부채널 파형 데이터를 사용한 머신러닝 암호 분류

권혁동[1], 김현지[1], 서화정[2]
[1]한성대학교 정보컴퓨터공학과
[2]한성대학교 IT융합공학부
korlethean@gmail.com, khj1594012@gmail.com, hwajeong84@gmail.com

Machine learning cipher classification using side-channel trace data

Hyeok-Dong Kwon[1], Hyun-Ji Kim[1], Hwa-Jeong Seo[2]
[1]Dept. of Information Computer Engineering, Hansung University
[2]Dept. of IT Convergence Engineering, Hansung University

요 약

부채널 분석은 하드웨어에서 발생하는 빛, 열, 전자기파와 같은 각종 부채널 정보를 이용하는 공격이다. 부채널 분석은 강력한 보안 위협에 속하지만, 부채널 정보 분석에 오랜 시간과 노력이 소요된다. 때문에 부채널 분석에 머신러닝을 접목하고자 하는 연구가 진행되었다. 머신러닝은 대량의 데이터를 학습하고 패턴을 파악하는데 용이하기 때문에 대량의 부채널 정보를 분석하는데 유리하다. 본 논문에서는 부채널 파형 데이터를 사용하여 암호 분류를 하는 머신러닝 모델을 소개한다.

1. 서론

부채널 분석(Side-channel attacks)은 암호 알고리즘이 가동되는 하드웨어에서 발생하는 부채널 정보를 습득하여 암호 키를 획득하는 공격이다. 부채널 분석은 대량의 데이터를 처리하기 때문에 공격자가 데이터를 분석하는데 많은 시간이 소요된다. 이를 해결하기 위해서 부채널 공격에 머신러닝을 사용하고자 하는 연구가 진행되었다. 본 논문에서는 부채널 데이터를 사용한 암호 분류 기법에 대해서 소개한다.

본 논문의 구성은 다음과 같다. 2장에서 부채널 공격과 머신러닝에 대한 사항을 확인한다. 3장에서 제안하는 머신러닝 모델과 성능에 대해서 소개한다. 마지막으로 4장에서 결론을 내린다.

2. 부채널 분석과 머신러닝

부채널 분석은 1996년 Kocher가 제안한 새로운 유형의 암호 알고리즘 공격 기법으로, 하드웨어에서 발생하는 부가적인 정보를 공격하는 암호 취약점 중 하나이다[1]. 부채널 정보에 해당되는 것들로는 빛, 열, 전자기파 등이 있다.

부채널 공격 중 가장 많이 사용되는 요소는 전력량으로, 흔히 전력 분석 공격(Power analysis attacks)이라 한다[2]. 전력 분석은 크게 세 종류로 나뉘며, 단순 전력 분석(Simple power analysis), 차분 전력 분석(Differential power analysis), 상관관계 분석(Correlation power analysis)이 있다.

부채널 공격을 단순하게 표현하자면 그림 1과 같이 표현이 가능하다.

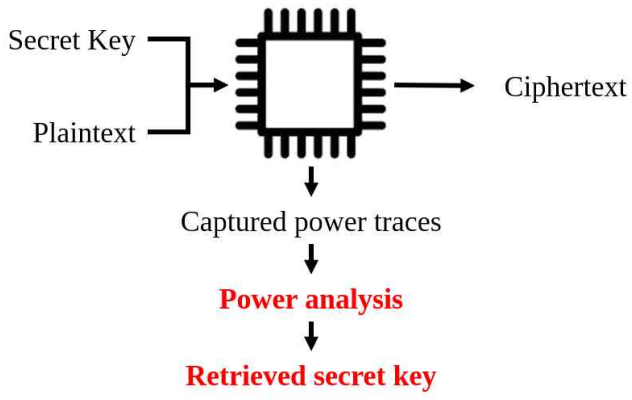

Fig. 1. Steps of power analysis attack.

머신러닝은 인공 신경망(Neural network) 개념에서 출발하였으며, 1958년 Rosenblatt가 제안한 퍼셉트론(perceptron)에 근간을 둔다[3]. 퍼셉트론은 선형 문제를 해결할 수 있지만, 비선형 문제에는 적합하

지 못했다. 이후 이를 해결할 수 있는 다층 퍼셉트론이 등장하였고, 이를 기반으로 머신러닝이 발전하기 시작하였다. 특히 다층 신경망을 사용하는 딥러닝 네트워크는 복잡한 문제, 대량의 데이터를 처리할 수 있게 되었다.

부채널 분석에는 다양한 장비가 있으며, 그 중에서 ChipWhisperer 장비는 저렴하면서도 사용하기 쉬운 장비로 알려져있다. ChipWhisperer는 캐나다의 NewAE Technology에서 개발한 부채널 장비 및 부채널 분석 소프트웨어의 명칭이다.

부채널 분석의 데이터는 상당한 데이터가 존재하기 때문에 머신러닝을 접목하는 다양한 연구 결과가 있다. ICISC 2011에서는 전력 분석 시 공격 성능에 영향을 주는 전력 모델링에 머신러닝을 사용하는 방법이 제안되었다[4]. SPACE 2016에서는 부채널 대응 기법 중 마스킹 기법이 적용된 AES 암호 분석에 머신러닝을 사용하였다[5,6]. 공개키 암호에 대한 분석도 진행되었다. CHES 2019에서는 RSA에 대해 중간 값 분석에 머신러닝을 사용하여 분석한 결과물이 제시되었다[7].

이처럼 부채널 분석의 다양한 방면에서 머신러닝이 동원되고 연구가 활발함을 알 수 있다.

3. 암호 분류 머신러닝 모델

머신러닝을 사용한 부채널 분석에서 모델은 주어진 전력 데이터를 분석하는데 사용된다. 이때 모델은 어떤 암호 알고리즘의 파형인지 미리 알고 있는 상태이다. 하지만 일반적으로 공격자는 하드웨어에서 어떤 암호 알고리즘이 가동되는지 알기가 어렵다.

암호 알고리즘은 각각 고유한 내부 구조를 가지며, 연산에서 발생하는 전력량도 다르다. 이 점에 착안하여 전력 파형 데이터를 가지고 암호 분류를 하는 모델을 작성한다. 대상으로 하는 암호 알고리즘은 AES-128, LEA-128, LEA-192, LEA-256, CHAM-64/128, CHAM-128/128, CHAM-128/256, PIPO-64/128, PIPO-128/128로 총 4종 9규격의 암호를 대상으로 한다.

우선 각각의 암호 알고리즘을 가동하면서 부채널 전력 데이터를 수집한다. 파형 데이터는 ChipWhisperer 부채널 장비를 사용하여 수집한다. 사용하는 장비는 ChipWhisperer-Lite 장비로, 파형 수집 장비인 오실로스코프와 암호 알고리즘 연산을 진행하는 타깃 프로세서가 하나의 보드로 연결되어 있다. 그림 2에서는 ChipWhisperer의 외형을 확인할 수 있다. ChipWhisperer을 사용하기 위해서 PC에 연결할 때는 USB를 통해 간단히 연결할 수 있다.

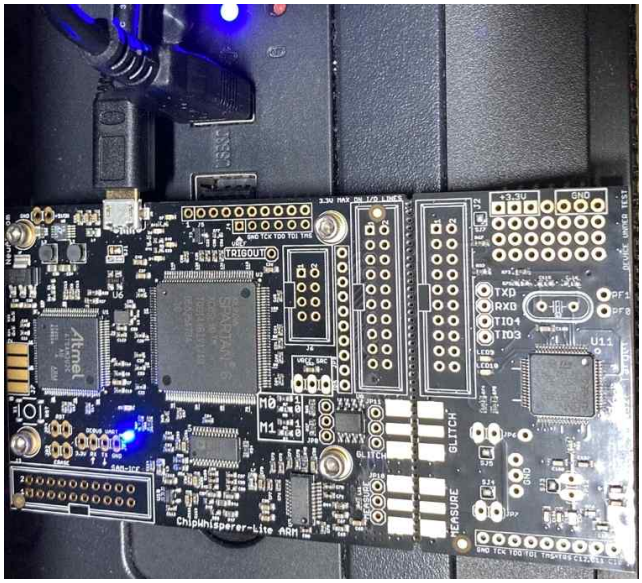

Fig. 2. Appearance of ChipWhisperer-Lite 32-bit.

암호 알고리즘을 가동하기 위해서 ChipWhisperer의 SimpleSerial을 사용하여 ChipWhisperer 하드웨어와 PC와 통신을 진행한다. 주고받는 정보는 크게 두 가지가 존재한다. 첫째는 구현한 암호 알고리즘이다. 구현한 암호 알고리즘을 컴파일하여 SimpleSerial을 통해서 컴파일한 데이터를 전송하여 타깃 프로세서에서 암호 알고리즘을 연산하도록 한다. 둘째는 타깃 프로세서에서 발생한 전력 값이다. 타깃 프로세서는 전송받은 암호 알고리즘을 연산하는데, ChipWhisperer의 Capture 보드에서 파형을 수집한다. 이후 수집된 파형 데이터는 SimpleSerial을 통해 PC로 전송되어 파형 데이터를 확인할 수 있게 된다.

본 논문에서는 실험을 위해 각 알고리즘 별로 5,000개의 파형 데이터를 수집하였으며, 각 파형별로 10,000개의 샘플 수를 지니게 하였다. 대상 암호 알고리즘의 수가 9종이므로 총 45,000개의 파형을 수집하였다.

머신러닝 네트워크 모델로는 CNN(Convolution Neural Network)을 사용한다. CNN은 이미지를 부분적으로 분석하며 학습하는데 유용한 네트워크이다. 수집한 파형 데이터를 matplotlib를 사용하여 그래프 이미지로 변환한 다음 해당 데이터를 머신러닝

모델에 학습, 검증, 시험 데이터로 사용한다. 기본 모델은 그림 3과 같이 구성하며, 레이어별로 뉴런의 수를 조정하거나 과적합을 방지하기 위해 Dropout 레이어를 삽입하는 변형을 준비한다. activation function으로는 relu를 사용하고 optimizer로 adam을 사용하였다.

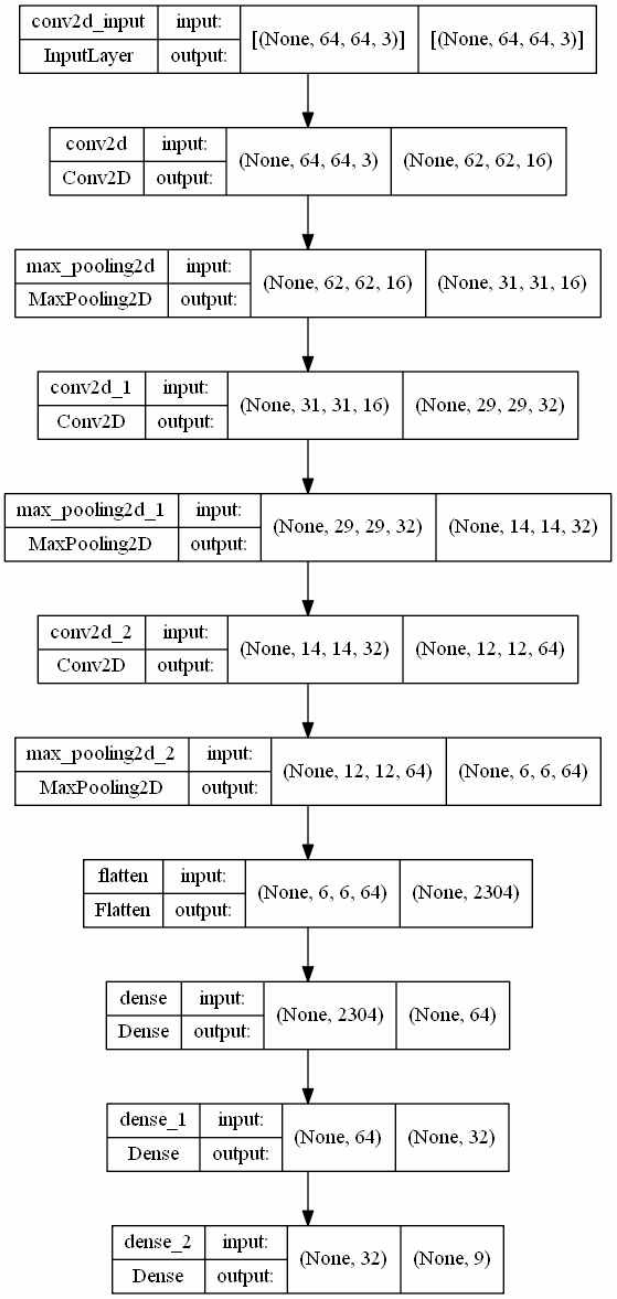

Fig. 3. Base model configuration.

수집한 데이터 45,000개는 학습, 검증, 시험 데이터로 분류하며, 이때 분류 비율은 6:2:2로 분류한다. 따라서 각각 27,000의 학습 데이터, 9,000개의 검증 데이터, 9,000개의 시험 데이터를 사용한다. 학습은 100회의 epoch를 거치며, 각각의 epoch별로 100회의 학습 과정 및 5회의 검증 과정을 거친다. 학습 결과는 표 1에서 정리한다.

Table 1. Learning results summary table. Params: Number of parameters, L-acc: Learning accuracy(%), L-loss: Learning loss, V-acc: Validation accuracy(%), V-loss: Validation loss, T-acc: Test accuracy(%).

	Params	L-acc.	L-loss	V-acc.	V-loss	T-acc.
1	173,481	93.10	0.125	94.38	0.108	92.19
2	173,481 +Dropout	77.90	0.452	85.00	0.167	88.75
3	1,303,305	90.60	0.152	93.12	0.136	93.44
4	1,303,305 +Dropout	90.35	0.142	92.50	0.122	90.62

모델은 크게 네 종류를 사용하였다. 1번 모델은 기본형 모델, 2번 모델은 기본형 모델에 Dropout 레이어를 추가하였다. 3번 모델은 기본형 모델에서 뉴런의 수를 늘린 모델, 4번 모델은 3번 모델에서 Dropout 레이어를 추가한 모델이다.

1번 모델의 학습 정확도는 약 93.1%로 가장 높은 학습 정확도를 보였다. 또한 검증 정확도 역시 약 94.38%로 다른 모델에 비해 높은 정확도를 보였고, 시험 정확도는 92.19%로 두 번째로 높은 정확도를 보였다.

2번 모델의 학습 정확도는 77.90%를 보였다. 이는 1번 모델의 학습 정확도가 과적합 되었다고 해석이 가능하다. 이는 학습 손실이 높고 검증 정확도가 낮은 것으로도 동일한 해석이 가능해진다. 최종적으로 학습 정확도는 88.75%로 가장 낮은 정확도를 보였다. 2번 모델은 1번 모델에 Dropout 레이어를 추가하여 과적합을 방지한 모델이므로, 1번 모델의 높은 정확도는 과적합에 의해 발생했다고 판단할 수 있다.

3번 모델은 1번 모델에서 레이어 별로 뉴런의 수를 늘린 모델이다. 학습 정확도와 검증 정확도는 각각 90.60%, 93.12%로 1번 모델에 비해 다소 낮아졌지만 시험 정확도는 93.44%로 4개의 모델 중 가장 좋은 정확도를 보였다.

마지막으로 4번 모델은 3번 모델에 Dropout 레이어를 추가한 모델이다. 하지만 Dropout 레이어를 추가했음에도 3번 모델에서 정확도가 크게 변하지 않은 것을 확인할 수 있다. 이는 3번 모델이 과적합되지 않고 적절한 수준으로 학습이 되었다고 판단할 수 있다.

따라서 3번 모델은 9종의 암호 알고리즘을 93.44%의 정확도로 분류할 수 있는 성능을 지녔다고 평가할 수 있다.

4. 결론

본 논문에서는 암호 알고리즘의 파형 데이터만으로 하드웨어에서 어떤 암호 알고리즘이 가동되는지 분류하는 머신러닝 모델을 작성하였다. 작성한 모델은 기본형에서 여러 종류의 파생형을 만들었고, 각각의 모델의 학습 상태를 확인하여 적합한 모델의 파라미터를 확인하였다.

구현한 모델은 파형 데이터의 그래프 이미지를 통해서 암호 알고리즘을 분류하였다. 추후 연구 과제로 그래프 이미지가 아닌, 파형 데이터의 수치 값을 통해서 분류하는 모델을 작성하여 이미지를 통한 분류와 수치를 통한 분류 모델 중 어느 것이 효과적인지를 제시한다.

5. Acknowledgement

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.2018-0-00264, IoT 융합형 블록체인 플랫폼 보안 원천 기술 연구, 100%).

참고문헌

[1] P.C.Kocher, “Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems,” In *Annual International Cryptology Conference*, Springer, Berlin, Heidelberg, pp. 104-113, 1996.

[2] P.Kocher, J.Jaffe, and B.Jun, “Differential power analysis,” In *Annual international cryptology conference*, Springer, Berlin, Heidelberg, pp. 388-397, August, 1999.

[3] F.Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychological Review*, 65(6), 386 - 408, 1958.

[4] S.yang, Y.Zhou, J.Liu,and, and D.Chen, “Back Propagation Neural Network Based Leakage Characterization for Practical Security Analysis of Cryptographic Implementations,” In *International Conference on Information Security and Cryptology*, pp. 169-185, 2011.

[5] R.Gilmore, N.Hanley, and M.O'Neill, “Neural network based attack on a masked implementation of AES,” In *IEEE International Symposium on Hardware Oriented Security and Trust*(HOST), pp. 106-111, 2015.

[6] Z.Martinasek, O.Zapletal, K.Vrba, and K.Trasy, “Power analysis attack based on the MLP in DPA Contest v4,” In *38th International Conference on Telecommunications and Signal Processing*(TSP), pp. 154-158, 2015.

[7] M.Carbone, V.Conin, M.-A.Cornélie, F.Dassance, G.Dufresne, C.Dumas, E.Prouff and A.Venelli, “ Deep Learning to Evaluate Secure RSA Implementations,” *IACR Transactions on Cryptographic Hardware and Embedded Systems*, pp. 132-161, 2019.