인공신경망 기반의 암호 분석 연구 동향

김현지[1], 강예준[1], 임세진[1], 김원웅[1], 서화정[1]
[1]한성대학교 IT융합공학과
khj1594012@gmail.com, etus1211@gmail.com, dlatpwls834@gmail.com,
dnjsdndeee@gmail.com, hwajeong84@gmail.com

Trends in Artificial Neural Network-based Cryptanalysis Technology

Hyun-Ji Kim[1], Yea-Jun Kang[1], Se-Jin Lim[1], Won-Woong Kim[1], Hwa-Jeong Seo[1]
[1]Dept. of IT Convergence Engineering, Hansung University

요 약

안전한 암호 시스템은 평문을 복원하거나 키를 유추해낼 수 없도록 설계된다. 암호 분석은 이러한 암호 시스템에서 평문과 키를 추정하는 것이며, 알려진 평문 공격, 선택 평문 공격, 차분분석 등 다양한 방법이 존재한다. 또한, 최근에는 데이터의 특징을 추출하고 학습해내는 인공신경망 기술을 기반으로 하는 암호 분석 기법들이 제안되고 있다. 현재는 라운드 축소된 S-DES, SPECK, SIMON, PRESENT 등의 경량암호 및 고전암호에 대한 공격이 대부분이며, 이외에도 암호 분석을 위한 active S-box의 수를 예측하는 등과 같이 다양한 측면에서 인공신경망이 적용되고 있다. 향후에는 신경망의 효율적 구현, full-round에 대한 공격과 그에 대한 암호학적 해석이 가능한 연구들이 진행되어야 할 것으로 생각된다.

1. 서론

암호 분석(Cryptanalysis)은 암호 알고리즘의 평문 또는 암호 키를 유추해내는 것이며, 안전한 암호시스템은 혼돈 및 확산을 통해 평문 복원하거나 사용된 키를 알아낼 수 없도록 설계된다. 암호 분석에는 차분 분석 (differential cryptanalysis), 선형 분석 (linear cryptanalysis) 등의 방법이 존재한다. 최근에는 인공지능 기술이 발달하면서 인공신경망을 기반으로 한 암호 분석에 관한 연구들이 활발히 진행되고 있다.

2. 관련 연구

2.1 인공신경망 (Artificial Neural Network)

인공 신경망은 생물학의 신경망에서 영감을 받은 학습 알고리즘이다. 신경망은 여러 개의 노드로 이루어진 레이어가 여러 층 쌓인 형태로 구성된다. 각 레이어에 존재하는 노드들은 자신과 연결된 이전 레이어의 노드 값과 가중치를 이용하여 가중합 연산을 수행하고, 비선형 함수인 활성화 함수를 통해 단일 값으로 계산된다. 이러한 방식으로 입력 데이터에 대해 모든 레이어를 적용한 후 실제 값과 예측 값의 차이인 손실 값을 계산한다. 그런 다음 역전파 과정을 통해 손실을 최소화하는 방향으로 신경망 내부의 가중치를 갱신한다. 이러한 과정을 반복하여 훈련되지 않은 데이터에 대한 일반화 성능을 보장하는 신경망을 구축할 수 있다. 실제 추론 시에는 학습된 신경망의 가중치가 고정된 상태로 입력 데이터에 대한 추론을 진행한다. 이를 통해 입력 데이터(이미지, 시계열, 언어, 그래프 등)의 특징을 추출하여 분류 및 회귀 등의 작업이 가능하다.

2.2 암호 분석 (Cryptanalysis)

암호 분석은 암호 알고리즘에 대한 평문 및 키 값을 유추해내는 공격 기법이다. 암호문 단독 공격, 선택 평문 공격, 선택 암호문 공격, 알려진 평문 공격, 차분 분석, 부채널 분석 등 다양한 방식이 존재한다. 암호문 단독 공격은 공격자가 암호문만 가지고 있는 상태에서 암호문의 통계적 특성이나 전수 조사를 통해 복호화하는 기법이다. 선택 평문 공격은 공격자가 많은 수의 무작위 평문을 암호문으로 만들 수 있는 상태에서 공격을 수행한다. 즉, 임의의 평문에 대한 암호문을 생성하고 비교해가며 키를 유추해낼 수 있다. 알려진 평문 공격은 이미 알고 있는 다수의 평문

과 암호문 쌍을 활용하여 키 값을 유추하는 암호분석 기법이다. 차분 분석은 선택된 평문 공격에서 사용할 수 있는 공격 기법이다. 암호 알고리즘의 경우, 평문을 더 작은 단위로 쪼갠 후, 치환(Substitution)과 전치(Permutation)을 반복적으로 적용하여 안전하게 설계한다. 치환 과정의 경우 비선형적이므로 전수조사하지 않는 한 키 값을 알아내기 어려웠으나, 해당 과정이 충분히 안전하지 않게 설계되었을 경우 차분이 유지된다는 성질을 이용한 차분 분석을 통해 효과적으로 키를 유추할 수 있게 되었다. Higher-order differential cryptanalysis, Truncated differential cryptanalysis, Impossible differential cryptanalysis 등과 같이 변형 기법이 존재한다.

3. 인공신경망 기반의 암호 분석 연구 동향

인공신경망 기반의 암호 분석은 고전 암호, S-DES, SPECK, SIMON 등에 대한 알려진 평문 공격, 암호문 단독 공격, 차분 공격 등이 주로 수행되었다.

3.1 암호문 단독 공격 (Ciphertext Only Attack)

[1]에서는 대체암호(substitution cipher)에 대한 최초의 암호문 단독 공격(Ciphertext Only Attack, COA)을 수행하였다. 암호문 단독 공격은 공격자가 암호문만을 수집하여 암호문에 대한 통계적 성질과 문장의 특성 등을 추정함으로써 그에 대한 평문과 키를 알아내는 암호 분석 기법이다. 해당 연구에서는 간단한 통계적 특성(암호문의 상대적 빈도)을 입력하여 신경망을 훈련하고, 이를 통해 키를 예측할 수 있도록 하였다. 또한, 암호문의 빈도수를 나타내는 히스토그램은 평문의 히스토그램을 평행이동 시킨 것과 동일하다. 신경망은 이러한 히스토그램의 상대적 이동을 인식하도록 훈련되었으므로 무차별 대입을 필요로 하지 않고 키를 찾아낼 수 있다.

3.2 알려진 평문 공격 (Known Plaintext Attack)

알려진 평문 공격을 통해 S-DES, Speck, Simon에 대한 암호 분석 기법이 연구되었다[2]. 해당 기법은 평문과 암호문 쌍을 비트로 표현한 후 연접하고, 신경망에 입력하면 그에 대응하는 키를 예측한 후 실제 키와 비교하여 손실을 계산한다. 이와 같은 방법으로 평문 및 암호문 쌍에 해당하는 키 값을 학습하도록 하여 추론 시에는 평문 및 암호문 쌍만을 가지고 키를 예측할 수 있도록 하였다. S-DES 및 Speck, Simon의 학습과 검증에 사용된 평문, 암호문 쌍의 개수는 각각 5만 개, 1만 개 그리고 50만 개, 100만 개이다.

또한, 해당 연구에서는 비트키와 텍스트키로 나누어 실험하였으며, 비트키는 모든 비트가 발생할 확률이 동일하며 키 공간이 이고 텍스트키는 아스키 코드 중 64개만을 사용하고 모든 비트가 발생할 확률이 동일하지 않다, 해당 기법의 성능 평가 시에는 수만 개의 데이터에 대해 학습하면 키 비트 예측 확률과 해당 비트의 발생 확률이 비슷해질 것이므로, 두 값 간의 차이를 계산하여 해당 값이 양수이면 암호 분석이 가능한 것으로 판단하였다. S-DES의 경우 비트키와 텍스트키 모두에 대해 암호 분석이 가능하였다. 비트키와 텍스트키에서 공통적으로 5번째, 8번째 비트가 공격에 취약하였으며, 6번째 비트는 안전하였다. Speck과 Simon은 비트키를 사용한 경우는 예측확률이 평균 0.5를 달성하여 암호 분석에 실패하였으며 텍스트 키에 대해서는 성공하였다.

[3]에서는 양자 머신러닝 중 Quantum Support Vector Machine(QSVM)을 활용하여 고전암호 Caesar에 대한 알려진 평문 공격을 수행하였다. QSVM은 기존의 SVM의 피처맵을 양자 컴퓨터 상에서 동작하는 양자회로로 설계한 것이며, SVM과 동일하게 초평면을 통해 데이터 포인트 간의 최적 경계를 찾는 머신러닝 기법이다. 양자 회로를 동작시키면 입력 데이터를 양자 상태로 인코딩한 후, 양자 회로를 구성하는 게이트들을 거치고 측정을 통해 하나의 값으로 결정된다. 해당 값을 기반으로 손실을 계산한 후, 게이트의 매개변수 값을 신경망의 가중치와 같이 변화시켜가며 손실을 최소화 하도록 학습한다. QSVM은 고차원의 데이터 작업에 유리하기 때문에 SVM이 처리하기 어려운 커널 최적화의 이점이 있으며, 일반적으로 기존 SVM보다 성능이 좋다. 해당 연구에서는 양자 회로 구성에 필요한 큐비트 등의 자원 및 메모리 문제로 2비트 및 3비트 평문 및 암호문 쌍과 키에 대해 실험하였다. 양자 시뮬레이터로 예측한 결과, 2비트 평문 및 암호문에 대해서는 큐비트 측정 횟수인 shots이 5일 때 1.00의 정확도를 달성하였으며 3비트 데이터셋에 대해서는 shots이 150일 때 0.84의 정확도를 달성하였다. 또한 2비트 평문 및 암호문에 대해 5 큐비트를 제공하는 실제 양자 하드웨어에서 회로를 동작시킨 결과, 0.93의 정확도를 얻었으며 이는 실제 양자 프로세서에서 발생하는 노이즈 등으로 인해 시뮬레이터의 정확도보다 0.07 감소하였

음을 알 수 있다.

3.3 차분 분석 (Differential Cryptanalysis)

CRYPTO 2019에서는 처음으로 round reduced SPECK에 대한 딥러닝 기반 차분 분석 기법이 제안되었다[4]. 해당 연구는 11라운드의 SPECK32/64에 대해 기존 방식의 차분 구별자를 뛰어 넘는 성능을 달성하였으며, 이를 통해 인공신경망 기술이 차분분석에서도 유의미하게 적용될 수 있음을 보였다. 제안 기법은 round reduced SPECK의 차분 속성을 활용하기 위해, 주어진 입력 차분에 대한 암호문 쌍과 무작위 쌍을 구분하도록 학습시킨다. 즉, 추측 키를 사용하여 복호화 한 결과가 실제 암호문인지 또는 랜덤하게 생성된 값인지 판별하여, 암호문인 경우 해당 키를 옳은 키로 판단하는 방식이다. 이를 신경망 구별자라고 하며, 학습에 사용된 데이터 셋은 훈련 데이터 900만 개와 검증데이터 100만 개이다. 신경망 구별자를 5, 6, 7, 8라운드의 SPECK32/64에 대해 학습시킨 결과, 모든 경우에서 고전적인 차분 구별자보다 약 0.01 이상씩 더 높은 정확도를 보였다. 또한, 7라운드 신경망 구별자에 2라운드 차분특성(0x4000/0x0000)을 추가하여 9라운드 구별자로 만든 후, 입력 평문 쌍이 0x4000/0x0000 차분특성으로 전파되도록 하여 1라운드를 확장시킨다. 즉, 10라운드 암호문에 대해 무작위 여부를 판별할 수 있게 된다. 그러므로 11라운드의 암호문 쌍을 추측 키를 사용하여 1라운드를 복호화 한 후, 해당 암호문쌍을 10라운드 신경망 구별자에 입력하여 얻는 결과 값이 특정 임계값보다 클 경우 옳은 키 후보로 판단한다. 이러한 방식으로 키 복구 공격을 수행한 결과, 100번 중 마지막 2라운드에 대한 서브키는 81번 복구하였고 마지막 라운드에 대한 서브키는 99번 복구에 성공하였다. 그러나 해당 연구 결과는 블랙박스 모델인 인공 신경망의 해석 가능성에 대한 문제를 제기하였으며, Eurocrypt2021에서는 [4]에 대한 해석이 발표되었다[5]. 이처럼 인공신경망 기반의 차분 분석 기법은 [4]를 기반으로 하여 SPECK, SIMON, PRESENT 등의 암호에 대해 진행되고 있다[6, 7, 8].

[6]는 [4]의 연구 결과로부터 영감을 받아 비 마르코프 암호 및 일체형 차분(동일한 입력 차이에서 모든 출력 차이를 고려)에 대한 인공신경망 기반의 암호 분석 기법을 제안하였다. 분석 대상 암호 알고리즘은 GIMLI, ASCON, KNOT 그리고 CHASKEY이며, 모두 round-reduced 버전이다.

[8]은 [6]의 차분 구별자를 기반으로 3~6라운드의 PRESENT에 대한 차분 분석 기법이다. 해당 연구에서는 무작위 차분 대신 [9]에서 제안된 차분을 사용하여 더 나은 결과를 얻었다. 해당 공격 기법은 5라운드까지 적용 가능하며, full-round PRESENT에 대한 공격은 불가능하다는 한계점이 존재한다.

이처럼 현재 수행된 대부분의 인공신경망 기반의 차분 분석 연구는 경량암호 및 round-reduced에 해당하므로, 향후에는 full-round에 대한 공격과 그에 대한 암호학적 해석이 가능한 연구가 진행되어야 할 것으로 생각된다.

3.4 Active S-box 수 예측

[10]에서는 GFS(generalized feistel block) 블록 암호의 active S-box 수를 예측하는 머신러닝 기법을 제안하였다. 해당 연구에서는 Active S-box의 수가 최소인 differential trail을 식별하는 것이 암호 분석의 목적이므로 이를 위해 active S-box의 수를 예측하였다. 라운드 수(r) 및 순열 패턴과 같은 블록 암호의 공통된 특징과, truncated differences(입력 및 출력 차분) 같은 차등 암호 분석 관련 기능을 사용하여 fully-connected 신경망을 훈련하여 활성화되는 S-box의 수를 예측한다. 입력 및 출력 차분, 라운드 수, 순열 패턴을 데이터 특징으로 사용하였으며, 각각의 특징들을 제거하며 예측 성능을 평가하였을 때 순열 패턴이 active S-box 예측에 가장 중요한 요소임을 알 수 있다. 또한, RMSE, R-squared(R^2)를 사용하여 신경망의 성능을 평가하였으며 전반적으로 RMSE 값이 1.96 이하로 active S-box의 수를 예측할 수 있다.

4. 결론

평문 및 키 값을 유추해내는 암호 분석에는 알려진 평문 공격, 선택된 평문 공격, 암호문 단독 공격, 차분 공격 등이 있다. 최근에는 고전적인 암호 분석 뿐만 아니라 인공신경망 기술을 기반으로 하는 암호 분석 기법들이 제안되고 있다. 현재는 고전암호, S-DES, SPECK, SIMON, PRESENT 등의 간단한 경량 암호에 대한 분석이 가능하며, 차분 분석 연구 결과에서 알 수 있듯이 round-reduced 형태의 암호를 대상으로 하고 있다. 암호 분석에는 Linear 레이어 또는 Convolution 레이어 등이 사용되며, Resnet과 같은 잔차 네트워크 구조가 사용되었다. 또한, 암호 분석의 성능을 개선하는 연구뿐만 아니라 비슷한 성

능을 보장하면서 인공 신경망의 구조를 최소화 하는 연구들도 진행되고 있다. 향후에는 인공 신경망의 결과를 암호학적 관점에서 해석하기 위한 설명 가능한 인공지능(eXplainable Artificial Intelligence, XAI) 기술의 적용과 경량 암호가 아닌 경우 및 전체 라운드에 대한 공격이 수행되어야 할 것으로 생각된다. 또한, Active S-box의 수를 예측하는 연구와 같이 다양한 측면에서 암호 분석을 위한 인공신경망의 적용이 가능할 것으로 보인다.

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] Focardi, Riccardo, and Flaminia L. Luccio. "Neural Cryptanalysis of Classical Ciphers." ICTCS. 2018.

[2] So, Jaewoo. "Deep learning-based cryptanalysis of lightweight block ciphers." Security and Communication Networks 2020 (2020).

[3] Kim, Hyun-Ji, et al. "Cryptanalysis of Caesar using Quantum Support Vector Machine." 2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia). IEEE, 2021.

[4] A. Gohr, “Improving attacks on round-reduced speck32/64 using deep learning”. Annual International Cryptology Conference, pp. 150-179. Springer, Cham, 2019.

[5] A. Benamira, D. Gerault, T. Peyrin, and Q. Q. Tan, “A deeper look at machine learning-based cryptanalysis.” Annual International Conference on the Theory and Applications of Cryptographic Techniques, pp. 805-835. Springer, Cham, 2021.

[6] A. Baksi, J. Breier, X. Dong, and C. Yi, “Machine learning assisted differential distinguishers for lightweight ciphers.” IACR Cryptol. ePrint Arch., vol. 2020, p. 571, 2020.

[7] Hou, Zezhou, Jiongjiong Ren, and Shaozhen Chen. "Cryptanalysis of round-reduced simon32 based on deep learning." Cryptology ePrint Archive (2021).

[8] Jain, Aayush, Varun Kohli, and Girish Mishra. "Deep learning based differential distinguisher for lightweight cipher PRESENT." Cryptology ePrint Archive (2020).

[9] M. Wang, “Differential cryptanalysis of reduced-round present,” in International Conference on Cryptology in Africa. Springer, 2008, pp. 40 - 49.

[10] Idris, Mohamed Fadl, et al. "A deep learning approach for active S-box prediction of lightweight generalized feistel block ciphers." IEEE Access 9 (2021): 104205-104216.