在新的侧信息泄露.

(2) 可能的防御方案:目前已有许多文章给出了防御侧信道攻击的手段,见表 4,对应不同的侧信道攻击手段采用相应的防护策略,一般来说,可以分成以下两个方面的解决方案:a) 软件方面:这里主要为源码层次的解决方案,可以多采用一些系统特性来预防和检测攻击,如:随机化技术、时间异常检测技术、检测可疑异常和中断等,同时采用并行技术,减少单一数据的侧信息泄露,增加噪声,以达到分析攻击的难度.b) 硬件层面:首先硬件层面的优化实现本身的工程量大,复杂度高,对应的优化实现解决方案也一直处于探索阶段.在此基础上,侧信道防护策略必将增加额外的资源消耗和设计难度,因此,本文推荐一种可能的解决方案:采用软/硬件协同方式进行实现,将耗时关键的核心算子使用硬件加速实现,其余部分使用软件实现,既保证了高效性,又降低了复杂度和工程量.另外,硬件实现虽然复杂度高,但目前也有一些全硬件实现的解决方案,因此可以参考目前已有的解决方案进而加以改善.

6.4 现存问题及未来前景展望

最后,我们对后量子密码存在的问题和未来的前景展望给出简单论述.目前,后量子存在的问题主要体现在性能、攻击、防御这 3 个方面,当然也存在其他因素,比如公钥尺寸大小、解密错误率等,这些因素都会影响算法标准的制定,但是性能和安全性是标准最核心的评价因素,比如,超奇异同源密码具有良好的公钥尺寸小的特点,但性能过低,相对于格密码,超奇异同源的性能会比最快的密码方案慢千倍之多,因此性能成为了超奇异同源密码最主要的瓶颈.NIST 第 2 轮中,也重点评比了性能和安全性,根据 NIST 官方的数据,目前加密方案中,综合性能、参数、公钥尺寸等因素,格密码具有最高的评价;在公钥大小、密文大小方面,超奇异同源最优;在性能方面,格密码最优,编码次之.而签名方案中,签名字节最小的是多变量,格密码排在中等;性能上,多变量和格密码较优;综合来看,多变量的签名方案综合评分最高,格密码次之.对于后量子密码未来的前景展望,由于量子计算机的研究不断地取得进展,传统的公钥密码算法领域将都会被后量子密码所取代,并且,在新型产业,如物联网、5G、卫星通信、军事国防、区块链、数字货币、数字签名等领域都是后量子密码广泛应用的领域,也是亟需安

全可靠的后量子密码来保障数据安全的领域.同时,后量子密码也衍生出新的研究领域,如同态加密、基于属性加密等,这些领域也是目前比较热门的研究领域.未来后量子密码的应用,无论是用于上述新兴产业领域还是其衍生领域,侧信道分析都是后量子密码的安全门关,在应用于这些领域之前,都会对于后量子密码进行侧信道安全测评,因此,侧信道安全分析和测评会是未来后量子密码重点研究的关键点.为了使用上述产业领域,多平台的侧信道分析和测评会是未来后量子密码的侧信道分析中必要的解决方案.

7 总 结

本文针对最新的后量子密码方案的侧信道攻击与防御进行了调研,通过分析针对各类后量子密码侧信道的攻击与防御策略,从攻击方法、攻击点、攻击评价等多方面对不同后量子密码算法进行了侧信道攻击的深入分析,并从安全性和防护代价两个维度对侧信道防御策略进行了论述,总结了针对不同类型后量子密码的攻击点,以及不同攻击类型的防护手段及代价,最后讨论了后量子密码未来可能的侧信道攻击、解决方案,并对未来的前景进行了展望.

References:

[1] Shor PW. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Review, Society for Industrial and Applied Mathematics, 1999,41(2):303–332.

[2] Bindel N, Buchmann J, Krämer J. Lattice-based signature schemes and their sensitivity to fault attacks. In: Proc. of the 2016 Workshop on Fault Diagnosis and Tolerance in Cryptography (FDTC). 2016. 63–77.

[3] Khalid A, Rafferty C, Howe J, Brannigan S, Liu W, O'Neill M. Error samplers for lattice-based cryptography-challenges, vulnerabilities and solutions. In: Proc. of the 2018 IEEE Asia Pacific Conf. on Circuits and Systems (APCCAS). 2018. 411–414.

[4] Khalid A, Oder T, Valencia F, O'Neill M, Güneysu T, Regazzoni F. Physical protection of lattice-based cryptography: Challenges and solutions. In: Proc. of the 2018 on Great Lakes Symp. on VLSI. Chicago: Association for Computing Machinery, 2018. 365–370.

[5] Roy KS, Kalita HK. A survey on post-quantum cryptography for constrained devices. Int'l Journal of Applied Engineering Research, 2019,14(11):2608–2615.

[6] Valencia F, Oder T, Güneysu T, Regazzoni F. Exploring the vulnerability of R-LWE encryption to fault attacks. In: Proc. of the 5th Workshop on Cryptography and Security in Computing Systems. Association for Computing Machinery, 2018. 7–12.

[7] Drăgoi V, Richmond T, Bucerzan D, Legay A. Survey on cryptanalysis of code-based cryptography: From theoretical to physical attacks. In: Proc. of the 7th Int'l Conf. on Computers Communications and Control (ICCCC). 2018. 215–223.

[8] Nejatollahi H, Dutt N, Ray S, Regazzoni F, Banerjee I, Cammarota R. Post-quantum lattice-based cryptography implementations: A survey. ACM Computing Surveys, 2019,51(6):129:1–129:41.

[9] Ti YB. Fault attack on supersingular isogeny cryptosystems. In: Lange T, Takagi T, eds. Post-quantum Cryptography. Cham: Springer Int'l Publishing, 2017. 107–122.

[10] Gélin A, Wesolowski B. Loop-abort faults on supersingular isogeny cryptosystems. In: Lange T, Takagi T, eds. Post-quantum Cryptography. Cham: Springer Int'l Publishing, 2017. 93–106.

[11] Koziel B, Azarderakhsh R, Jao D. Side-channel attacks on quantum-resistant supersingular isogeny Diffie-Hellman. In: Adams C, Camenisch J, eds. Selected Areas in Cryptography—SAC 2017. Cham: Springer International Publishing, 2018. 64–81.

[12] Computer Security Division ITL. Round 2 submissions-post-quantum cryptography|csrc. CSRC|NIST, 2017-01-03. (2017-01-03) [2020-05-25]. https://csrc.nist.gov/projects/post-quantum-cryptography/round-2-submissions

[13] https://sfjs.cacrnet.org.cn/site/term/list_77_1.html

[14] Kocher PC. Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In: Koblitz N, ed. Advances in Cryptology—CRYPTO'96. Berlin, Heidelberg: Springer-Verlag, 1996. 104–113.

[15] Kocher P, Jaffe J, Jun B. Differential power analysis. In: Wiener M, ed. Proc. of the Advances in Cryptology—CRYPTO'99. Berlin, Heidelberg: Springer-Verlag, 1999. 388–397.

[16] Brier E, Clavier C, Olivier F. Correlation power analysis with a leakage model. In: Joye M, Quisquater J-J, eds. Proc. of the Cryptographic Hardware and Embedded Systems—CHES 2004. Berlin, Heidelberg: Springer-Verlag, 2004. 16–29.

[17] Chari S, Rao JR, Rohatgi P. Template attacks. In: Kaliski BS, Koççetin K, Paar C, eds. Proc. of the Cryptographic Hardware and Embedded Systems—CHES 2002. Berlin, Heidelberg: Springer-Verlag, 2003. 13–28.

[18] Biham E, Shamir A. Differential fault analysis of secret key cryptosystems. In: Kaliski BS, ed. Proc. of the Advances in Cryptology—CRYPTO'97. Berlin, Heidelberg: Springer-Verlag, 1997. 513–525.

[19] Mathan SA, Koedinger KR. Fostering the intelligent novice: Learning from errors with metacognitive tutoring. Educational Psychologist, Routledge, 2005,40(4):257–265.

[20] Lyubashevsky V, Peikert C, Regev O. On ideal lattices and learning with errors over rings. In: Gilbert H, ed. Proc. of the Advances in Cryptology–EUROCRYPT 2010. Berlin, Heidelberg: Springer-Verlag, 2010. 1–23.

[21] Langlois A, Stehlé D. Worst-case to average-case reductions for module lattices. Designs, Codes and Cryptography, 2015,75(3): 565–599.

[22] Saarinen M-JO. Arithmetic coding and blinding countermeasures for lattice signatures. Journal of Cryptographic Engineering, 2018, 8(1):71–84.

[23] Pessl P. Analyzing the shuffling side-channel countermeasure for lattice-based signatures. In: Dunkelman O, Sanadhya SK, eds. Proc. of the Progress in Cryptology–INDOCRYPT 2016. Cham: Springer Int'l Publishing, 2016. 153–170.

[24] Primas R, Pessl P, Mangard S. Single-trace side-channel attacks on masked lattice-based encryption. In: Fischer W, Homma N, eds. Proc. of the Cryptographic Hardware and Embedded Systems—CHES 2017. Cham: Springer Int'l Publishing, 2017. 513–533.

[25] Kim S, Hong S. Single trace analysis on constant time CDT sampler and its countermeasure. Applied Sciences, Multidisciplinary Digital Publishing Institute, 2018,8(10):1809.

[26] Pessl P, Primas R. More practical single-trace attacks on the number theoretic transform. In: Schwabe P, Thériault N, eds. Proc. of the Progress in Cryptology—LATINCRYPT 2019. Cham: Springer Int'l Publishing, 2019. 130–149.

[27] Huang W-L, Chen J-P, Yang B-Y. Power analysis on NTRU prime. IACR Trans. on Cryptographic Hardware and Embedded Systems, 2020, 123–151.

[28] Ravi P, Roy SS, Chattopadhyay A, Bhasin S. Generic side-channel attacks on CCA-secure lattice-based PKE and KEMS. IACR Trans. on Cryptographic Hardware and Embedded Systems, 2020,2019:307–335.

[29] Bruinderink LG, Pessl P. Differential fault attacks on deterministic lattice signatures. IACR Trans. on Cryptographic Hardware and Embedded Systems, 2018, 21–43.

[30] Ravi P, Roy DB, Bhasin S, Chattopadhyay A, Mukhopadhyay D. Number "not used" once—practical fault attack on PQM4 implementations of nist candidates. In: Polian I, Stöttinger M, eds. Proc. of the Constructive Side-channel Analysis and Secure Design. Cham: Springer Int'l Publishing, 2019. 232–250.

[31] McCarthy S, Howe J, Smyth N, Brannigan S, O'Neill M. BEARZ attack falcon: implementation attacks with countermeasures on the falcon signature scheme. In: Proc. of the SECRYPT. 2019. 61–71.

[32] Valencia F, Polian I, Regazzoni F. Fault sensitivity analysis of lattice-based post-quantum cryptographic components. In: Pnevmatikatos DN, Pelcat M, Jung M, eds. Proc. of the Embedded Computer Systems: Architectures, Modeling, and Simulation. Cham: Springer Int'l Publishing, 2019. 107–123.

[33] Albrecht MR, Deo A, Paterson KG. Cold boot attacks on ring and module LWE keys under the NTT. IACR Trans. on Cryptographic Hardware and Embedded Systems, 2018, 173–213.

[34] D'Anvers J-P, Tiepelt M, Vercauteren F, Verbauwhede I. Timing attacks on error correcting codes in post-quantum secure schemes. IACR Cryptology ePrint Archive, 2019,2019:292.

[35] Espitau T, Fouque P-A, Gérard B, Tibouchi M. Side-channel attacks on Bliss lattice-based signatures: Exploiting branch tracing against strong swan and electromagnetic emanations in microcontrollers. In: Proc. of the 2017 ACM SIGSAC Conf. on Computer and Communications Security. Dallas: Association for Computing Machinery, 2017. 1857–1874.

[36] Roy SS, Reparaz O, Vercauteren F, Verbauwhede I. Compact and side channel secure discrete gaussian sampling. IACR Cryptology ePrint Archive, 2014,2014:591.

[37] Oder T, Schneider T, Pöppelmann T, Güneysu T. Practical CCA2-secure and masked ring-LWE implementation. IACR Trans. on Cryptographic Hardware and Embedded Systems, 2018, 142–174.

[38] Reparaz O, Sinha Roy S, Vercauteren F, Verbauwhede I. A masked ring-LWE implementation. In: Güneysu T, Handschuh H, eds. Proc. of the Cryptographic Hardware and Embedded Systems—CHES 2015. Berlin, Heidelberg: Springer-Verlag, 2015. 683–702.

[39] Reparaz O, De Clercq R, Roy SS, Vercauteren F, Verbauwhede I. Additively homomorphic ring-LWE masking. In: Takagi T, ed. Proc. of the Post-quantum Cryptography. Cham: Springer Int'l Publishing, 2016. 233–244.

[40] Barthe G, Belaïd S, Espitau T, Fouque P-A, Grégoire B, Rossi M, Tibouchi M. Masking the GLP lattice-based signature scheme at any order. In: Nielsen JB, Rijmen V. Proc. of the Advances in Cryptology—EUROCRYPT 2018. Cham: Springer Int'l Publishing, 2018. 354–384.

[41] Espitau T, Fouque P-A, Gérard B, Tibouchi M. Loop-abort faults on lattice-based Fiat-Shamir and hash-and-sign signatures. In: Avanzi R, Heys H, eds. Proc. of the Selected Areas in Cryptography—SAC 2016. Cham: Springer Int'l Publishing, 2017. 140–158.

[42] Howe J, Khalid A, Martinoli M, Regazzoni F, Oswald E. Fault attack countermeasures for error samplers in lattice-based cryptography. In: Proc. of the 2019 IEEE Int'l Symp. on Circuits and Systems (ISCAS). 2019. 1–5.

[43] Khalid A, Howe J, Rafferty C, O'Neill M. Time-independent discrete Gaussian sampling for post-quantum cryptography. In: Proc. of the 2016 Int'l Conf. on Field-Programmable Technology (FPT). 2016. 241–244.

[44] Micciancio D, Walter M. Gaussian sampling over the integers: Efficient, generic, constant-time. In: Katz J, Shacham H, eds. Proc. of the Advances in Cryptology—CRYPTO 2017. Cham: Springer Int'l Publishing, 2017. 455–485.

[45] Karmakar A, Roy SS, Reparaz O, Vercauteren F, Verbauwhede I. Constant-time discrete Gaussian sampling. IEEE Trans. on Computers, 2018,67(11):1561–1571.

[46] Karmakar A, Roy SS, Vercauteren F, Verbauwhede I. Pushing the speed limit of constant-time discrete Gaussian sampling: A case study on the falcon signature scheme. In: Proc. of the 56th Annual Design Automation Conf. 2019. Las Vegas: Association for Computing Machinery, 2019. 1–6.

[47] Barthe G, Belaïd S, Espitau T, Fouque P-A, Rossi M, Tibouchi M. GALACTICS: Gaussian sampling for lattice-based constant-time implementation of cryptographic signatures, revisited. In: Proc. of the 2019 ACM SIGSAC Conf. on Computer and Communications Security. London: Association for Computing Machinery, 2019. 2147–2164.

[48] Walters M, Roy SS. Constant-time bch error-correcting code. IACR Cryptology ePrint Archive, 2019,2019:155.

[49] McEliece RJ. A public-key cryptosystem based on algebraic. Coding THV, 1978,4244:114–116.

[50] Overbeck R, Sendrier N. Code-based cryptography. In: Bernstein DJ, Buchmann J, Dahmen E, eds. Proc. of the Post-quantum Cryptography. Berlin, Heidelberg: Springer-Verlag, 2009. 95–145.

[51] Huffman WC, Pless V. Fundamentals of ERror-correcting Codes. Cambridge: Cambridge University Press, 2010.

[52] Von Maurich I, Güneysu T. Towards side-channel resistant implementations of QC-MDPC McEliece encryption on constrained devices. In: Mosca M, ed. Proc. of the Post-quantum Cryptography. Cham: Springer Int'l Publishing, 2014. 266–282.

[53] Chen C, Eisenbarth T, Von Maurich I, Steinwandt R. Differential power analysis of a McEliece cryptosystem. In: Malkin T, Kolesnikov V, Lewko AB, Polychronakis M, eds. Proc. of the Applied Cryptography and Network Security. Cham: Springer Int'l Publishing, 2015. 538–556.

[54] Chen C, Eisenbarth T, Von Maurich I, Steinwandt R. Masking large keys in hardware: A masked implementation of McEliece. In: Dunkelman O, Keliher L, eds. Proc. of the Selected Areas in Cryptography—SAC 2015. Cham: Springer Int'l Publishing, 2016. 293–309.

[55] Chou T. QcBits: Constant-time small-key code-based cryptography. In: Gierlichs B, Poschmann AY, eds. Proc. of the Cryptographic Hardware and Embedded Systems—CHES 2016. Berlin, Heidelberg: Springer-Verlag, 2016. 280–300.

[56] Rossi M, Hamburg M, Hutter M, Marson ME. A side-channel assisted cryptanalytic attack against QCbits. In: Fischer W, Homma N, eds. Proc. of the Cryptographic Hardware and Embedded Systems—CHES 2017. Cham: Springer Int'l Publishing, 2017. 3–23.

[57] Sim B-Y, Kwon J, Choi KY, Cho J, Park A, Han D-G. Novel side-channel attacks on quasi-cyclic code-based cryptography. IACR Trans. on Cryptographic Hardware and Embedded Systems, 2019, 180–212.

[58] Strenzke F, Tews E, Molter HG, Overbeck R, Shoufan A. Side channels in the McEliece PKC. In: Buchmann J, Ding J, eds. Proc. of the Post-quantum Cryptography. Berlin, Heidelberg: Springer-Verlag, 2008. 216–229.

[59] Eaton E, Lequesne M, Parent A, Sendrier N. QC-MDPC: A timing attack and a CCA2 KEM. In: Lange T, Steinwandt R, eds. Proc. of the Post-quantum Cryptography. Cham: Springer Int'l Publishing, 2018. 47–76.

[60] Paiva TB, Terada R. A timing attack on the HQC encryption scheme. In: Paterson KG, Stebila D, eds. Proc. of the Selected Areas in Cryptography—SAC 2019. Cham: Springer Int'l Publishing, 2020. 551–573.

[61] Wafo-Tapa G, Bettaieb S, Bidoux L, Gaborit P. A practicable timing attack against HQC and its countermeasure. Cryptology ePrint Archive, Report, 2019/909, 2019.

[62] Danner J, Kreuzer M. A fault attack on the niederreiter cryptosystem using binary irreducible GOPPA codes. arXiv: 2002.01455 [cs, math], 2020.

[63] Petrvalsky M, Richmond T, Drutarovsky M, Cayrel P-L, Fischer V. Differential power analysis attack on the secure bit permutation in the McEliece cryptosystem. In: Proc. of the 26th Int'l Conf. Radioelektronika (RADIOELEKTRONIKA). 2016. 132–137.

[64] Chen C, Eisenbarth T, Von Maurich I, Steinwandt R. Horizontal and vertical side channel analysis of a McEliece cryptosystem. IEEE Trans. on Information Forensics and Security, 2016,11(6):1093–1105.

[65] Lamport L. Constructing digital signatures from a one-way function. Technical Report, CSL-98, SRI Int'l, 1979.

[66] Buchmann J, Dahmen E, Klintsevich E, Okeya K, Vuillaume C. Merkle signatures with virtually unlimited signature capacity. In: Katz J, Yung M, eds. Proc. of the Applied Cryptography and Network Security. Berlin, Heidelberg: Springer-Verlag, 2007. 31–45.

[67] Buchmann J, Dahmen E, Hülsing A. XMSS—a practical forward secure signature scheme based on minimal security assumptions. In: Yang B-Y, ed. Proc. of the Post-quantum Cryptography. Berlin, Heidelberg: Springer-Verlag, 2011. 117–129.

[68] Genet A. Hardware attacks against hash-based cryptographic algorithms. Infoscience, 2017-08-18. (2017-08-18)[2020-05-19]. https://infoscience.epfl.ch/record/253317

[69] Castelnovi L, Martinelli A, Prest T. Grafting trees: A fault attack against the sphincs framework. In: Lange T, Steinwandt R, eds. Proc. of the Post-quantum Cryptography. Cham: Springer Int'l Publishing, 2018. 165–184.

[70] Genêt A, Kannwischer MJ, Pelletier H, McLauchlan A. Practical fault injection attacks on sphincs. IACR Cryptology ePrint Archive, 2018,2018:674.

[71] Kannwischer MJ, Genêt A, Butin D, Krämer J, Buchmann J. Differential power analysis of XMSS and sphincs. In: Fan J, Gierlichs B, eds. Proc. of the Constructive Side-channel Analysis and Secure Design. Cham: Springer Int'l Publishing, 2018. 168–188.

[72] Mozaffari-Kermani M, Azarderakhsh R, Aghaie A. Fault detection architectures for post-quantum cryptographic stateless hash-based secure signatures benchmarked on asic. ACM Trans. on Embedded Computing Systems, 2016,16(2):59:1–59:19.

[73] Courtois NT. The security of hidden field equations (HFE). In: Naccache D, ed. Proc. of the Topics in Cryptology—CT-RSA 2001. Berlin, Heidelberg: Springer-Verlag, 2001. 266–281.

[74] Kipnis A, Patarin J, Goubin L. Unbalanced oil and vinegar signature schemes. In: Stern J, ed. Proc. of the Advances in Cryptology —EUROCRYPT'99. Berlin, Heidelberg: Springer-Verlag, 1999. 206–222.

[75] Matsumoto T, Imai H. Public quadratic polynomial-tuples for efficient signature-verification and message-encryption. In: Barstow D, Brauer W, Brinch Hansen P, Gries D, Luckham D, Moler C, Pnueli A, Seegmüller G, Stoer J, Wirth N, Günther C G, eds. Proc. of the Advances in Cryptology—EUROCRYPT'88. Berlin, Heidelberg: Springer-Verlag, 1988. 419–453.

[76] Patarin J. Cryptanalysis of the Matsumoto and Imai public key scheme of Eurocrypt'88. In: Coppersmith D, ed. Proc. of the Advances in Cryptology—CRYPT0'95. Berlin, Heidelberg: Springer-Verlag, 1995. 248–261.

[77] Patarin J. The oil and vinegar signature scheme. In: Proc. of the Dagstuhl Workshop on Cryptography. 1997.

[78] Park A, Shim K-A, Koo N, Han D-G. Side-channel attacks on post-quantum signature schemes based on multivariate quadratic equations: Rainbow and UOV. IACR Trans. on Cryptographic Hardware and Embedded Systems, 2018, 500–523.

[79] Krämer J, Loiero M. Fault attacks on UOV and rainbow. In: Polian I, Stöttinger M, eds. Proc. of the Constructive Side-channel Analysis and Secure Design. Cham: Springer Int'l Publishing, 2019. 193–214.

[80] Shim K-A, Koo N. Algebraic fault analysis of UOV and rainbow with the leakage of random vinegar values. IEEE Trans. on Information Forensics and Security, 2020, 1.

吴伟彬(1996－),男,学士,CCF 学生会员,主要研究领域为侧信道攻击与防御,密码工程,后量子密码.

杨昊(1997－),男,学士,CCF 学生会员,主要研究领域为密码工程,后量子密码.

刘哲(1986－),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为密码工程,后量子密码,侧信道攻击与防御.

张吉鹏(1999－),学士,主要研究领域为密码工程,后量子密码.