정 보 보 호 학 회 지
제 30 권 제1호, 2020. 2

딥러닝을 이용한 부채널 분석 기술 연구 동향

진 성 현*, 김 희 석**

요 약

딥러닝 기술의 발달로 인해 다양한 응용 분야에서 해당 기술 활용 시 좋은 성능을 보임에 따라 부채널 분석 분야에서도 딥러닝 기술을 적용하는 연구들이 활발히 진행되고 있다. 초기 딥러닝 기술은 데이터 분류 문제를 해결해야 하는 템플릿 공격과 같은 프로파일링 기반의 부채널 공격에 집중되어 적용되었지만 최근에는 프로파일링 기반의 부채널 분석 뿐만 아니라 상관 전력 분석 등과 같은 논프로파일링 기반 부채널 공격, 파형 인코딩 및 전처리, 부채널 누출신호 탐색 등으로 연구범위가 확대되어지고 있다. 본 논문에서는 딥러닝을 이용한 부채널 분석 기술의 최신 연구 동향을 분야별로 체계적으로 정리 및 분석하고자 한다.

Ⅰ. 서 론

센서 같은 소형 장비를 넘어 모든 장비들이 연결되는 사물인터넷 시대가 도래함에 따라 IoT 기술을 이용한 새로운 서비스 등을 통해 사용자들은 다양한 편의성을 제공받고 있지만, 편의성과 별개로 개인 프라이버시 등의 관점에서 많은 보안 취약점들이 발생하고 있다. 부채널 분석은 IoT 환경에서 보안을 취약하게 만들 수 있는 대표적인 기술로서 최근 다양한 평가 기준에서 부채널 분석에 대한 안전성을 필수적으로 요구하고 있으며, 그러한 요구에 따라 부채널 분석에 대한 안전성과 연관된 연구들이 활발히 진행되고 있다.

딥러닝은 머신러닝의 한 종류로 20세기 중반부터 연구가 시작된 기술이다. 하지만 컴퓨팅 환경의 부재, 데이터 부족 등의 이유로 다른 머신러닝보다 좋은 성능을 내지 못하고 있었으나 21세기 초부터 딥러닝 알고리즘의 발달과 빅데이터 시대, 그리고 클라우드 등 컴퓨팅 파워의 발달과 함께 딥러닝 기술의 성능이 급속히 증가하여 영상 처리, 자연어 처리, 기계번역, 질병 진단 등 다양한 분야에서 좋은 성능을 보이고 있다[1].

딥러닝 기술이 다양한 분야에서 성공적인 성과를 거두고 있고, 딥러닝 기술과 프로파일링 기반 부채널 분석 기술이 데이터를 분류하는 태스크로써 같은 철학을 공유함에 따라 자연스럽게 딥러닝을 부채널 분석에 적용한 연구들이 진행되었으며 안전성 평가의 객관화 가능성과 딥러닝 기반 부채널 분석의 성능을 입증한 다양한 연구 결과들이 최근들어 발표되고 있다. 평가자 관점에서 딥러닝을 부채널 분석에 이용할 경우 기존 부채널 분석 과정에서 필수적인 단계인 파형 전처리, 정렬 그리고 유의미한 시점(Points of Interest, POIs) 선택 등을 딥러닝의 특징을 이용하여 생략할 수 있다. 이는 부채널 분석에 대한 안전성 평가에서 평가자의 역량을 배제하고 객관적인 안전성 평가를 할 수 있다는 가능성을 암시한다. 또한, 딥러닝을 이용한 부채널 분석의 성능이 기존 부채널 분석보다 좋을 수 있음이 확인되었다. 이는 딥러닝을 이용한 부채널 분석이 더 강력한 공격이 될 수 있으며 기존 부채널 분석에서 사용하지 못한 정보를 추가로 이용한 것으로 보여질 수 있다. 이러한 흐름에 따라 딥러닝을 이용한 부채널 분석에 대한 연구가 활성화되어 활발히 진행되고 있는 추세이다.

본 논문에서는 기반 지식으로 딥러닝과 부채널분석을 간략히 소개한 후에 딥러닝을 이용한 부채널 분석을 용도에 따라 체계적으로 분류하여 정리하고 분석하였다. 이를 통해 현재까지 연구 흐름을 확인할 수 있으며 향후 연구 방향에 대한 도움이 될 것으로 기대된다.

이 성과는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구 재단의 지원을 받아 수행된 연구임(No.NRF-2019R1A2C2088960).
* 고려대학교 정보보호대학원 정보보호학과, 고려대학교 정보보호연구원 (대학원생, sunghyunjin@korea.ac.kr)
** 고려대학교 사이버보안학과 (교수, 80khs@korea.ac.kr)

Ⅱ. 딥러닝

딥러닝은 신경망(뉴럴 네트워크, Neural Network) 알고리즘에 기반한 머신러닝의 한 종류이다. 신경망은 순차적인 데이터 표현을 추상화하는 계산 모델이다[1]. Universal approximation theorem에 의해 특정 조건을 만족하는 활성화 함수를 갖춘 하나 이상의 은닉층으로 구성된 신경망은 임의의 Borel measurable functions f를 근사 가능하며 각 층의 순차적인 연산을 통해 데이터 표현을 추상화한다는 성질이 알려져 있다[1,2]. 이러한 성질에 따라 신경망을 식 (1)과 같은 함수 $\hat{f}(\bullet)$로 표현할 수 있으며 함수 $\hat{f}$를 특정 함수 f로 근사하는 과정인 학습 과정을 식 (2)와 같이 표현 가능하다.

$$\hat{f} = O \circ A_n \circ \lambda_n \circ \cdots \circ A_1 \circ \lambda_1 \circ I \quad (1)$$

$$\hat{\theta} = \arg\min_{\theta} L(\hat{f}(X;\theta), Y = f(X)) \quad (2)$$

식 (1)에서 I, O는 각각 입력과 출력에 대한 어떠한 함수이다. 일반적으로 I는 항등함수가 사용되며 O는 항등함수 혹은 소프트맥스(softmax) 함수가 사용된다. λ_i는 선형 함수를 의미하며 A_i는 활성화 함수(activation function)로 sigmoid, ReLU 등의 비선형 함수가 사용된다. 식 (2)의 L은 손실함수, (X, Y)는 각각 데이터와 라벨을 의미한다. 신경망은 데이터를 입력으로 할때 함수 $\hat{f}$의 출력이 의도한 결과인 라벨이 출력되도록 λ_i의 가중치(weight)를 조정한다. 이러한 과정을 학습이라 한다. 학습을 통해 함수 $\hat{f}$를 함수 f로 근사시킬 수 있게 된다.

손실함수가 감소하도록 학습하기 위해 가중치를 조정하는 방법으로 경사하강법(gradient descent method)을 이용한다. 손실함수에 대한 그래디언트를 이용하여 손실함수가 최소가 되는 방향을 알아낸 후 가중치를 조정하는 방법이다. 그래디언트를 조정하는 정도를 학습률로 조절 가능하며 가중치를 변경하는 방식인 학습 방법으로 Momentum[3,4], RMSProp[5], Adam[6] 등이 있다.

학습으로 인해 변경되는 가중치를 제외한 신경망의 구조, 학습 방법 등과 같이 사용자가 선택 가능한 모든 것을 하이퍼 파라미터라 하며 이를 적절하게 설정하여야만 신경망이 의도한 태스크를 처리하도록 특정 함수를 근사시킬 수 있다. 그러나 하이퍼 파라미터를 쉽게 알 수 있는 방법은 없으며 이를 머신러닝 분야에서는 “공짜 점심은 없다(There is no free lunch)”라고 표현한다[7].

신경망을 학습하기 위해 사용되는 데이터셋은 각각 훈련, 검증, 테스트 데이터셋으로 나뉘어진다. 각 데이터셋은 서로 중복되어서는 안되며 각각 데이터셋마다 고유의 특성이 있다. 학습 과정에서 훈련 데이터셋을 이용하여 가중치를 조정하며 학습이 제대로 되었는지를 검증하기 위해 검증 데이터셋을 이용한다. 이 때, 검증 데이터셋에 대해서는 가중치를 조절하지 않는다. 테스트 데이터셋은 실제 학습된 신경망을 적용하기 위한 데이터셋으로 실제 목표를 위해 사용되는 데이터이다.

학습 데이터셋과 검증 데이터셋이 너무 작을 경우 신경망이 편향되어 학습될 가능성이 있다. 이러한 경우를 피하기 위해 교차검증(K-fold cross validation)이란 기법이 이용된다[2]. 또한 학습 과정에서 신경망의 구조나 학습량 등으로 인해 신경망이 테스트 데이터셋에 대한 실제 성능보다 훈련 데이터에 특화되어 학습될 경우가 있다. 이러한 현상을 오버피팅(Overfitting)이라 한다. 이를 방지하기 위해 weight decay[8,9], dropout[10], batch nomalization[11] 등의 기법을 이용하기도 한다.

Ⅲ. 부채널 분석

증명 가능한 안전성이나 계산적 안전성을 보장 받는 암호 시스템이더라도 하드웨어에서 실제로 운용될 때 의도치 않은 정보가 발생한다. 이러한 의도치 않은 정보는 하드웨어에서 동작하는 연산과 처리되는 데이터에 따라 발생하는데 전력 소모량, 연산 소요시간, 전자파 방출 등이 대표적인 예이며 이러한 것을 부채널 누출이라 하고 부채널 누출을 이용하여 비밀정보를 분석하는 기법을 부채널 분석이라 한다[12]. 1996년 P. Kocher가 부채널 분석에 대한 개념을 도입한 이후로 분할 정복이 가능하다는 특징으로 인해 실질적이며 강력한 공격임이 잘 알려져 있어 현재까지 부채널 분석에 대한 연구가 활발히 수행되고 있다.

부채널 분석은 공격자 환경에 따라 논프로파일링과

프로파일링 공격으로 구분된다. 부채널 분석의 두 가지 공격유형 모두에서 파형 정렬, 잡음 제거, 유의미한 시점 선택 등의 과정을 공격자가 직접 수행해야만 한다는 특징이 있다.

본 장의 나머지에서는 논프로파일링 공격과 프로파일링 공격에 대해 소개하고 부채널 대응기법 및 관련된 이슈에 대해 소개한다. 이후 내용 서술시 부채널 누출과 부채널 정보, 부채널 신호, 파형을 혼용하여 사용한다.

3.1. 논프로파일링 공격

공격자가 공격 대상 장비로부터만 부채널 정보를 얻는 환경을 논프로파일링 환경이라 하며 논프로파일링 공격은 이러한 환경하에 수행되는 공격이다. 논프로파일링 공격은 다시 두 가지 유형으로 분류될 수 있다.

첫 번째 유형은 단일 혹은 소수의 파형만을 가지고 비밀 정보를 추론하는 분석하는 기법이며 대표적으로 단순전력분석(simple power analysis, SPA)이 있다[13].

두 번째 유형은 다량의 파형을 가지고 통계적 분석을 통해 비밀 정보를 알아내는 방법이다. 대표적인 방법으로는 차분전력분석(differential power analysis, DPA)[13], 상관전력분석(correlation power analysis, CPA)[14], 상호 정보량 분석(mutual information analysis, MIA)[15] 등이 있다. 이러한 유형의 공격에서는 키를 추측하여 중간값을 계산하고 그에 대한 전력모델값을 실제 전력과 통계량을 비교함으로써 키를 찾아낸다. 이로 인해 통계량을 계산하기 위해 시점 일치, 즉 파형 정렬이 되어져 있어야 한다[12,16,17]. 또한, 해밍웨이트나 해밍디스턴스 등의 전력 모델과 같이 연산 혹은 처리되는 데이터가 부채널 정보를 어떤 방식으로 결정하는지에 대한 모델링이 성능에 큰 영향을 미친다.

3.2. 프로파일링 공격

프로파일링 공격은 공격 대상 장비와 동일하며 프로그래밍 가능한 프로파일링 장비가 주어진 공격자 환경에서의 공격이며 이러한 환경을 프로파일링 환경이라 한다. 공격자는 동일한 장비를 이용하여 암호시스템이 동작할 때 연산과 데이터에 따라 발생하는 부채널 정보를 특징화할 수 있다. 공격자는 실제 공격 대상 장비에서 암호 시스템에 대한 부채널 분석시 특징화한 정보를 이용하여 비밀 정보를 알아낼 수 있다. 프로파일링 공격으로는 템플릿공격(template attack)[18]과 stochastic model[19] 등이 있다.

대표적인 프로파일링 공격인 템플릿 공격은 다음과 같은 절차로 수행된다. 프로파일링 장비로부터 유의미한 시점에 해당하는 정보를 추출한다. 실제 유의미한 시점들에서 각 중간값의 부채널 정보는 특정 다변수 분포를 형성한다. 템플릿 공격에서는 유의미한 시점들에 대한 분포를 정규분포로 가정함으로써 다변수 정규분포의 통계량인 평균과 공분산 행렬을 추정한다. 평균과 공분산 행렬을 템플릿이라 한다. 실제 공격 대상 장비를 분석할 때 같은 유의미한 시점으로부터 얻은 부채널 정보를 프로파일링 단계에서 계산한 템플릿과 최대우도추정법(maximum likelihood approach)을 이용하여 중간값을 추측할 수 있다. 중간값을 추측한 후에 알고있는 데이터를 이용하여 비밀키를 복구 할 수 있다.

템플릿 공격에서도 논프로파일링 공격에서와 마찬가지로 파형의 정렬이 필수적으로 요구된다. 유의미한 시점이 많아질수록 연산복잡도 및 공간복잡도가 지수적으로 증가하게 된다. 이를 완화하기 위해 PCA(주성분분석, Principal Component Analysis), LDA(선형판별분석, Linear Discriminant Analysis) 등이 적용되기도 한다[20,21]. 또한, 최대우도추정법을 위해 공분산 행렬의 역행렬들을 계산할 때 수치연산 문제가 발생하는데 이를 방지하기 위해 공통 공분산행렬(pooled covariance matrix)등을 사용하는 방법이 제안되었다[22].

3.3. 대응기법

부채널 분석을 방지하기 위해 마스킹 대응기법과 하이딩 대응기법등이 사용된다[12]. 마스킹 대응기법은 실제 랜덤 값을 이용하여 중간값을 랜덤하게 보이도록 함으로써 중간값을 올바르게 추론하지 못하도록 함으로써 부채널 분석을 방지한다[23,24]. 하이딩 대응기법은 신호대잡음비(Signal-to-Noise, SNR)를 감소시켜 부채널 정보와 연산 및 처리되는 데이터 간의 관계성

을 감소시키는 기법이다. 랜덤 jitter나 연산 순서를 바꾸는 방식으로 가능하며[25,26,27], 또 다른 방법으로는 연산과 데이터에 무관하게 일정한 부채널 정보가 발생하도록 하는 방법이 있다[28,29,30,31].

부채널 분석이 실제 하드웨어에서 암호알고리즘을 운용할 때 설계자가 의도치 않은 정보가 발생하여 가능하듯이, 대응기법을 적용하더라도 부채널 취약점이 발생할 수 있다. 이를 방지하기 위해 대응기법이 정확히 적용되었는지를 TVLA(Test Vector Leakage Assessment)와 같은 방법을 통해 검증하여야만 한다[32].

Ⅳ. 딥러닝을 이용한 부채널 분석 기술 연구 동향

본 장에서는 딥러닝을 이용한 부채널 분석 기술 동향을 체계적으로 소개하기 위해 딥러닝을 부채널 분석에 적용한 용도별로 분류하여 소개한다. 그림 1은 딥러닝 기반 부채널 분석 기술의 공격자 환경 및 용도별 분류를 간단히 정리한 그림이다.

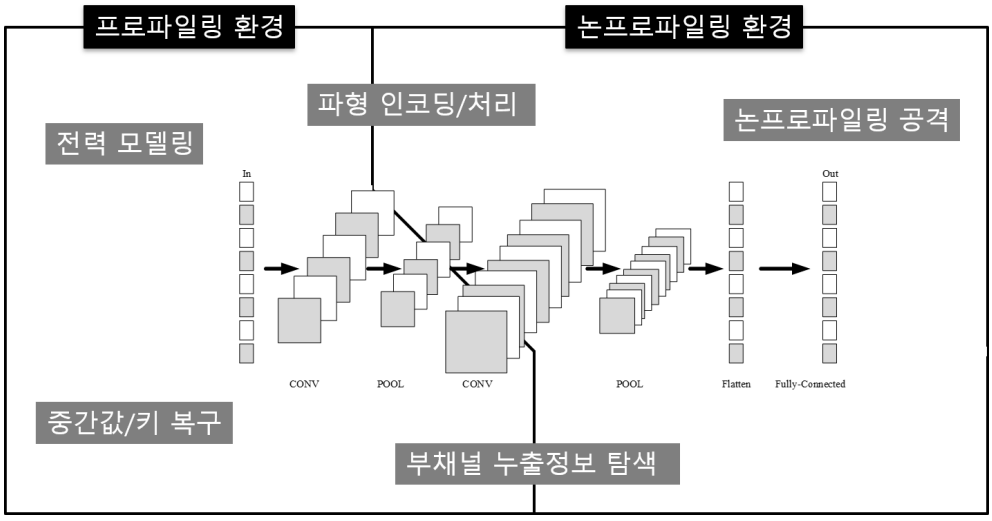
4.1. 전력 모델링

ICISC 2011에서 Yang등은 부채널 분석 공격 성능에 중요한 영향을 미치는 전력 모델을 향상시키기 위해 신경망을 전력모델링에 사용하는 방법을 제안하였다[33]. 분석 대상인 중간값을 입력으로 하고 실제 전력값을 라벨로 사용하였다. 학습주기마다 입력에 대한 신경망의 출력값과 라벨 간의 상관계수를 계산하여 일정 정도 수치를 넘을 때까지 신경망을 학습하였다. 신경망이 비선형 함수를 학습할 수 있음을 가정하여 입력으로 사용되는 중간값의 각 비트값이 전력모델에 미치는 영향을 학습할 수 있다는 성질을 이용하였다. 학습을 마치고 난 후에 신경망을 전력모델로 사용하면 신경망의 비선형성으로 인해 일반적으로 사용되는 해밍웨이트 혹은 선형회귀를 통해 계산한 전력모델보다 성능이 좋음을 실험적으로 확인하였다.

4.2. 대칭키암호에 대한 프로파일링 중간값/키 분석

프로파일링 부채널 공격에서 중간값/키 분석을 하는 것은 신경망을 이용하여 데이터를 분류하는 것과 같은 작업을 하는 것으로 초기 딥러닝을 이용한 부채널 분석은 이러한 방향의 연구가 주를 이루었다. 딥러닝을 이용한 프로파일링 중간값/키 분석은 다음과 같은 과정으로 수행된다. 프로파일링 단계(학습단계)에서 중간값/키를 분석하기 위해 파형을 신경망의 입력으로 사용하고 중간값/키를 라벨로 사용하여 신경망을 학습하면 학습된 신경망은 중간값/키 분류기로써 공격 단계(테스트 단계)에서 사용된다.

Martinasek등이 처음으로 신경망을 프로파일링 중간값/키 분석에 사용하였다[34]. 파형의 잡음을 제거하고 학습 후 성능을 향상시키기 위해 평균파형만을 입



〔그림 1〕 딥러닝 기반 부채널 분석의 용도별 분류

력으로 사용하였으며 학습해야하는 특징을 살리기 위해 평균파형의 평균을 각 평균파형에 빼주어 입력으로 사용하는 기법을 제안하였다[35]. 또한, 초기에는 신경망이 특징을 자동으로 학습하는 성질을 이용하는 대신 기존에 사용되었던 PCA와 같은 부채널 전처리를 사용하는 연구들이 주를 이루었다[36,37].

마스킹 대응기법이 적용된 AES에 대해 딥러닝을 이용하기도 하였다. 처음엔 마스킹 값 혹은 마스킹된 중간값을 라벨로 사용하는 방식으로 사용되었으나[36,37], Maghrebi등이 SPACE 2016에서 단일 신경망을 이용할 때 마스킹을 모를 때 실제 중간값을 라벨로 사용하더라도 분석이 가능함을 보였으며 PCA를 적용하지 않은 MLP(Multilayer perceptron)가 PCA를 적용한 MLP보다 성능이 좋음을 실험 결과를 제시하였다[38]. 이는 부채널 전처리없이 신경망이 자체의 성질을 이용하여 유의미한 시점을 학습 가능함을 보여주는 첫 결과이다.

Cagli등은 CHES2017에서 처음으로 jitter 기반의 하이딩 대응기법에 대해 CNN(Convolutional Neural Network)을 이용하여 분석이 가능함을 보였다[39]. CNN을 이용하여 하이딩 대응기법이 적용된 파형에 대한 학습을 진행할 시 부족한 데이터 양을 늘리고 성능을 향상시킬 수 있는 기법인 Data augmentation 기법을 함께 사용할 것을 제안하였으며 실험을 통해 효과가 있음을 보였다.

Cagli등의 CNN기반 부채널 분석 기술이 발표된 이후 중간값/키 분석의 성능을 향상시키는 기법들에 대한 연구들이 발표되었다. 첫째로 Domain Knowledge(DK)라는 뉴런을 CNN에서 Fully-connected layer 전에 추가적인 입력으로 사용하는 기법이다[40]. 평문 혹은 암호문과 같은 알려진 데이터를 DK 뉴런의 입력으로 사용하여 성능 향상을 도모하였다. 두 번째로는 파형의 입력 형태를 변환하는 기법이다. 1차원 파형을 2차원의 스펙트로그램으로 변환하여 입력으로 사용하는 기법[41], denoising autoencoder와 같이 파형에 잡음을 추가하여 입력으로 사용함으로써 일반화가 더 잘되게하는 기법[42], 커널 기법과 유사하게 1차원 파형을 이미지화하여 입력으로 사용하는 기법[43]등이 대표적이다.

템플릿 공격의 프로파일링/공격 단계에서의 장비가 서로 다를 때 발생하는 교차 장비간의 포팅 문제가 신경망을 이용한 프로파일링 중간값/키 분석 기법에서도 동일하게 발생한다. 이를 극복하기 위해 여러 장비에서 수집된 파형을 함께 학습에 사용하는 기법이 제안되었다[44,45]. 여러 장비에서 수집된 파형을 학습에 사용하면 각 장비별 특징 차이가 극복되도록 신경망을 일반화하여 학습이 가능하다.

Zaid등은 부채널 누출신호 탐색에 사용된 기법을 이용하여 효율적인 CNN을 설계하는 방법을 제안하였다[46]. CNN 각 층의 weight visualization, gradient visualization, heatmap등을 이용하여 CNN 기반 부채널분석에서 하이퍼 파라미터가 성능에 영향을 미치는 정도를 분석함으로써 부채널 분석에 적합한 CNN 하이퍼 파라미터를 선택하는 방법을 제안하였다. 이를 통해 효율적인 CNN 하이퍼 파라미터를 선택할 수 있다.

4.3. 공개키암호에 대한 프로파일링 중간값/키 분석

공개키 암호에 대해 신경망을 이용하여 중간값을 분석한 결과들도 발표되었다. CHES 2019에 Carbone등은 EAL4+인증 받은 아시아 지역 회사에서 생산된 IC에서 메시지, 지수, 모듈러 블라인딩이 모두 적용된 RSA에 대해 중간값 혹은 사용된 레지스터를 학습함으로써 중간값을 분석한 결과를 발표하였다[47]. Weissbart등은 EdDSA에 대해 신경망을 이용하여 분석하였다[48]. Zhou등은 montgomery ladder 스칼라 곱셈 알고리즘에 대해 신경망을 이용한 분석 결과를 제시하였다[49]. Fully Convolutional Networks를 사용하였으며 유의미한 시점 선택과정이 없을 때보다 SNR 기반의 유의미한 시점을 선택할 경우 성능이 높음을 확인하였다. 현재까지 발표된 신경망 공개키 암호 분석 연구들에서는 정렬된 파형을 입력으로 사용하였다. Zhou등이 딥러닝 기반 부채널 분석시 파형 정렬이 여전히 필요함을 말한 이유[49,50] 그리고 공개키 암호에서 기본 연산에 파형에 노이즈가 더 많기 때문에 정렬된 파형을 사용한 것으로 추측된다.

4.4. 논프로파일링 중간값/키 분석

Timon은 처음으로 딥러닝을 이용한 논프로파일링 공격 기법을 제안하였다[51]. 이 기법은 논프로파일링 환경 하에 신경망의 학습을 진행한다. 프로파일링을 이

용한 중간값 분석에서처럼 파형과 중간값을 각각 입력과 라벨로 사용하여 학습하는 방식을 이용한다. 부채널 분석 관점에서 옳은 키를 가지고 계산한 중간값은 파형과 연관성이 있고 틀린 키를 가지고 계산한 중간값은 파형과 연관성이 없을 것이다. Timon의 기법은 각 키 가정에 따라 신경망을 학습할 때 연관성이 있을 때는 학습이 잘되고 연관성이 없을 때는 학습이 안 될 것이라는 경향성을 이용하여 비밀키를 찾을 수 있다. 이로 인해 이 기법을 Differential deep learning Analysis(DDLA)라 명명하였다.

DDLA에서는 다른 중간값/키 분석을 하는 딥러닝 기반 부채널 분석 기법들과 다르게 중간값을 일대일 함수인 원핫인코딩으로 라벨화시키지 않고 일대일 함수가 아닌 중간값의 HW, LSB, MSB 값에 대한 원핫 인코딩 결과를 라벨로 사용한다. 이는 Timon의 기법이 파티션 기반 부채널 분석이므로 일대일대응 라벨화 방법을 이용할 경우 옳은 키와 틀린 키에 상관없이 단순한 클러스터링 문제가 되므로 학습 경향에서 구분되지 않기 때문이다[52,53].

4.5. 파형 인코딩/전처리

신경망을 파형 인코딩 혹은 파형 전처리하는 방법은 세 가지가 제안되어있다. 첫 번째 기법은 논프로파일링 공격의 성능이 향상되도록 신경망을 파형 인코더로 학습한 기법으로 Robyns 등에 의해 CHES 2019에 발표되었다[54]. 파형을 입력으로 사용할 때 출력된 노드가 중간값 혹은 중간값의 전력모델값과 상관계수가 커지도록 미니배치 단위의 상관계수를 이용한 손실함수를 사용하였다. 또한, 정렬되지 않은 파형에 대해 FFT 변환을 하여 주파수 도메인의 파형을 사용할 경우 CNN이 아닌 은닉층이 한 층인 얕은 MLP를 이용해도 충분히 분석 가능함을 보였다.

두 번째 기법은 논프로파일링 환경하에 오토인코더(autoencoder)를 잡음 제거하는 전처리기로 사용한 기법이다[55]. 파형을 입력으로 사용하고 같은 데이터에 해당하는 파형들의 평균 파형을 출력으로 사용하여 오토인코더로 학습할 경우 잡음이 제거되도록 신경망이 학습될 수 있다는 성질을 이용한 기법이다. denoising autoencoder와 유사하지만 부채널 분석에 사용되는 파형에 잡음을 추가할 때 잡음 추가는 심한 잡음을 유발하여 오히려 잡음 제거를 못하게 되는 경향이 발생할 수 있다. 해당 기법은 이런 경향을 회피하기 위해 반대로 라벨로 잡음을 없앤 평균 파형을 이용하였다.

세 번째 기법은 프로파일링 환경에서 대응기법을 제거하도록 파형을 변환하는 기법이다[56]. 두 번째 기법과 유사한 방식으로, 대응기법을 잡음으로 간주하여 입력으로는 대응기법이 적용된 파형을 사용하고 라벨로는 대응기법이 없는 파형을 사용하여 신경망을 학습한다. 실험을 통해 실제 학습된 신경망이 대응기법을 무력화하도록 파형을 변환시킬 수 있음을 보였다.

4.6. 부채널 누출신호 탐색

설명가능한 인공지능(explainable AI, XAI) 기술의 일종인 attribution method를 이용하여 부채널 누출신호를 탐색하는 기법들이 독립적으로 제안되었다. 먼저 프로파일링 환경하에 누출신호를 탐색하기 위해 Hettwer등은 Saliency map, Layer-wise Relevance Propagation, Occlusion 방법들을 이용하였다[57]. Masure등은 sensitivity analysis 방법을 이용하는 gradient visualization 기법을 제안하였으며, 얻어낸 부채널 누출신호의 위치를 템플릿 공격의 유의미한 시점으로 사용하면 기존 부채널 분석에서 사용되던 유의미한 시점을 선택하는 방법보다 성능이 높아짐을 실험을 통하여 확인하였다[58]. Timon은 논프로파일링 환경에서 학습된 MLP의 첫 번째 은닉층의 가중치의 절댓값 합을 이용하거나 MLP/CNN에 대해 sensitivity analysis를 이용하여 누출신호를 탐색하는 방법을 제안하였다[51]. 앞선 3가지 기법은 부채널 누출신호 탐색뿐만 아니라 딥러닝 기반 부채널 분석에서 신경망이 실제 부채널 정보를 자체적으로 학습이 가능하다는 것을 알 수 있음을 내포한다. 또한, 이를 통해 신경망이 실제로 학습을 유효하게 하였는지를 판별 할 수 있다.

Wegener등은 딥러닝을 이용하여 부채널 누출신호 평가에 사용하는 기법인 DL-LA(Deep Learning Leakage Assessment)를 제안하고 기존 기술인 t-test와 χ^2-test보다 성능이 우수할 수 있음을 실험을 통해 확인하였다[59]. DL-LA 기법은 신경망의 특성으로 인해 평가자가 평가할 파형에 대한 위치, 정렬 상태, 누출의 통계적 차수 등을 고려할 필요가 없다는 특징을 가진다.

4.7. 이외 동향

Benadjila등은 딥러닝 기반 부채널 분석 연구가 다양해짐에 따라 그에 대한 성능을 검증하기 위해 벤치마킹용 데이터인 ASCAD를 공개하였다[60]. 마스킹 대응기법이 적용된 AES의 전자파 방출 파형이며 jitter 시뮬레이션한 파형도 제공한다.

Masure등은 교차 엔트로피(cross entropy)를 손실함수로 사용할 때 교차 엔트로피와 perceived information과의 연관성을 확인함으로써 계산이 힘든 perceived information을 효율적으로 계산 가능하며 실제 신경망이 제대로 학습되었는지 여부를 확인 가능함을 보였다[61]. 또한, 이를 기반으로 기존 부채널 대응기법인 불 마스킹, 실행 순서 셔플링 등이 딥러닝 기반 부채널 분석에서도 유효함을 확인하였다.

암호에 대한 부채널 분석에 딥러닝을 이용하는 것뿐만 아니라 부채널 분석을 통해 신경망의 가중치를 알아내거나 입력되는 데이터를 복구하는 연구들도 진행되었다[62,63,64].

Ⅴ. 결 론

본 논문에서는 딥러닝을 이용한 부채널 분석 기술의 최신 연구 동향을 체계적으로 정리하였다. 다양한 응용분야에서 좋은 성능을 보인 딥러닝 기술이 부채널분석에서도 좋은 성능을 보임이 확인되면서 딥러닝 기반 부채널 분석의 새로운 기술들이 제안되었으며 그 성능 검증에 대한 연구들이 진행되고 있다. 이러한 흐름은 아직 초기 수준이며 더 많은 검증이 필요하므로 앞으로도 활발한 연구가 진행될 것으로 보인다.

참 고 문 헌

[1] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436.

[2] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

[3] Polyak, Boris T. "Some methods of speeding up the convergence of iteration methods." USSR Computational Mathematics and Mathematical Physics 4.5 (1964): 1-17.

[4] Sutskever, Ilya, et al. "On the importance of initialization and momentum in deep learning." International conference on machine learning. 2013.

[5] Hinton, Geoffrey, Nitish Srivastava, and Kevin Swersky. "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent." Neural Netw. Machine Learn., Coursera MOOC, 2012. https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

[6] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[7] Wolpert, David H., and William G. Macready. "No free lunch theorems for optimization." IEEE transactions on evolutionary computation 1.1 (1997): 67-82.

[8] Tikhonov, Andrey Nikolayevich. "On the stability of inverse problems." Dokl. Akad. Nauk SSSR. Vol. 39. 1943.

[9] Tibshirani, Robert. "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society: Series B (Methodological) 58.1 (1996): 267-288.

[10] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The journal of machine learning research 15.1 (2014): 1929-1958.

[11] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." arXiv preprint arXiv:1502.03167 (2015).

[12] Mangard, Stefan, Elisabeth Oswald, and Thomas Popp. Power analysis attacks: Revealing the secrets of smart cards. Vol. 31. Springer Science & Business Media, 2008.

[13] Kocher, Paul, Joshua Jaffe, and Benjamin Jun. "Differential power analysis." Annual International Cryptology Conference. Springer, Berlin, Heidelberg, 1999.

[14] Brier, Eric, Christophe Clavier, and Francis Olivier. "Correlation power analysis with a leakage model." International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 2004.

[15] Gierlichs, Benedikt, et al. "Mutual information analysis." International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 2008.

[16] van Woudenberg, Jasper GJ, Marc F. Witteman, and Bram Bakker. "Improving differential power analysis by elastic alignment." Cryptographers' Track at the RSA Conference. Springer, Berlin, Heidelberg, 2011.

[17] Muijrers, Ruben A., Jasper GJ van Woudenberg, and Lejla Batina. "RAM: Rapid alignment method." International Conference on Smart Card Research and Advanced Applications. Springer, Berlin, Heidelberg, 2011.

[18] Chari, Suresh, Josyula R. Rao, and Pankaj Rohatgi. "Template attacks." International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 2002.

[19] Schindler, Werner, Kerstin Lemke, and Christof Paar. "A stochastic model for differential side channel cryptanalysis." International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 2005.

[20] Archambeau, Cédric, et al. "Template attacks in principal subspaces." International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 2006.

[21] Standaert, François-Xavier, and Cédric Archambeau. "Using subspace-based template attacks to compare and combine power and electromagnetic information leakages." International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 2008.

[22] Choudary, Omar, and Markus G. Kuhn. "Efficient template attacks." International Conference on Smart Card Research and Advanced Applications. Springer, Cham, 2013.

[23] Chari, Suresh, et al. "Towards sound approaches to counteract power-analysis attacks." Annual International Cryptology Conference. Springer, Berlin, Heidelberg, 1999.

[24] Goubin, Louis, and Jacques Patarin. "DES and differential power analysis the "Duplication" method." International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 1999.

[25] Coron, Jean-Sébastien, and Ilya Kizhvatov. "An efficient method for random delay generation in embedded software." International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 2009.

[26] Coron, Jean-Sébastien, and Ilya Kizhvatov. "Analysis and improvement of the random delay countermeasure of CHES 2009." International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 2010.

[27] Veyrat-Charvillon, Nicolas, et al. "Shuffling against side-channel attacks: A comprehensive study with cautionary note." International Conference on the Theory and Application of Cryptology and Information Security. Springer, Berlin, Heidelberg, 2012.

[28] Tiri, Kris, Moonmoon Akmal, and Ingrid Verbauwhede. "A dynamic and differential CMOS logic with signal independent power consumption to withstand differential power analysis on smart cards." Proceedings of the 28th European solid-state circuits conference. IEEE, 2002.

[29] Popp, Thomas, and Stefan Mangard. "Masked dual-rail pre-charge logic: DPA-resistance without routing constraints." International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 2005.

[30] Chen, Cong, et al. "Balanced encoding to mitigate power analysis: a case study." International Conference on Smart Card Research and Advanced Applications. Springer, Cham, 2014.

[31] Maghrebi, Houssem, Victor Servant, and Julien Bringer. "There is wisdom in harnessing the strengths of your enemy: customized encoding to thwart side-channel attacks." International Conference on Fast Software Encryption. Springer, Berlin, Heidelberg, 2016.

[32] Becker, George, et al. "Test vector leakage assessment (TVLA) methodology in practice." International Cryptographic Module Conference. Vol. 1001. 2013.

[33] Yang, Shuguo, et al. "Back propagation neural network based leakage characterization for practical security analysis of cryptographic implementations." International Conference on Information Security and Cryptology. Springer, Berlin, Heidelberg, 2011.

[34] Martinasek, Zdenek, and Vaclav Zeman. "Innovative method of the power analysis." Radioengineering 22.2 (2013): 586-594.

[35] Martinasek, Zdenek, Jan Hajny, and Lukas Malina. "Optimization of power analysis using neural network." International Conference on Smart Card Research and Advanced Applications. Springer, Cham, 2013.

[36] Gilmore, Richard, Neil Hanley, and Maire O'Neill. "Neural network based attack on a masked implementation of AES." 2015 IEEE International Symposium on Hardware Oriented Security and Trust (HOST). IEEE, 2015.

[37] Martinasek, Zdenek, et al. "Power analysis attack based on the MLP in DPA Contest v4." 2015 38th International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2015.

[38] Maghrebi, Houssem, Thibault Portigliatti, and Emmanuel Prouff. "Breaking cryptographic implementations using deep learning techniques." International Conference on Security, Privacy, and Applied Cryptography Engineering. Springer, Cham, 2016.

[39] Cagli, Eleonora, Cécile Dumas, and Emmanuel Prouff. "Convolutional neural networks with data augmentation against jitter-based countermeasures." International Conference on Cryptographic Hardware and Embedded Systems. Springer, Cham, 2017.

[40] Hettwer, Benjamin, Stefan Gehrer, and Tim Güneysu. "Profiled power analysis attacks using convolutional neural networks with domain knowledge." International Conference on Selected Areas in Cryptography. Springer, Cham, 2018.

[41] Yang, Guang, et al. "Convolutional Neural Network Based Side-Channel Attacks in Time-Frequency Representations." International Conference on Smart Card Research and Advanced Applications. Springer, Cham, 2018.

[42] Kim, Jaehun, et al. "Make Some Noise. Unleashing the Power of Convolutional Neural Networks for Profiled Side-channel Analysis." IACR Transactions on Cryptographic Hardware and Embedded Systems (2019): 148-179.

[43] Won, Yoo-Seung, and Jong-Yeon Park. "Non-Profiled Side Channel Attack based on Deep Learning using Picture Trace." IACR Cryptology ePrint Archive, Report 2019/1242, 2019.

[44] Golder, Anupam, et al. "Practical Approaches Toward Deep-Learning-Based Cross-Device Power Side-Channel Attack." IEEE Transactions on Very Large Scale Integration (VLSI) Systems 27.12 (2019): 2720-2733.

[45] Das, Debayan, et al. "X-DeepSCA: Cross-device deep learning side channel attack." Proceedings of the 56th Annual Design Automation Conference 2019. ACM, 2019.

[46] Zaid, Gabriel, et al. "Methodology for efficient CNN architectures in profiling attacks." IACR Transactions on Cryptographic Hardware and

Embedded Systems (2020): 1-36.

[47] Carbone, Mathieu, et al. "Deep learning to evaluate secure RSA implementations." IACR Transactions on Cryptographic Hardware and Embedded Systems (2019): 132-161.

[48] Weissbart, Leo, Stjepan Picek, and Lejla Batina. "One Trace Is All It Takes: Machine Learning-Based Side-Channel Attack on EdDSA." International Conference on Security, Privacy, and Applied Cryptography Engineering. Springer, Cham, 2019.

[49] Zhou, Yuanyuan, and François-Xavier Standaert. "Simplified Single-Trace Side-Channel Attacks on Elliptic Curve Scalar Multiplication using Fully Convolutional Networks."

[50] Zhou, Yuanyuan, and François-Xavier Standaert. "Deep learning mitigates but does not annihilate the need of aligned traces and a generalized ResNet model for side-channel attacks." Journal of Cryptographic Engineering (2019): 1-11.

[51] Timon, Benjamin. "Non-profiled deep learning-based side-channel attacks with sensitivity analysis." IACR Transactions on Cryptographic Hardware and Embedded Systems (2019): 107-131.

[52] Standaert, François-Xavier, Benedikt Gierlichs, and Ingrid Verbauwhede. "Partition vs. comparison side-channel distinguishers: An empirical evaluation of statistical tests for univariate side-channel attacks against two unprotected cmos devices." International Conference on Information Security and Cryptology. Springer, Berlin, Heidelberg, 2008.

[53] Whitnall, Carolyn, Elisabeth Oswald, and François-Xavier Standaert. "The myth of generic DPA… and the magic of learning." Cryptographers' Track at the RSA Conference. Springer, Cham, 2014.

[54] Robyns, Pieter, Peter Quax, and Wim Lamotte. "Improving CEMA using correlation optimization." IACR Transactions on Cryptographic Hardware and Embedded Systems (2019): 1–24.

[55] 권동근, et al. "비프로파일링 기반 전력 분석의 성능 향상을 위한 오토인코더 기반 잡음 제거 기술." 정보보호학회논문지 29.3 (2019): 491-501.

[56] Wu, Lichao, and Stjepan Picek. "Remove Some Noise: On Pre-processing of Side-channel Measurements with Autoencoders." IACR Cryptology ePrint Archive, Report 2019/1474, 2019.

[57] Hettwer B., Gehrer S., Güneysu T. "Deep Neural Network Attribution Methods for Leakage Analysis and Symmetric Key Recovery." Selected Areas in Cryptography – SAC 2019. Springer, Cham, 2019.

[58] Masure, Loïc, Cécile Dumas, and Emmanuel Prouff. "Gradient visualization for general characterization in profiling attacks." International Workshop on Constructive Side-Channel Analysis and Secure Design. Springer, Cham, 2019.

[59] Wegener, Felix, Thorben Moos, and Amir Moradi. "DL-LA: deep learning leakage assessment: A modern roadmap for SCA evaluations." IACR IACR Cryptology ePrint Archive, Report 2019/505, 2019.

[60] Benadjila, Ryad, et al. "Deep learning for side-channel analysis and introduction to ASCAD database." Journal of Cryptographic Engineering (2019): 1-26.

[61] Masure, Loïc, Cécile Dumas, and Emmanuel Prouff. "A comprehensive study of deep learning for side-channel analysis." IACR Transactions on Cryptographic Hardware and Embedded Systems (2020): 348-375.

[62] Hua, Weizhe, Zhiru Zhang, and G. Edward Suh. "Reverse engineering convolutional neural networks through side-channel information leaks." 2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC). IEEE, 2018.

[63] Wei, Lingxiao, et al. "I know what you see:

Power side-channel attack on convolutional neural network accelerators." Proceedings of the 34th Annual Computer Security Applications Conference. ACM, 2018.

[64] Batina, Lejla, et al. "Csi neural network: Using side-channels to recover your artificial neural network information." arXiv preprint arXiv:1810.09076 (2018).

〈저 자 소 개〉

진 성 현 (Sunghyun Jin)

학생회원

2015년 2월 : 서울시립대학교 수학과, 컴퓨터과학 학사

2017년 2월 : 고려대학교 정보보호학과 석사

2017년 3월~현재 : 고려대학교 정보보호학과 박사과정

<관심분야> 정보보호, 부채널 공격, 머신러닝 기반 암호분석

김 희 석 (HeeSeok Kim)

정회원

2006년 : 연세대학교 수학과 학사

2008년 : 고려대학교 정보보호대학원 석사

2011년 : 고려대학교 정보보호대학원 박사

2011년 9월~2012년 12월 : Bristor University 박사후 연구원

2013년~2016년 8월 : 한국과학기술정보연구원(KISTI) 선임연구원

2015년~2016년 8월 : 과학기술연합대학원대학교(UST) 조교수

2016년 9월~현재 : 고려대학교 과학기술대학 사이버보안전공 부교수

<관심분야> 부채널 공격, 암호시스템 안전성 분석 및 고속구현, 암호칩 설계 기술, 보안관제, 네트워크 보안