Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.33, NO.1, Feb. 2023 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2023.33.1.33

CRYSTALS-Dilithium 대상 비프로파일링 기반 전력 분석 공격 성능 개선 연구*

장 세 창,[1†] 이 민 종,[2] 강 효 주,[2] 하 재 철[3‡]
[1,2,3]호서대학교 (대학원생, 학생, 교수)

A Study on Performance Improvement of Non-Profiling Based Power Analysis Attack against CRYSTALS-Dilithium*

Sechang Jang,[1†] Minjong Lee,[2] Hyoju Kang,[2] Jaecheol Ha[3‡]
[1,2,3]Hoseo University (Graduate student, Student, Professor)

요 약

최근 미국의 국립표준기술연구소(NIST: National Institute of Standards and Technology)는 양자 내성 암호(PQC: Post-Quantum Cryptography, 이하 PQC) 표준화 사업을 진행하여 4개의 표준 암호 알고리즘을 발표하였다. 본 논문에서는 전자서명 분야에서 표준화가 확정된 CRYSTALS-Dilithium 알고리즘을 이용하여 서명을 생성하는 과정에서 동작하는 다항식 계수별 곱셈 알고리즘을 대상으로 비프로파일링 기반 전력 분석 공격인 CPA(Correlation Power Analysis)나 DDLA(Differential Deep Learning Analysis) 공격에 의해 개인키가 노출될 수 있음을 실험을 통해 증명한다. ARM-Cortex-M4 코어에 알고리즘을 탑재하여 실험 결과, CPA 공격과 DDLA 공격에서 개인 키 계수를 복구할 수 있음을 확인하였다. 특히 DDLA 공격에서 StandardScaler 전처리 및 연속 웨이블릿 변환을 적용한 전력 파형을 이용하였을 때 공격에 필요한 최소 전력 파형의 개수가 줄어들고 NMM(Normalized Maximum Margin) 값이 약 3배 증가하여 공격 성능이 크게 향상됨을 확인하였다.

ABSTRACT

The National Institute of Standards and Technology (NIST), which is working on the Post-Quantum Cryptography (PQC) standardization project, announced four algorithms that have been finalized for standardization. In this paper, we demonstrate through experiments that private keys can be exposed by Correlation Power Analysis (CPA) and Differential Deep Learning Analysis (DDLA) attacks on polynomial coefficient-wise multiplication algorithms that operate in the process of generating signatures using CRYSTALS-Dilithium algorithm. As a result of the experiment on ARM-Cortex-M4, we succeeded in recovering the private key coefficient using CPA or DDLA attacks. In particular, when StandardScaler preprocessing and continuous wavelet transform applied power traces were used in the DDLA attack, the minimum number of power traces required for attacks is reduced and the Normalized Maximum Margines (NMM) value increased by about 3 times. Conseqently, the proposed methods significantly improves the attack performance.

Keywords: CRYSTALS-Dilithium, Power Analysis Attack, Hardware Security, Deep Learning, Wavelet Transform

Received(12. 16. 2022), Modified(01. 25. 2023), Accepted(01. 26. 2023)

* 본 논문은 2021년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다.(No. 2021RIS-004)

* 이 논문은 2022년도 한국정보보호학회 동계 학술대회에 발표한 우수논문을 개선 및 확장한 것임.

† 주저자, wkdehreo55@naver.com

‡ 교신저자, jcha@hoseo.edu(Corresponding author)

I. 서 론

최근 양자 컴퓨터의 연산 능력 향상은 기존 암호 시스템들의 안전성을 위협하고 있다. 양자 컴퓨터에서 Shor 알고리즘을 이용하면 소인수 분해를 기반한 RSA와 같은 공개 키 암호시스템을 다항식 시간 안에 공격할 수 있다[1]. 또한, Grover 알고리즘을 이용하면 대칭 키 암호시스템의 비밀 키를 탐색하는데 필요했던 $O(N)$의 연산 작업을 $O(N^{1/2})$만의 연산으로 줄일 수 있다[2]. 따라서 기존 암호시스템들은 양자 컴퓨터 공격 능력으로부터 보호받기 위해 비밀 키의 길이를 늘이거나 긴 길이의 연산체를 사용해야 한다.

한편, 이와 같은 양자 컴퓨터의 등장에 대응하고자 NIST는 2016년부터 양자 컴퓨팅 환경에서 안전한 양자 내성 암호 표준화 사업을 진행 중이며, 3라운드가 종료되면서 표준화 확정 알고리즘이 발표되었다[3]. PKE/KEM 분야에서는 유일하게 CRYSTALS-KYBER 알고리즘이 확정되었으며, 전자서명 분야에서는 CRYSTALS-Dilithium, FALCON, SPHINCS$^+$ 알고리즘의 표준화가 확정되었다.

본 논문에서는 NIST PQC 전자서명 분야 표준화 알고리즘인 CRYSTALS-Dilithium을 대상으로 비프로파일링 기반 전력 분석 공격을 시도해 보고 그 안전성을 확인하고자 한다. Dilithium은 송신자가 자신의 개인 키로 서명을 생성하고 수신자는 송신자의 공개 키로 서명을 검증하게 된다. 따라서 공격은 서명 생성 단계에서 진행되며, 서명 생성 연산 중 다항식 계수별 곱셈 연산을 공격 지점으로 선정하고 최종적으로 개인 키의 다항식 계수를 복구한다.

공격은 비프로파일링 기반 전력 분석 공격의 일종인 CPA 공격과 DDLA 공격을 진행한다. 특히, DDLA 공격에서는 소비 전력 파형에 StandardScaler 표준화 및 연속 웨이블릿 변환을 이용함으로써 공격 성능을 크게 향상시켰다.

본 논문의 구성은 다음과 같다. 2장에서는 부채널 공격, CRYSTALS-Dilithium 전자서명 알고리즘, 연속 웨이블릿 변환에 대해 서술하며, 3장에서 Dilithium 비프로파일링 기반 전력 분석 공격 취약점 검증 실험을 진행하고, 4장에서 결론을 맺는다.

II. 배경 지식

2.1 부채널 공격

1996년 P. Kocher에 의해 처음으로 제안된 부채널 공격은 암호용 하드웨어에서 누출되는 부채널 정보를 분석하여 비밀 키 등의 비밀 정보를 복구하는 공격이다[4]. 이때 부채널 정보는 하드웨어에서 암호 알고리즘이 실행될 때 누설되는 부가적인 정보이며 대표적으로 전력, 전자파, 시간차, 소리 등이 있다. 이러한 부채널 공격은 암호용 하드웨어 뿐만 아니라 상용 컴퓨터 프로세서 내의 비밀 정보까지 찾아낼 수 있는 위협적인 공격이다[5, 6].

전력 분석 공격은 부채널 공격 분야에서 가장 활발히 연구되고 있는 공격이다. 전력 분석 공격은 마이크로 컨트롤러(MCU: MicroController Unit)와 같은 암호용 하드웨어에서 암호 알고리즘이 수행되는 동안 누출되는 소비 전력을 측정하여 비밀 정보를 복구한다. 이러한 공격이 가능한 이유는 암호용 하드웨어에서 동작 중인 명령어와 사용되는 데이터의 정보가 소비 전력 파형과 연관성이 있기 때문이다. 이러한 전력 분석 공격은 비밀 정보를 복구하는 방법에 따라 크게 프로파일링 기반 전력 분석 공격과 비프로파일링 기반 전력 분석 공격으로 나눌 수 있다.

2.1.1 프로파일링 기반 전력 분석 공격

프로파일링 기반 전력 분석 공격은 공격자가 공격 대상 장비와 동일한 장비를 보유하고 있어 암호 알고리즘에 입력되는 평문 또는 암호문에 매칭되는 중간 값을 알 수 있고, 그에 대응하는 소비 전력 파형을 측정할 수 있는 환경이 필요하다.

프로파일링 환경에서 공격자는 관심 지점(PoI: Point of Interest, 이하 PoI)의 전력 파형과 중간 값을 이용해 템플릿을 생성하는 템플릿 공격[7]과 전력 파형의 PoI 지점 혹은 전체 파형을 신경망 모델의 입력으로 사용하고 중간 값을 라벨로 설정하여 딥러닝 모델을 학습시키는 딥러닝 기반 프로파일링 공격[8, 9, 10]을 진행할 수 있다. 이러한 공격은 공격 준비 시 수많은 소비 전력 파형이 필요하지만, 실제 공격 대상의 매우 적은 수의 소비 전력 파형을 이용해 비밀 정보를 복구할 수 있다는 특징이 있다.

2.1.2 비프로파일링 기반 전력 분석 공격

비프로파일링 기반 전력 분석 공격은 공격자가 오직 공격 대상 장비에서 측정한 소비 전력 파형만을 통계 기법 등을 이용해 분석하는 공격 방법이다. 해당 공격은 프로파일링 대상 장비가 없기 때문에 추측 키를 이용해 공격 지점 중간 값을 계산한 후 전력 모델인 해밍 무게(HW: Hamming Weight, 이하 HW) 혹은 해밍 거리(HD: Hamming Distance) 모델을 통해 소비 전력을 가정하는 과정이 필요하다. 차분 전력 분석(DPA: Differential Power Analysis)[11]은 공격 대상 장비에서 수집한 다량의 전력 파형을 평균의 차를 이용하여 분석하는 공격 방법이며, 상관 전력 분석(CPA: Correlation Power Analysis, 이하 CPA)[12]은 공격 대상 장비에서 수집한 다량의 전력 파형을 상관계수를 이용하여 분석하는 공격 방법이다. 또한 DDLA(Differential Deep Learning Analysis) 공격[13]은 전력 파형을 신경망 모델의 입력으로 사용하고 중간 값의 HW, 최상위비트(MSB: Most Significant Bit, 이하 MSB), 최하위비트(LSB: Least Significant Bit, 이하 LSB), HW-based Binary[14] 등을 라벨로 설정하여 딥러닝 모델을 학습시켜 신경망의 학습 경향성을 토대로 비밀 정보를 복구하는 딥러닝 기반 비프로파일링 공격이다.

2.2 CRYSTALS-Dilithium 전자서명 알고리즘

CRYSTALS-Dilithium은 Module-LWE (Learning With Errors)와 SIS(Short Integer Solution)를 수학적 난제로 사용하는 격자 기반 전자서명 알고리즘으로 CRYSTALS-KYBER와 기본 특성 및 구조를 공유한다[15]. 또한, Gaussian sampling을 사용하는 FALCON[16]과 비교해 안정성과 효율성이 높은 Uniform sampling을 사용한다.

Dilithium은 다항식 환 $\mathcal{R}_q$를 $\mathbb{Z}_q[x]/(X^n+1)$와 같이 정의한다. 본 논문에서 사용한 Dilithium2는 $k=4$, $l=4$, $n=256$, $q=8{,}380{,}417$를 사용한다. 이때 랜덤 행렬 A의 차원을 구성하는 k와 l은 Dilithium의 보안 강도에 따라 다르다.

키 생성 단계에서는 먼저 난수 발생기(DRBG: Deterministic Random Bit Generator, 이하 DRBG)를 통해 임의의 a를 생성한 후 SHAKE 해시 함수에 a를 입력으로 하여 seed를 생성한다. 이후 생성한 seed를 이용해 랜덤 행렬 A를 생성한 뒤 샘플링을 통해 s_1, s_2를 생성한다. 마지막으로 As_1+s_2를 계산하고 t를 획득하여 최종적으로 공개키 $pk=(A,t)$와 개인 키 $sk=(A,t,s_1,s_2)$를 생성한다. Fig. 1은 Dilithium의 전체 알고리즘을 나타낸 것이다.

서명 생성 단계는 본 논문에서 진행한 공격의 공격 대상 연산이 진행되는 단계이다. 먼저 DRBG를 통해 랜덤한 y를 생성한다. 이때 SHAKE 해시 함수를 사용하며 비밀 키와 의존성 해소를 위해 rejection sampling을 사용한다. 이어서 $A \cdot y$를 계산한 뒤 반올림을 통해 상위 비트를 마스킹한 w_1을 계산한다. 이후 주어진 메시지와 w_1을 사용하여 해시 값 c를 계산하고 최종적으로 서명 $z=y+cs_1$을 생성하고 2회의 서명 검증 절차를 거친 뒤 서명을 패킹(packing)을 한다. 이때, 최종 서명 z를 계산하는 과정에서 해시 값 c와 개인 키 값 s_1 사이의 다항식 계수별 곱셈이 진행되며 공격자는 서명 값으로 해시 값 c를 알아낼 수 있고, 전력 분석 공격을 통해 해당 지점에서 s_1을 복구할 수 있다.

서명 검증 단계에서는 수신한 서명 값 z, c를 이용해 $Az-ct$를 연산한 후 반올림을 통해 w'_1을 계산한다. 이후 z의 범위($\leq \gamma_1-\beta$)를 확인하고 M과

Key Generation(pk, sk)
1. $A = R_q^{k\times t}$
2. $(s_1, s_2) = S_q^{\ t} \times S_q^{k}$
3. $t = A s_1 + s_2$
4. return($pk = (A,t)$, $sk = (A,t,s_1,s_2)$)

Sign(sk, M, σ)
1. $z = \perp$
2. while $z = \perp$ do {
3. $y = S^{l}_{\gamma 1 - 1}$
4. $w_1 = HighBits(Ay, 2\gamma_2)$
5. $c = H(M \| w1) \in B_\tau$
6. $z = y + c s_1$
7. if($(\|z\|_\infty \geq \gamma_1 - \beta)$ or $(\|LowBits(Ay - cs_2, 2\gamma_2)\|_\infty \geq \gamma_2 - \beta)$ then $z = \perp$
8. return($\sigma = (z, c)$)

Verify(pk, M, $\sigma = (z,c)$)
1. $w' = HighBits(Az - ct, 2\gamma_2)$
2. if($(\|z\|_\infty < \gamma_1 - \beta)$ and $(c = H(M \| w_1')))$ then accept signature

Fig. 1. Dilithium signature scheme

w'_1를 이용하여 해시 값 c'을 생성한 뒤 수신한 c와 계산한 c'을 비교하여 서명을 검증한다.

2.3 연속 웨이블릿 변환

과거에는 시간에 대한 신호를 주파수 영역으로 변환하기 위해 푸리에 변환(FT: Fourier Transform)을 주로 사용하였다. 푸리에 변환을 이용하면 시계열 신호에 대해 특정 주파수를 추출할 수 있지만, 특정 시점의 정보를 얻기는 어렵다. 이러한 단점은 시간을 분할한 다음 푸리에 변환을 적용하는 단시간 푸리에 변환(STFT: Short-Time Fourier Transform, 이하 STFT)이 개발됨에 따라 극복되었다. 하지만 STFT 역시 시간 해상도와 주파수 해상도 중 하나를 선택해야 하는 상충관계가 존재한다.

STFT의 한계점을 극복하기 위해 등장한 웨이블릿 변환(wavelet transform)은 시간에 따라 변하는 주파수를 가지는 소비 전력 파형과 같이 서로 다른 스케일에서 특징이 변하는 데이터를 다중 해상도의 시간-주파수 기반으로 분석하는 기법이다. 이러한 웨이블릿 변환은 STFT와 다르게 유한한 길이의 기저 함수(basis function)를 이용하여 기존의 시계열 데이터를 주파수 도메인으로 변환하며, 웨이블릿 변환의 기저 함수를 모 웨이블릿(mother wavelet)이라 한다. 연속 웨이블릿 변환 함수 W는 시간 t에 대한 입력 신호를 $x(t)$, 규모 매개변수를 s, 전이 매개변수를 τ 그리고 모 웨이블릿을 ψ 이라 할 때 수식 (1)과 같이 정의할 수 있다.

$$W_x(s,\tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t)\psi(\frac{t-\tau}{s})dt \qquad (1)$$

웨이블릿 변환은 규모 매개변수(scale parameter)가 이산적인지, 연속적인지에 따라 이산 웨이블릿 변환(DWT: Discrete Wavelet Transform, 이하 DWT)과 연속 웨이블릿 변환(CWT: Continuous Wavelet Transform, 이하 CWT)으로 나뉜다. DWT는 웨이블릿의 규모 매개변수를 정밀하지 않게 이산화하여 나타내며, 주로 사용되는 모 웨이블릿의 종류는 Haar, Daubechies, Gauss 등이 있다.

본 논문의 실험에서 사용한 CWT는 웨이블릿의 규모 매개변수를 더욱 미세하게 이산화하여 나타낸다. 이때 주파수를 단계별로 필터링하고 전력 파형을 압축하여 입력 데이터의 길이와 동일한 길이를 가진 데이터를 출력하는 DWT와 달리, CWT는 특정 스케일에 대응하는 주파수의 정보를 추출한 뒤 전력 파형을 분해하여 입력 데이터에서 차원이 1개 늘어난 데이터를 출력한다. 따라서 CWT는 스케일의 개수만큼 데이터 크기가 증가하여 분석 시간이 증가하는 단점이 존재하지만, 각 주파수에 대한 정보를 늘림으로써 전력 파형 분석 시 파형의 특징을 더욱 자세히 나타내어 공격 성능 향상을 노릴 수 있다. 주로 사용되는 모 웨이블릿의 종류는 Mexican hat, Morlet, Paul 등이 있으며, 본 논문에서는 Mexican hat을 사용하였다.

III. Dilithium 비프로파일링 기반 전력 분석 공격

3.1 실험 환경

본 실험의 공격 지점은 서명 생성 과정 중 NTT가 적용된 해시 값 c와 개인 키 값 s_1 간의 다항식 계수별 곱셈 연산이다. 격자 기반 암호에서 다항식은 256개의 계수로 표현되며, 모든 계수는 modulus q 상에서 존재하므로 하나의 계수는 8,380,417 ($2^{23}-2^{13}+1$) 미만의 값으로 23비트 값이다. 해당 계수들은 같은 차수의 계수들과 곱해지는 다항식 계수별 곱셈 연산을 진행하는데, 이때 곱셈 결과를 부호 있는 정수(signed int) 형태로 저장하며 그 범위는 $(-\frac{q-1}{2} \sim \frac{q-1}{2})$이다. 이는 최소한의 modulus 감산을 위한 형태로서 modulus 연산을 최소한으로 진행하는 lazy reduction이 가능하다.

본 논문에서는 CRYSTALS-Dilithium에 대한 전력 분석 공격 실험을 위해서 NIST에서 제공하는 공식 소스 코드인 Dilithium2 reference 버전을 ARM-Cortex-M4 코어가 탑재된 32비트 프로세서인 STM32F3 MCU 상에서 구현하여 사용하였다. 개인 키는 칩 내부에 내장되어 있다고 가정하였으며, 프로세서가 암호 연산을 수행할 때 소비되는 전력 파형은 ChipWhisperer Lite를 통해 29.5MS/s 속도로 측정하였다.

3.2 Dilithium 전력 분석 공격 지점

본 논문에서는 총 2가지의 전력 분석 공격을 진행하였다. Basic 전력 분석 공격은 전체 추측 키 범위

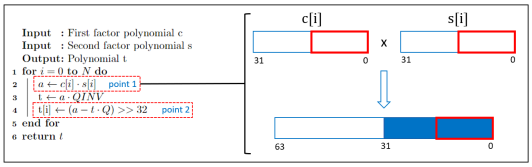


Fig. 2. Polynomial coeffcient-wise multiplication in Dilithium signature

를 이용하여 한 번에 개인 키를 복구하는 공격으로, Fig. 2의 point2를 공격 지점으로 설정한다. 이는 한 번의 전력 분석 공격으로 개인 키를 복구할 수 있지만, Dilithium의 큰 modulus q로 인해 많은 연산량을 가진다는 단점이 존재한다.

상기한 문제를 해결하기 위해 큰 사이즈의 비밀 키 일부 비트를 먼저 복구한 뒤 나머지 비트를 복구하는 공격 방법이 제안되었으며[17], 2021년 Z. Chen 등은 Dilithium 개인 키 계수를 대상으로 실험을 진행하였다[18]. 이 실험에서는 총 10,000개의 전력 파형을 이용하여 ARM-CortexM4 코어에서 CPA만을 이용하여 개인 키를 복구할 수 있음을 보였다.

본 논문에서 진행하는 Fast 전력 분석 공격도 Z. Chen 등의 실험에서와 유사하게 총 두 번의 전력 분석을 진행한다. 1단계에서는 Fig. 2의 point1을 공격 지점으로 설정하여 개인 키 계수의 하위 16비트(LSP: Least Significant Part, 이하 LSP)를 먼저 복구해 LSP 후보를 특정한다. 2단계에서는 point2를 공격 지점으로 설정하고, 1단계에서 특정한 LSP를 통해 개인 키 계수의 상위 7비트(MSP: Most Significant Part, 이하 MSP)를 예측하여 최종적으로 전체 개인 키 계수 23비트를 복구하는 공격이다.

IV. Dilithium 전력 분석 공격 실험

본 논문에서는 비프로파일링 공격인 CPA 공격과 딥러닝 기반의 DDLA 공격을 수행하여 Dilithium에 대한 부채널 공격 취약점이 있음을 검증한다.

4.1 Dilithium 대상 CPA 공격

4.1.1 Basic CPA 공격

Dilithium의 개인 키 계수의 범위는 $[0 \sim q-1]$이다. 이때, 상기한 lazy reduction의 특성으로 인해 개인 키 계수의 추측 범위를 $[0 \sim \frac{q-1}{2}]$로 줄일 수 있다. 추측 개인 키 계수가 $\frac{q-1}{2}$일 때 다항식 계수별 곱셈 연산 결과 값이 rt라면, $\frac{q+1}{2}$일 때 결과 값은 $-rt$로 절댓값은 같고 부호만 다른 값이 도출된다. 이때 rt가 양의 상관 계수를 갖는다면 $-rt$는 rt의 상관 계수와 대칭되는 음의 상관 계수를 갖는다. 따라서 개인 키 계수의 추측 범위를 줄인 뒤 CPA를 진행해 상관 계수를 확인하였을 때 높은 양의 상관 계수가 관찰되면 해당 키를 올바른 추측 키로, 높은 음의 상관 계수가 관찰되면 해당 키를 올바른 추측 키의 대응 키 ck로 특정한다. 대응 키가 관찰되었을 때에는 $q-ck$를 계산하여 실제 키를 특정한다. Fig. 3은 $\frac{q-1}{2}=4{,}190{,}208$를 추측 키 범위로 하여 진행한 Basic CPA 공격 결과이다.

실험에 사용된 개인 키 계수는 0x0079D76B (7,985,003)이며, 다항식 계수별 곱셈 연산 지점 파형 100개를 사용하였다. CPA 공격 실험 결과, 오직 하나의 계수에서 PCC(Pearson Correlation Coefficoent)의 절댓값 $|PCC| > 0.8$을 만족하는 음의 상관 계수가 측정되었으며 해당 계수 값은 0x60896(395,414)이다. 음의 상관 계수가 측정되었으므로 $q-ck$를 계산하면 올바른 개인 키 계수인 7,985,003가 계산되는 것을 확인할 수 있으므로 개인 키 계수 복구에 성공하였다.

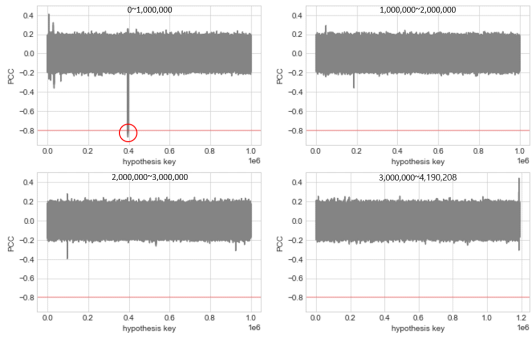


Fig. 3. Experimental results of Basic CPA

4.1.2 Fast CPA 공격

개인 키 계수의 LSP를 먼저 복구하기 위해 Fig. 2의 point1 지점을 공격한다. 해당 지점은 단순 곱셈 연산으로 모든 추측 키를 이용하여 공격을 진행하면 비트 시프트로 인한 유사한 HW 존재 및 상위 비트의 노이즈로 인해 오탐이 존재한다. 따라서 해당 공격을 진행하여 개인 키 계수의 LSP 후보를 특정한다. Fig. 4는 point1을 공격 지점으로 전력 파형 10,000개를 이용하여 진행한 1단계 CPA 공격 결과이다. $PCC > 0.3$을 만족하는 12개의 계수가 관찰되어 이를 후보 계수로 선정하였다. 논문에서는 실험을 통하여 12개의 후보 키를 선택하는데 필요한 상관 계수가 0.3정도임을 확인하였다. 이때 $PCC < -0.3$인 계수들이 존재하는데 이는 모두 상위 비트 노이즈로 인한 오탐 값이므로 후보 계수 선정에서 제외하였다.

이어서 획득한 12개의 LSP 후보를 이용해 MSP를 예측하고 최종적으로 개인 키 계수를 복구하는 2단계 CPA 공격을 진행하였다. Fig. 5는 point2를 공격 지점으로 전력 파형 100개를 이용하여 진행한 2단계 CPA 공격 결과이다. 공격 결과 오직 하나의 계수에서 $|PCC| > 0.8$을 만족하는 양의 상관 계수가 측정되었으며 해당 계수 값은 올바른 개인 키 계수인 0x0079D76B(7,985,003)로 복구에 성공하였다.

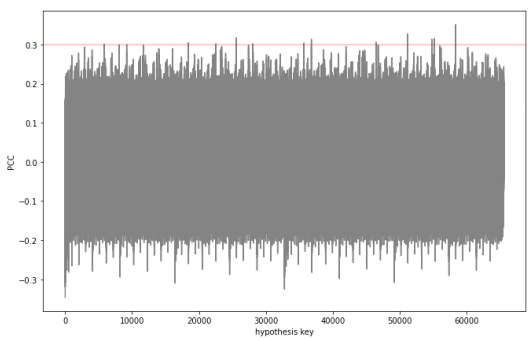


Fig. 4. Fast CPA result on LSP

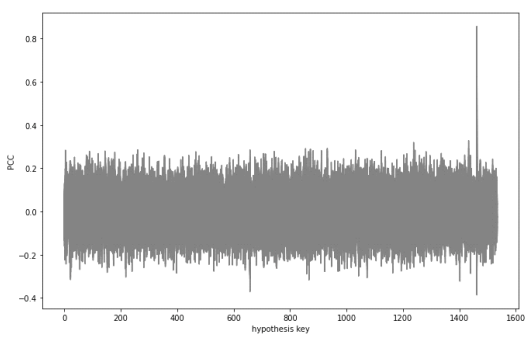


Fig. 5. Fast CPA result on MSP

4.2 Dilithium 대상 DDLA 공격

다음으로 DDLA 기법을 이용하여 공격을 수행하였는데 매우 큰 추측 키 범위로 인해 한 번에 개인 키를 복구하는 것은 불가능하며, Fast CPA와 같이 총 두 번의 DDLA 공격을 통해 개인 키의 하위 16비트를 먼저 찾아내고 이후 전체 개인 키를 복구한다. 또한, 전력 파형에 각 특성의 평균을 0, 분산을 1로 변경하는 StandardScaler 전처리 과정과 상기한 CWT를 적용하여 공격 성능 개선 여부를 확인한다.

공격에 사용된 딥러닝 모델은 다층 퍼셉트론(MLP)을 사용하였으며, StandardScaler 전처리와 CWT가 적용된 파형을 입력으로 사용할 때에는 이미지 처리에 효과적인 합성곱 신경망(CNN)을 사용하였다. LSP 복구 공격에서는 10,000개, MSP 복구 공격에서는 500개의 전력 파형을 입력으로 사용하였으며, 라벨은 추측 키로부터 생성한 HW, MSB, LSB, HW-based binary를 각각 사용하여 공격을 시도하였다. 또한, 딥러닝 모델의 과적합 문제를 회피하고 추측 키의 학습률을 일반화 성능으로 측정하기 위해 validation loss를 학습 경향성의 지표로 사용한다.

MSP 복구 공격의 성능은 평가 기법인 NMM(Normalized Maximum Margin)[19]을 이용해 평가한다. NMM은 올바른 키와 잘못된 키의 학습 평가 지표 값의 차이를 표준편차 σ 단위로 나타낸 것이다. 만약 올바른 개인 키 계수의 NMM이 0보다 크다면 올바른 키를 복구할 수 있음을, 0보다 작다면 올바른 키를 복구할 수 없음을 의미한다.

4.2.1 LSP 대상 Fast DDLA 공격

개인 키 계수의 LSP를 먼저 복구하는 DDLA 공격을 진행한다. StandardScaler 전처리가 적용된 전력 파형 10000개를 MLP 모델의 입력으로 하고, 중간 값의 MSB를 라벨로 설정하여 공격을 진행하였다. 공격 결과는 Fig. 6과 같으며, 추측 키 중 학

Table 1. LSP candidate coefficient selection

Valid candidate coefficients								
0x12a0	0x2895	0x44a8	0x6ed6	0x912a	0xb580	0xbb58	0xc7da	**0xd76b**
High-bit noise (false positives)								
0x1000	0x4000	0x5000	0x5800	0x6000	0x8000	0xa000	0xb000	0xc000	0xe000	0xf000

습이 이루어져 상대적으로 낮은 validaion_loss를 가지는 키들이 구분되어 임계치를 설정하고 후보 계수를 특정하였다. 이때 Table. 1과 같이 유효 후보 계수 뿐만 아니라 상위 비트 노이즈에 의한 오탐 값도 후보 계수에 포함되어 이러한 오탐 값은 후보 계수에서 제외하였고, 9개의 후보 계수를 선정하였다.

또한, Table. 2과 같이 모든 라벨을 이용하여 진행한 LSP 복구 결과 MSB 이외의 다른 라벨을 이용하였을 경우 추측 키 사이에 학습 경향성의 차이가 드러나지 않아 후보 계수 특정에 실패하였다.

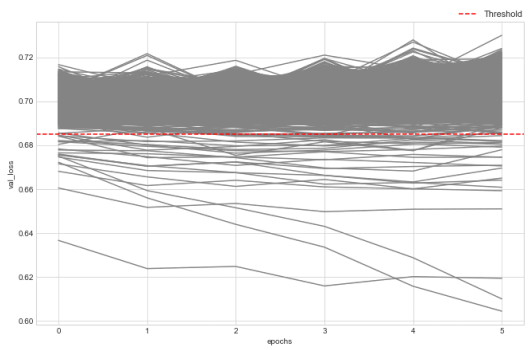


Fig. 6. Fast DDLA result on LSP using MSB label

Table 2. Fast DDLA result on LSP using all labels

	HW	MSB	LSB	HW_B
Threshold (val_loss)	-	<0.685	-	-
Number of candidate coefficient	Unable	9	Unable	Unable

4.2.2 MSP 대상 Fast DDLA 공격

이어서 개인 키 계수의 MSP를 복구하는 DDLA 공격을 진행한다. 이때, 상기한 LSP 복구 실험에서 유일하게 후보 계수를 선정할 수 있었던 MSB 라벨을 이용하여 특정한 후보 계수를 이용해 MSP를 복구한다. 이때 파형 전처리 과정인 StandardScaler 표준화 및 CWT 적용 여부에 따른 성능을 비교한다. 이때 추측 키는 $9\times2^7=1{,}152$로 설정하였으며 파형 전처리와 각 라벨에 따른 출력층 노드 개수를 제외한 모든 딥러닝 모델 파라미터는 동일하게 설정하였다.

먼저 파형 전처리가 적용되지 않은 소비 전력 파형을 이용하여 개인 키 계수의 MSP를 복구하는 Fast DDLA 공격을 진행하였다. Fig. 7은 전처리가 적용되지 않은 소비 전력 파형 500개를 이용하여 개인 키의 MSP 복구를 시도하는 Fast DDLA 공격을 진행한 뒤 이를 NMM으로 평가한 것이다. 공격 결과 모든 라벨에서 올바른 개인 키 계수와 잘못된 개인 키 계수의 학습 경향성이 구분되지 않아 올바른 개인 키 계수의 NMM 값이 모두 0보다 작으므로 공격에 실패하였다.

이어서 StandardScaler를 이용해 전처리한 파형을 이용하여 공격을 진행하였다. Fig. 8의 (a)는 StandardScaler가 적용된 소비 전력 파형 500개를 이용하여 Fast DDLA 공격을 진행한 뒤 이를 NMM으로 평가한 것이다. 공격 결과 MSB, HW-based binary 라벨의 경우 학습 시작 직후 올바른 키의 NMM 값이 0보다 큰 값을 가져 곧바

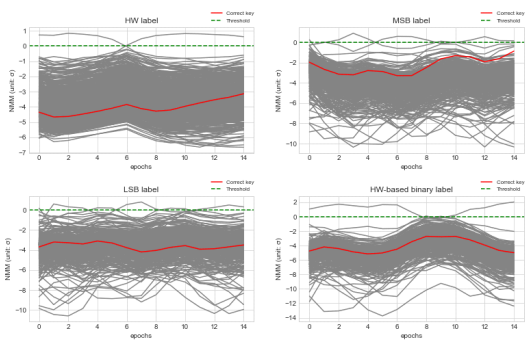


Fig. 7. Fast DDLA result on MSP using plain traces without preprocessing

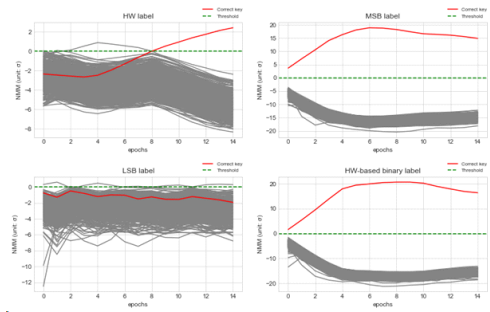
로 구분되므로 성능이 뛰어난 것을 확인하였고, HW 라벨의 경우 8 에포크(epoch)에서부터 올바른 키가 구분되었으며, LSB 라벨은 올바른 키 구분에 실패한 것을 확인하였다.

StandardScaler가 적용된 파형을 25개부터 500개까지 늘리며 학습한 모델의 NMM 평가 결과는 Fig. 8의 (b)와 같다. 실험 결과 MSB, HW-based binary 라벨의 경우 50개의 파형을 이용하였을 때부터, HW 라벨의 경우 450개의 파형을 이용하였을 때부터 개인 키 계수의 MSP를 복구하는 공격이 모두 성공함을 확인하였다. LSB 라벨은 모든 구간에서 올바른 키 구분에 실패하였다.

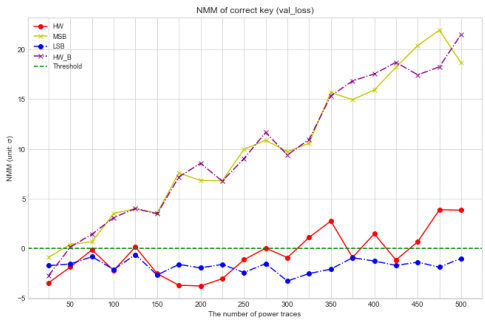
마지막으로 StandardScaler 전처리와 CWT를 모두 적용한 파형을 이용하여 공격을 진행하였다. 이때 파형은 시간-주파수 도메인의 2차원 데이터이며, 이러한 형태의 파형을 CNN의 입력으로 사용하였다. Fig. 9의 (a)는 StandardScaler와 CWT가 모두 적용된 소비 전력 파형 500개를 이용하여 Fast DDLA 공격을 진행한 뒤 이를 NMM으로 평가한 것이다. 공격 결과 MSB, HW-based binary 라벨의 경우 이전 공격 결과와 유사하게 학습 시작 직후 올바른 키의 NMM 값이 0보다 큰 값을 가져 곧바로 구분되는 것을 확인하였다. 이때 HW 라벨의 경우 1 에포크에서부터 올바른 키가 구분되었으며 이전 공격 결과와 비교해 성능이 향상된 것을 확인하였다. 단, LSB 라벨은 이 경우에도 올바른 키 구분에 실패한 것을 확인하였다.

StandardScaler와 CWT가 모두 적용된 파형을 25개부터 500개까지 늘리며 학습한 모델의 NMM 평가 결과는 Fig. 9의 (b)와 같다. 실험 결과 MSB, HW-based binary 라벨은 이전 공격과 동일하게 50개의 파형을 이용하였을 때부터 공격에 성공하였다. 이때 HW 라벨의 경우 225개의 파형을 이용하였을 때부터 공격이 모두 성공하는 것을 확인하여 지난 공격과 비교해 공격에 필요한 최소 파형 수가 줄어들어 공격 성능이 개선된 것을 확인하였다.

Table. 3은 Dilithium 다항식 계수별 곱셈 연산의 파형 전처리를 달리 한 3번의 Fast DDLA 공

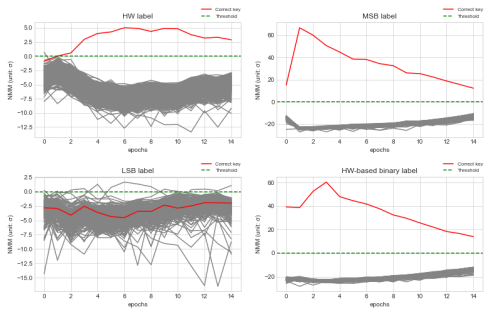


(a) NMM calculated by epochs using standardized traces

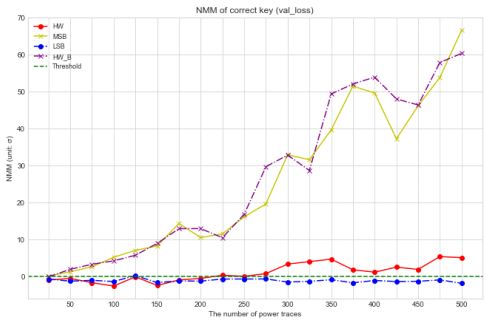


(b) NMM calculated by number of traces using standardized traces

Fig. 8. NMM calculated in Fast DDLA result on MSP using standardized traces



(a) NMM calculated by epochs using standardized and CWT applied traces



(b) NMM calculated by number of traces using standardized and CWT applied traces

Fig. 9. NMM calculated in Fast DDLA result on MSP using standardized and CWT applied traces

Table 3. NMM calculated in Dilithium Fast DDLA attack result on correct key coefficient

	Plain	Standard Scaler	StandardScaler + CWT
HW	-3.13	2.41	5.02
MSB	-0.86	18.88	**66.64**
LSB	-3.10	-0.49	-1.92
HW_B	-2.71	20.67	60.35

격에서 측정한 올바른 개인 키 계수의 NMM 값을 정리한 것이다. 표에서 확인할 수 있듯이, 이러한 공격이 성공하기 위해선 파형 전처리가 필수적임을 확인하였다. 특히 StandardScaler와 CWT를 파형에 모두 적용한 경우 StandardScaler만을 적용한 경우와 비교해 공격 성능이 약 3배 증가한 것을 확인하여 DDLA 공격 시 연속 웨이블릿 변환 기법을 사용하면 성능 향상을 기대해 볼 수 있다.

V. 결 론

양자 컴퓨터의 등장으로 기존의 수학적 난제를 기반한 알고리즘들을 다항식 시간 안에 연산할 수 있게 되었다. 따라서, NIST에서는 이러한 양자 컴퓨터에 대응하고자 양자 내성 암호 표준화 작업을 진행하고 있으며 최근에 표준화가 확정된 알고리즘을 발표하였다. CRYSTALS-Dilithium은 전자서명 분야에서 표준화가 확정된 Module-LWE 문제에 기반한 격자 기반의 양자 내성 암호이다. 하지만 CRYSTALS-Dilithium의 서명 생성 단계에서 부채널 누출이 존재한다.

본 논문에서는 CRYSTALS-Dilithium에 대해 비프로파일링 기반 전력 분석 공격의 일종인 CPA 공격과 DDLA 공격 실험을 진행하였으며, 두 공격 모두 Dilithium의 개인 키 계수를 복구하는 데 성공하였다. 특히, DDLA의 경우 파형 전처리 여부에 따라 공격 결과가 크게 달라지는 것을 확인하였다. StandardScaler와 CWT를 파형에 모두 적용할 경우 StandardScaler만을 적용했을 경우와 비교해 약 3배의 성능 향상이 관찰되었으며, 그중에서도 MSB 라벨을 사용할 경우 66.64의 높은 NMM 값을 가져 높은 성능으로 개인 키 계수를 복구할 수 있음을 보였다. 따라서 딥러닝을 이용한 전력 분석 공격에서 파형 전처리의 중요도를 확인하였다. 결론적으로 본 논문을 통해 Dilithium 서명 생성 단계에서 부채널 정보 누출이 존재하여 개인 키 계수를 복구할 수 있다는 것을 확인하였으므로 향후 이에 대한 대응책이 마련되어야 할 것이다.

References

[1] P. Shor, "Polynomial time algorithms for discrete logarithms and factoring on a quantum computer," SIAM Journal on Computing, Vol. 26, Issue 5, pp. 1484-1509, 1997.

[2] L. Grover, "A fast quantum mechanical algorithm for database search," ACM symposium on Theory of Computing(STOC'96), pp. 212-219, July. 1996.

[3] D. Moody, G. Alagic, D. Cooper, Q. Dang, T. Dang, J. Kelsey, J. Lichtinger, Y. Liu, C. Miller, R. Peralta, R. Perlner, A. Robinson, D. Smith-Tone and D. Apon, "Status Report on the Third Round of the NIST Post-Quantum Cryptography Standardization Process," National Institute of Standards and Technology, July. 2022.

[4] P. Kocher, "Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems," CRYPTO'96, LNCS 1109, pp. 104-113, Aug. 1996.

[5] D. Genkin, I. Pipman and E. Tromer, "Get your hands off my laptop: physical side-channel key-extraction attacks on PCs," Journal of Cryptographic Engineering, Vol. 5, Issue 2, pp. 95-112, May. 2015.

[6] D. Genkin, A. Shamir and E. Tromer, "RSA key extraction via low-bandwidth acoustic cryptanalysis," CRYPTO'14, LNCS 8616, pp. 444-461, 2014.

[7] S. Chari, J. Rao and P. Rohatgi, "Template attacks," CHES'02, LNCS

2523, pp. 13 - 28, 2003.
[8] G. Hospodar, B. Gierlichs, E. Mulder, I. Verbaushede and J. Vandewalle, "Machine learning in side-channel analysis: a first study," Journal of Cryptographic Engineering, Vol. 1, No. 4, pp. 293-302, 2011.
[9] L. Lerman, R. Poussier, G. Bontempi, O. Markowitch and F. Standaert, "Template attacks versus machine learning revisited and the curse of dimensionality in side-channel analysis," COSADE'15, LNCS 9064, pp. 20 - 33, 2015.
[10] L. Lerman, G. Bontempi and O. Markowitch, "A machine learning approach against a masked AES," Journal of Cryptographic Engineering, Vol. 5, pp. 123-139, 2015.
[11] P. Kocher, J. Jaffe and B. Jun, "Differential power analysis," Advances in Cryptology, CRYPTO' 99, LNCS 1666, pp. 388-397, 1999.
[12] E. Brier, C. Clavier and F. Olivier, "Correlation power analysis with a leakage model," CHES'04, LNCS 3156, pp. 16-29, 2004.
[13] B. Timon, "Non-profiled deep learning-based side-channel attacks with sensitivity analysis," IACR Transactions on Cryptographic Hardware and Embedded Systems, Vol. 2019, Issue 2, pp. 107-131, 2019.
[14] D. Bae, J. Hwang, H. Lee and J. Ha, "Non-profiling deep learning side-channel attack with Hamming weight-based binary labels", Conference on Information Security and Cryptography(CISC-W'20), pp. 2020.
[15] L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, P. Schwabe, G. Seiler and D. Stehlé, "CRYSTALS-Dilithium: A Lattice-Based Digital Signature Scheme," IACR Transactions on Cryptographic Hardware and Embedded Systems, Vol. 2018, Issue 1, pp. 238-268, Feb. 2018.
[16] P. A. Fouque, J. Hoffstein, P. Kirchner, V. Lyubashevsky, T. Pornin, T. Prest, T. Ricosset, G. Seiler, W. Whyte and Z. Zhang, "Falcon: Fast-Fourier Lattice-based Compact Signatures over NTRU, Specification v1.2," NIST Post-Quantum Cryptography Standardization Round 3, 2020.
[17] M. Tunstall, N. Hanley, R. McEvoy, C. Whelan, C. Murphy and W. Marnane, "Correlation Power Analysis of Large Word Sizes," IET Irish Signals and Systems Conference (ISSC), pp. 145-150, 2007.
[18] Z. Chen, E. Karabulut, A. Aysu, Y. Ma and J. Jing, "An Efficient Non-Profiled Side-Channel Attack on the CRYTSTALS-Dilithium Post-Quantum Signature," 2021 IEEE 39th International Conference on Computer Design (ICCD), pp. 583-590, 2021.
[19] D. Bae and J. Ha, "Performance Metric for Differential Deep Learning Analysis," Journal of Internet Services and Information Security (JISIS), 11(2), pp. 22-33, 2021.

〈저자소개〉

장 세 창 (Sechang Jang) 학생회원
2022년 2월: 호서대학교 정보보호학과 학사
2022년 3월~현재: 호서대학교 정보보호학과 석사과정
〈관심분야〉 부채널 공격, 정보보호, 인공지능 보안

이 민 종 (Minjong Lee) 학생회원
2018년 3월~현재: 호서대학교 컴퓨터공학부 학부과정
〈관심분야〉 인공지능 보안, 부채널 공격, 네트워크 보안

강 효 주 (Hyoju Kang) 학생회원
2020년 3월~현재: 호서대학교 컴퓨터공학부 학부과정
〈관심분야〉 인공지능 보안, 부채널 공격, 네트워크 보안

하 재 철 (Jaecheol Ha) 종신회원
1989년 2월: 경북대학교 전자공학과 학사
1993년 8월: 경북대학교 전자공학과 석사
1998년 2월: 경북대학교 전자공학과 박사
1998년 3월~2007년 2월: 나사렛대학교 정보통신학과 교수
2007년 3월~현재: 호서대학교 컴퓨터공학부 교수
2009년 1월~현재: 한국산학기술학회 이사
2013년 1월~현재: 한국정보보호학회 수석부회장
2023년 1월~현재 :국제차세대융합기술학회 부회장
〈관심분야〉 암호학, 부채널 공격, 네트워크 보안, 정보보호