딥러닝 기술을 활용한 PRESENT 평문 복구 공격*

김동훈*, 권동근*, 김성겸*, 홍득조**, 성재철***, 홍석희*
*고려대학교 정보보호학과, **전북대학교 IT정보공학과, ***서울시립대학교 수학과,
dhkim85@korea.ac.kr, donggeun.kwon@gmail.com, jeffgyeom@korea.ac.kr,
jcsung@uos.ac.kr, deukjo.hong@jbnu.ac.kr, shhong@korea.ac.kr

Plaintext Recovery Attack of PRESENT Using Deep Learning

Donghoon Kim*, Donggeun Kwon*, Seonggyeom Kim*, Deukjo Hong**,
Jaechul Sung***, Seokhie Hong*
*School of Cybersecurity, Korea University
**Department of Information Technology, Chunbuk National University
***Department of Mathematics, University of Seoul

요 약

최근, 딥러닝 기술을 활용하여 암호 알고리즘 식별, 평문 복구, 이론적 암호분석을 향상시키는 방법 등이 제안되고 있다. 이 중, 2019년에 Xiao 등이 암호학적 특성을 고려되지 않고 2-라운드 DES의 평문복구 공격에 딥러닝을 적용하는 방법을 제안하였다. 본 논문에서는 이러한 기법을 향상하여 암호문과 평문의 선형 연관성을 고려한 평문 복구 공격을 딥러닝을 통해 수행하는 방법을 제안한다. 이를 활용하여, PRESENT의 평문 복구 공격을 5-라운드까지 가능함을 보인다.

1. 서론

4차산업혁명시대의 도래와 COVID-19로 인한 언택트 사회로 전환되면서 정보통신망을 이용한 정보의 소통이 급속도로 증가하고 있다. 이에 따라, 민감한 데이터에 대한 기밀성 보호가 주요 이슈로 화두되고 있다. 이러한 민감 데이터를 보호해주는 정보보호의 핵심이 암호기술이다. 암호는 입력데이터를 혼돈(Confusion)과 확산(Diffusion)의 방법을 사용하여 제 3자가 암호문을 통해 평문을 복원하거나 암호키를 추측할 수 없도록 설계한다[1].

Xiao 등은 DES의 평문을 딥러닝 기술을 활용하여 복원하는 기법을 제안하였다[2]. 이 실험에서 2라운드까지 64-비트 평문의 각 비트 위치에 해당하는 값을 평균적으로 66%의 정확도로 복원하는데 성공하였다. 그러나, 이 방법은 암호의 평문·암호문 사이에 대한 선형 연관성이 고려되지 않아 상대적으로 적은 라운드만 분석 가능하였다. 이에, 본 논문에서는 암호분석기법 중 하나인 선형분석(LC)[3] 및 다중선형분석(MLC), 다차원선형분석(MDLC)[4]에서 사용하는 선형 특성(Linear Characteristic)을 도고려한 평문복원공격에 딥러닝을 활용하는 방안을 제시한다. 그리고, 이를 PRESENT[5]에 적용하여 유의미한 결과를 얻는 최대 라운드를 도출해보았다.

본 논문의 제안기법은 비트순열의 약한 확산효과로 인하여 발생하는 다중선형특성을 구성하여 5-라운드 PRESENT로 암호화된 암호문으로부터 평문 8-비트 복구공격에 성공하였다.

본 논문의 구성은 다음과 같다. 2장에서는 PRESENT에 대한 간단한 소개와 선형특성을 활용한 평문복구공격 방안을 제시하였으며, 암호분석에 딥러닝 적용 방법들을 소개하였으며, 3장에서는 본 논문의 제안방법 및 해당 기법을 PRESENT에 적용한 방법을 서술하였다. 4장에서는 실험의 결과를 분석하고 마지막 장에서는 결론을 제시하며 마무리하였다.

2. 배경지식

2.1 딥러닝을 활용한 암호분석

다양한 암호분석기법 중 통계기반 분석기법으로 차분분석이나 선형분석이 가장 널리 알려져있다. 이러한 방법들은 암호가 가지고 있는 통계특성 확률을 분석하는 것으로부터 시작된다.

* 본 연구는 고려대 암호기술 특화연구센터(UD170109ED)를 통한 방위사업청과 국방과학연구소의 연구비 지원으로 수행되었습니다.

그러나, n-비트 블록크기를 갖는 블록암호에 대한 특성 확률을 정확히 구하기 위해서는 2^n개의 전체 입·출력 쌍을 요구할뿐만 아니라, $O(2^{2n})$의 시간 복잡도가 요구된다. 대다수 블록암호의 경우 64-비트 이상의 블록크기를 가지고 있어 분석에 천문학적 시간이 필요하다. 이에, 암호를 구성하고 있는 S-box와 같은 작은 구성요소들로 테이블을 구성하고 연결한 경로를 통해 전체 암호의 특성을 설명하는 방법이 주를 이룬다. 그러나, 현재 암호 분석에서 사용되는 통계 특성은 해당 특성의 추정된 모수의 값만 활용하는 경우가 대부분으로 암호의 특성을 완전히 파악하는데 한계가 있다. 딥러닝 기술은 이러한 암호의 통계적 특성 파악을 향상시켜줄 가능성을 보이고 있으며, 지속적으로 연구되고 있다. 최근에는 Crytpo2019에서 Aron Gohr가 SPECK 32/64에 대한 기존 차분 구별자를 딥러닝을 활용하여 향상시킨 결과를 제시하였다[6].

2.2 선형 특성을 활용한 평문 복원 방법

선형 특성은 선형공격에서 사용하는 마스터키 K에 따라 정해지는 암호화 함수 E_K의 통계적 특성으로 평문P, 각 라운드 키k_i 그리고 암호문$C = E_K(P)$ 사이의 연관성을 나타내는 선형근사식을 통해 구성한다. 각 라운드키의 선형근사식과 연관된 비트값들의 XOR합을 ck라고 하고 각 평문과 암호문에 대한 마스킹값을 m_P, m_C라고 할 때, 선형근사식은 다음의 형태로 나타낼 수 있다.

$$m_P \cdot P \oplus m_C \cdot C \oplus ck = 0 \quad \cdots \quad (식\text{-}1)$$

선형공격에서는 식-1의 확률 $\Pr(m_P, m_C;K)$의 편향성(bias) $\left|\Pr(m_P, m_C;K) - 1/2\right|$에 따라서 공격에 필요한 데이터 복잡도가 결정된다. 이에 따라, 확률의 평균값 $\Pr(m_P, m_C)$을 선형경로 및 파일링업(Piling-up) 정리 기반으로 도출하여 추정하는 것이 선형공격에 있어서 우선적으로 해결되어야 하는 과정이다.

이러한 통계 특성이 주어져 있는 경우 주어진 평문·암호문쌍(p_i, c_i) 으로부터 ck의 정보를 높은 확률로 복원해 낼 수 있으며, 해당 알고리즘은 1993년에 M.Matsui가 제안한 **Algorithm 1**으로 잘 알려져 있다. 본 논문에서는 (p_i, c_i)에서 ck를 복원한 뒤에 평문 $p_{target} \notin \{p_i\}$의 부분 정보가 암호문 c_{target}로 부터 노출됨을 보인다.

일반성을 잃지 않고, $\Pr(m_P, m_C) > 1/2$인 선형근사식을 활용하여 **Algorithm 1**로부터 $ck = 0$임을 알아내고, $m_C \cdot c_{target} = 0$ 임을 가정한다. 식-1에 주어진 값을 대입하면 확률이 $\Pr(m_P, m_C)$로 같은 식-2를 얻을 수 있다.

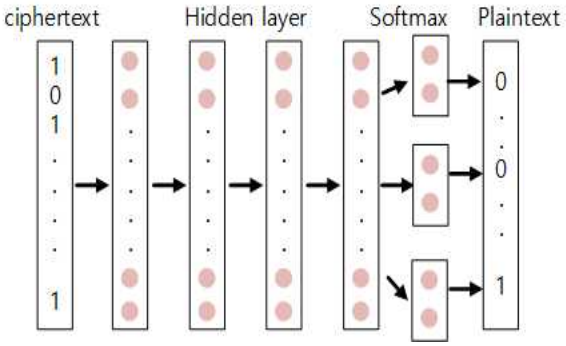
$$m_P \cdot P \oplus 0 \oplus 0 = 0 \quad \cdots \quad (식\text{-}2)$$

이는 주어진 정보$((p_i, c_i),\ c_{target})$가 알려지지 않은 평문 p_{target}의 1-비트 정보값$(m_P \cdot p_{target})$의 불확실성을 떨어트림을 알아낼 수 있고, 궁극적으로 해당 1-비트 정보 값의 결정에 영향을 준다. 더욱이, 식-1에서의 (m_P, m_C)를 변경하면 다수의 선형근사식을 구성할 수 있으며, 주어진 평문·암호문 데이터를 재활용하여 p_{target}의 다른 비트 복원 및 그 정확도를 향상시킬 수 있다.

2.3 Xiao 등의 평문 복원 방법

Xiao 등은 암호화 함수의 입출력에 대한 연관성을 딥러닝을 활용하여 파악하고 주어진 암호문에 해당하는 평문을 복원하는 방법을 제안하였다. 그림 1은 해당 방법을 도식화한 것으로 각 Softmax Layer들이 평문의 각 비트에 대한 출력을 도출하는 것을 확인할 수 있다.

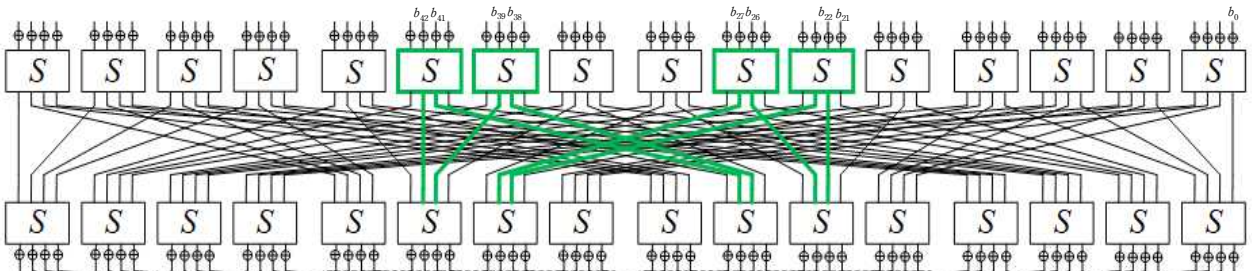
이러한 방법을 선형 특성을 활용한 평문 복원 방법으로 해석해 보면, m_P가 단위 벡터(standard unit vector, e_i)인 경우만 고려했다고 볼 수 있다. 이는, 식-1과 같은 선형특성을 고려하지 않은 기법으로 이를 개선한다면 향상된 결과를 얻을 수 있을 것으로 분석할 수 있다.



(그림 1) Ya Xiao의 DES 평문 복원

2.4 PRESENT

PRESENT는 Bogdanov 등이 2007년도에 발표한 SPN구조의 국제표준 경량블록암호 알고리즘이다. 64-비트의 블록사이즈와 80-비트와 128-비트의 키 사이즈를 사용하고 있으며, 16개의 4-비트 S-box와 64-비트 비트순열로 구성되어 있다. PRESENT는 총 31라운드로 구성되어 있으며, 하드웨어구현에 친



(그림 2) PRESENT Cipher의 약한 확산 효과

화적으로 설계되어 있다.

2.5 PRESENT의 약한 확산효과

그림 2에서 보는 것처럼 PRESENT는 S-box와 비트순열을 반복하며 이루어진다. 이 때, PRESENT의 5, 6, 9, 10번째 S-box에서 출력으로 나오는 초록색 선의 8개 비트는 서로 대칭을 이루면서 라운드가 반복되어도 5, 6, 9, 10번째 S-box에 입력된다. 따라서, 다수의 라운드를 거쳐도 동일위치의 S-box 출력이 동일입력에만 영향을 미쳐 확산 효과는 S-box에 의존해서 이루어지며 데이터 간에 높은 선형 연관성을 가지게 된다. B. Collard가 2009년에 제안한 Statistical Saturation Attack은 이러한 약한 확산효과를 활용하여 PRESENT를 공격하였다[7].

3. 딥러닝을 활용한 평문복원 설계

3.1 선형특성을 고려한 m_P 선택 방법

2.2에서 언급하였듯이, 하나의 선형근사식은 평문 p_{target}의 1-비트 정보값($m_P \cdot p_{target}$)결정에 해당 근사식의 편향성 크기만큼 영향을 준다. 이 평문 복구 방법의 확장 방법은 다수의 선형근사식들을 활용하는 것으로, 분석 암호에 m_P가 각기 다른 둘 이상의 근사식들을 사용할 경우 평문 2비트 이상의 값에 대한 연관성을 한번에 분석할 수 있다.

이에 선형 특성을 고려한 평문 복구 공격에 사용되는 m_p의 집합($\overline{M_P}$)은 다음과 같은 방법을 통해 선택할 수 있다.

1) 높은 편향성을 갖는 선형근사식들의 m_P집합 $M_P = \{m_P^0, m_P^1, \dots, m_P^{l-1}\}$ 도출

2) M_P를 포함하는 가장 작은 선형공간 $\overline{M_P}$ 도출

3.2 PRESENT의 $\overline{M_P}$ 도출 및 해당 데이터 생성

PRESENT의 $\overline{M_P}$ 도출을 위해, 우리는 2.5에서 제시한 약한 확산 효과로부터 얻을 수 있는 PRESENT의 선형근사식들을 고려하였다. 편향성 크기는 선형근사식이 거치는 각 라운드 S-box 개수와 반비례하는데, 약한 확산 효과로 인해 높은 편향성을 기대할 수 있는 경로들이 존재하게 된다. 결과적으로 우리는 PRESENT에 입력되는 k번째 위치에 해당하는 비트를 b_k라고 정의할 때, 5, 6, 9, 10번째 S-box에 입력되는 비트 중 $b_{21}, b_{22}, b_{25}, b_{26}, b_{37}, b_{38}, b_{41}, b_{42}$의 8-비트 정보를 고정시키고 나머지 56-비트를 랜덤하게 구성하여 평문 2^{15}개를 생성하였다. 이때, 고정한 8-비트의 값을 1씩 증가시키면서 구성하여 총 2^{23}개의 m_P를 생성하였다. 그리고 PRESENT 1라운드부터 1라운드씩 증가시키면서 대응하는 암호문을 구성하였다. 이후, 각 암호문을 딥러닝 모델의 입력 데이터로 제공하고 고정한 평문 비트의 값을 레이블로 구성한 후 딥러닝 모델이 256개 값 중 실제 고정한 값을 복원할 수 있는지에 대한 실험을 진행하였다. 트레이닝 데이터로는 전체 데이터의 약 70%인 5,88,800개를 사용하였으며, 검증 데이터로는 나머지 2,516,480개 사용하였다.

3.3 딥러닝 모델 구성

딥러닝 네트워크는 비트 사이의 상관관계를 통계적으로 분류하기에 적합한 완전연결계층(Fully Connected layer)를 사용하였다. 5개의 Dense layer를 구성하였으며, 과적합을 방지하기 위해서 3, 5번째 layer에서는 Dropout을 추가하여 주었다. 각 Dense layer에서는 sigmoid함수를, 마지막 Dense 층에서는 다중분류에 적합한 softmax 함수를 사용하였다.

딥러닝 프로그램은 케라스를 사용하였으며, 에포크는 100회, 배치 사이즈는 64로 설정하였다. 최적화 함수는 RMS prop을, 손실 함수는 'categorical crossentropy'를 사용하였다. 실험에 사용한 환경으로는 GeForce GTX 1080 Ti(6GB) GPU, Intel Core i7-9700 CPU와 32GB RAM을 탑재한 PC 1대를 사용하였다.

3.4 평문복원공격 성공판단 기준

딥러닝 모델의 성능을 암호학적으로 분석하기 위

하여 가능한 전체 비트값 갯수인 256중 실제 고정한 값을 딥러닝 모델이 정확히 추측할 확률로 1/256인 0.39%를 기준점으로 삼았다. 또한, 정확도를 판단하는 기준점은 검증 데이터에 대한 평가 정확도로 확인하였다.

4. 실험결과

학습과정에서는 별도의 과적합은 일어나지 않고 원활하게 학습되었다. 라운드 별로 검증 데이터의 정확도를 측정한 결과는 표 1과 같다. 표 1에서 보는 것처럼 적은 라운드 수에서는 암호문을 통해 평문의 8-비트를 높은 확률로 복원가능하였다.

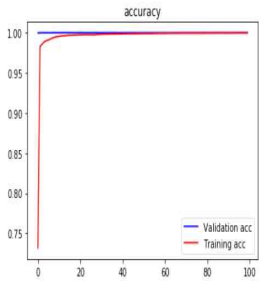
(accuracy(%))

라운드 수	1	2	3	4	5	6	7
정확도	100	100	96.3	1.1	1.1	0.39	0.39

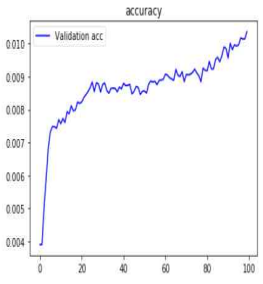
(표 1) PRESENT 라운드별 평문복원 정확도

그림 3에서 보는 것처럼 적은 에포크에서도 학습이 높은 정확도로 이루어졌는데 이는 PRESENT의 약한 확산효과로 인하여 고정한 비트값 들이 S-box만으로는 충분히 확산되지 않기 때문에 복원이 가능한 것으로 판단하였다. 특히, 선형 특성의 값이 라운드를 거치면서 줄어드는 것을 고려시, 딥러닝 모델이 추가적인 정보를 획득해서 평문을 복원하는 것으로 판단할 수 있다.

4~5라운드 PRESENT의 평문복원 결과를 살펴보면 랜덤하게 결과 값을 선택하는 것 보다 약 3배정도의 좋은 결과를 얻는 것을 볼 수 있었으며, 6라운드 부터는 실제 랜덤한 선택 결과와 거의 동일한 결과값을 보이고 있어 딥러닝 모델이 실질적인 구별을 하지 못하는 것을 알 수 있었다.



(그림 3) 2-라운드 PRESENT 평문복구공격 결과



(그림 4) 4-라운드 PRESENT 평문복구공격 결과

따라서, 실험에 사용한 딥러닝 구별자는 5라운드까지의 PRESENT에 대해서 평문을 복원할 수 있음을 알 수 있었다.

5. 결론

본 논문의 실험을 통하여 우리가 구성한 신경망 모델이 PRESENT의 평문을 일부 복원할 수 있음을 알 수 있었다. 실제 암호분석을 할 때에는 PRESENT의 세부적인 정보를 참고하면서 많은 계산을 통해서 키를 복원하고 옳은 키를 찾을 경우 평문을 복원할 수 있다. 그러나, 딥러닝 모델은 암호문과 평문에 대한 정보만으로도 PRESENT의 평문을 복원할 수 있음을 보여주었다. 이는, 암호분석에 있어 하나의 도움이 될 수 있는 도구로 딥러닝 모델이 활용될 수 있음을 보여준다.

또한, 여러 개의 약한 비트순열을 통해서 다차원을 구성할 경우 더 많은 비트의 평문을 복원도 가능할 것으로 판단된다.

차후에는 데이터의 양이나 딥러닝 네트워크 구조들에 따라서 더 많은 라운드의 PRESENT나 더 많은 평문을 복원할 수 있는지에 대한 연구를 진행해 볼 예정이다.

참고문헌

[1] Shannon, Claude E. "Communication theory of secrecy systems." The Bell system technical journal 28.4 (1949): 656-715.

[2] Xiao, Ya, Qingying Hao, and Danfeng Daphne Yao. "Neural cryptanalysis: Metrics, methodology, and applications in cps ciphers." 2019 IEEE Conference on Dependable and Secure Computing (DSC). IEEE, 2019.

[3] Matsui, Mitsuru. "Linear cryptanalysis method for DES cipher." Workshop on the Theory and Application of of Cryptographic Techniques. Springer, Berlin, Heidelberg, 1993.

[4] Hermelin, Miia, Joo Yeon Cho, and Kaisa Nyberg. "Multidimensional linear cryptanalysis of reduced round Serpent." Australasian Conference on Information Security and Privacy. Springer, Berlin, Heidelberg, 2008.

[5] Bogdanov, Andrey, et al. "PRESENT: An ultra-lightweight block cipher." International workshop on cryptographic hardware and embedded systems. Springer, Berlin, Heidelberg, 2007.

[6] Collard, Baudoin, and F-X. Standaert. "A statistical saturation attack against the block cipher PRESENT." Cryptographers' Track at the RSA Conference. Springer, Berlin, Heidelberg, 2009.

[7] Gohr, Aron. "Improving Attacks on Round-Reduced Speck32/64 Using Deep Learning." Annual International Cryptology Conference. Springer, Cham, 2019.