FPE의 딥러닝 기반의 신경망 구별자

한성대학교 융합보안학과 23213703 김덕영

한성대학교 HANSUNG UNIVERSITY

CryptoCraft LAB

Contents

서론

- Motivation
  - 암호 알고리즘은 특정 확률로 특정 차분 특성을 가짐
  - 그중에서도 신용카드 번호와 같은 민감한 데이터를 암호화하는 형태보존암호는 취약 할 수 있음
  - 구별자 공격은 이러한 사실을 기반으로 랜덤 데이터들로부터 암호 데이터를 구별해내는 작업
  - 데이터에 대한 확률적인 예측을 수행하는 딥러닝 기술은 이에 대한 좋은 솔루션이 될 수 있음

- Contribution
  - 본 논문에서는 NIST 형태보존암호인 FF1, FF3-1에 대한 딥러닝 기반의 신경망 구별자를 최초로 제안
  - 랜덤과 하나의 차분을 비교하는 싱글 모델과 여러개의 차분을 비교하는 멀티 모델을 구현함
  - 숫자, 소문자 알파벳 도메인을 이용하여 여러 데이터와 라운드를 넣어 정확도를 실험함

관련연구 - 형태보존암호

- 기존의 암호화 기법과 다르게 데이터를 암호화할 때 내부 데이터의 내용을 숨기면서 외부에서는 데이터의 형태나 구조를 보존할 수 있도록 함
  → 평문과 암호문의 길이 및 형태가 동일 (17비트의 평문 → 17비트의 암호문)

- 종류
  - FF1, FF3, FPE, FEA ..

- 본 연구에서는 미국 NIST의 표준 형태보존 암호인 FF1, FF3-1를 사용

관련연구 - 차분 특성 및 신경망 구별자

- 차분
  - 서로 다른 평문과 암호문을 XOR한 값
    - 평문 $P_0 \oplus P_1 = P_2$, 암호문 $C_0 \oplus C_1 = C_2$

      → 차분 = (P_2, C_2)
- 차분 특성
  - 데이터의 변화를 측정하거나 분석하는 과정에서 사용되는 특성
- 차분 분석
  - 암호 분석 기법 중 하나로, 암호화 알고리즘 또는 키를 해독하기 위해 사용되는 기법

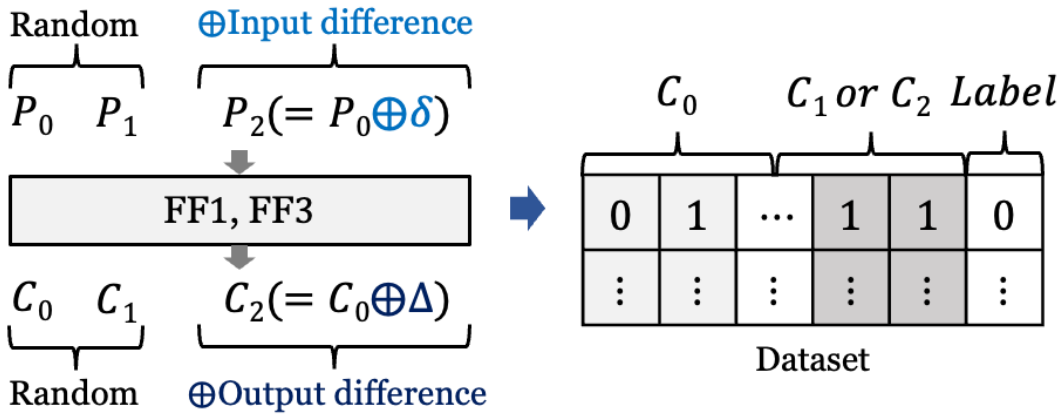
    → 차분 특성을 활용하여 암호화에 사용된 키 일부 정보를 도출하는 방법
- 딥러닝 기반의 신경망 구별자
  - 기존의 암호문에 나타나는 차분 특성을 신경망을 통해 학습
  - 차분을 갖는 암호문과 랜덤 데이터를 구별

    ➔ 차분 특성을 갖는 데이터 수집 → 딥러닝 모델 구축 → 학습 → 차분을 갖는 데이터 구별

제안기법 (ModelOne)

< FF1, FF3의 딥러닝 기반의 ModelOne 데이터 셋 생성 과정 >

- 우선 랜덤 평문 P_0, P_1 을 생성한다. 입력 차분을 만족하는 평문 쌍을 만들어야 하므로 P_0에 δ(입력 차분)을 XOR 하여 평문 P_2를 구함
- 그 후, 각 평문 P_0, P_1, P_2를 암호화하여 암호문 C_0, C_1, C_2 를 구함 여기서 C_0와C_1 은 차분 관계가 아닌 랜덤 평문을 암호화한 암호문으로, 두 값을 연접한 결과를 0으로 라벨링함
- C_0와 C_2 는 δ(입력 차분)을 만족하는 평문의 암호문이므로 특정 확률로 Δ(출력 차분)을 만족하는 암호 데이터이므로, 두 값을 연접한 값을 1으로 라벨링한다.

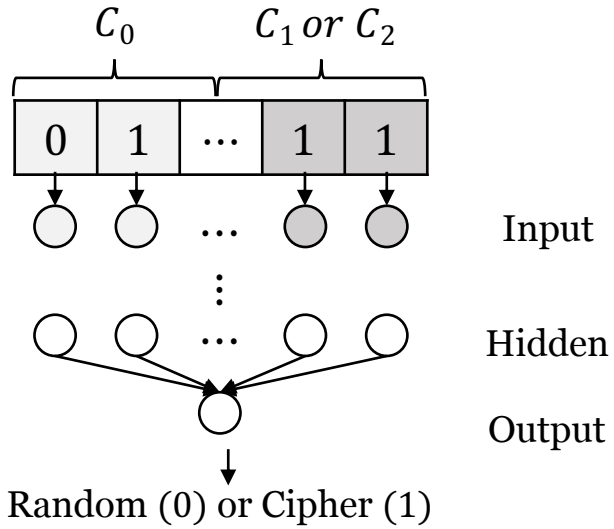


<ModelOne 차분 데이터 셋 >

제안기법 (ModelOne)

< FF1, FF3-1 딥러닝 기반의 ModelOne 구성 >

- 데이터 셋의 각 비트는 입력 레이어의 각 뉴런에 할당
  - 출력 레이어에 Sigmoid 활성화 함수 적용
    - 0~1 사이의 값으로 예측
- 출력 값과 실제 정답 (0 또는 1)간의 손실 계산

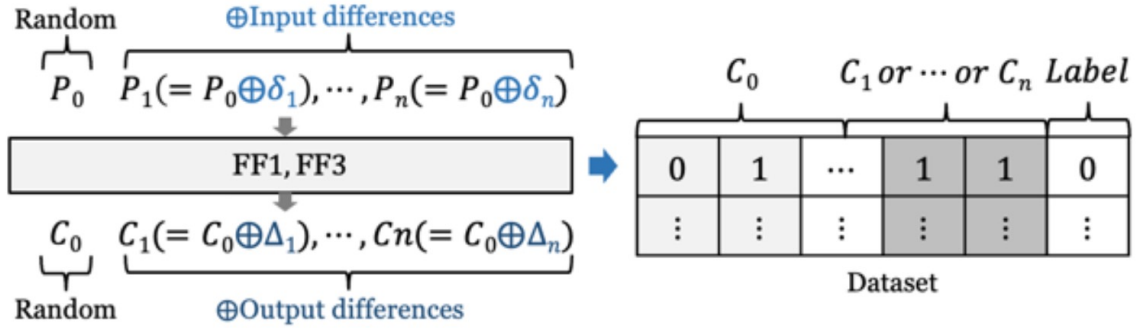


< ModelOne FF1, FF3-1 신경망 구별자의 구조 >

제안기법 (ModelMul)

< FF1, FF3-1의 딥러닝 기반의 ModelMul 데이터 셋 생성 과정 >

- ModelOne과 유사하게, 랜덤 평문 P_0을 생성 후 입력 차분 δ_n을 적용하여 평문 P_n을 생성한다.($P_n = P_0$ XOR δ_n)
- 그 후, 이를 암호화하여 암호문 C_n 을 생성한다. 암호문 C_0와 C_n 을 연결($C_0 \| C_n$)하여 학습 데이터로 사용한다.
- 이때, C_n이 δ_n에 해당하면 이를 클래스 n-1로 할당한다(예: δ_3 에 대응하는 C_3은 클래스 2로 분류).

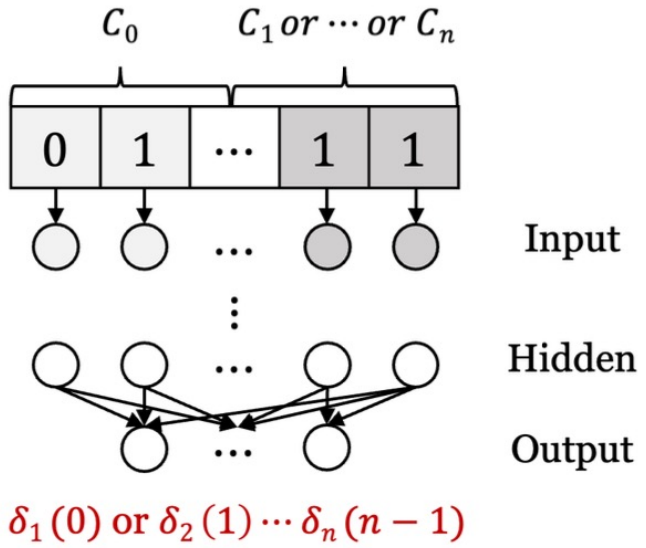


<ModelMul 차분 데이터 셋 >

제안기법 (ModelMul)

< FF1, FF3-1의 딥러닝 기반의 ModelMul 모델 구성 >

- 데이터 셋의 각 비트는 입력 레이어의 각 뉴런에 할당
- 출력 레이어에 Softmax 활성화 함수 적용하여 다중 클래스 확률을 예측
- 예측 값은 각 입력 차분에 대한 확률로 출력되며, 출력 노드는 클래스 개수(n)와 동일
- Categorical cross-entropy 손실 함수를 사용해 예측 값과 실제 정답 간의 손실을 계산



< ModelMul FF1, FF3 신경망 구별자의 구조 >

제안기법

< ModelOne, ModelMul 의 딥러닝 기반의 데이터셋 >

- 본 논문에서 데이터의 도메인을 숫자 (0 ~ 9), 소문자(a ~ z)로 나누어 사용하였으며 '0x0||K' (K는 0~F사이의 16진수)를 사용할 경우 더 높은 차분 확률을 가지므로 제시된 입력 차분들에 대한 실험을 진행

(내부 암호화 함수와 별개로 FF1, FF3-1에 독립적으로 0x0||K 사용 가능)

제안기법

< ModelOne, ModelMul 딥러닝 기반의 하이퍼파라미터 >

- ModelOne은 차분을 갖는 데이터와 랜덤 데이터를 구별해야하므로 이진분류 수행
  → "binary_crossentropy" 손실함수 사용
- ModelMul은 여러 차분을 만족하는 암호문 쌍을 분류하는 다중 클래스 분류 수행
  → "categorical cross-entropy" 손실 함수 사용
- 최적화 함수는 Adam을 사용
- 더욱 정교한 학습을 진행하기 위해 학습률을 0.001 ~ 0.0001까지 감소시킴

Model	***ModelOne***	***ModelMul***
Schemes	FF1/FF3	FF1/FF3
Epochs	20/15	20/15
Loss function	Binary cross-entropy	Categorical cross-entropy
Optimizer	Adam (0.001 to 0.0001, learning rate decay)	
Activation function	ReLu (hidden)	
	Softmax (output)	Sigomid (output)
Batch size	32	
Hidden layers	5/4 hidden layers (with 64/128 units)	
Parameters	173,956/74,497	173,956/75,787

< 신경망 구별자 모델의 하이퍼파라미터 >

실험 결과(ModelOne)

< FF1-1, FF3-1의 딥러닝 기반 ModelOne 구별자 실험결과 >

- Number) FF1 입력차분으로 0x0F을 사용하였을 경우, 최대 10 라운드까지 구별 가능하며 0.855의 높은 정확도 달성 FF3-1 입력차분으로 0x08을 사용하였을 경우, 최대 8라운드까지 구별 가능하며 0.987의 높은 정확도 달성
- Lowercase) 평문 및 암호문의 경우의 수가 2^{26}으로 증가함에 따라 최대 2라운드까지 구별 가능하며, FF1 입력 차분 0x09에 대해 0.522의 정확도를 달성 FF3-1 입력 차분 0x08에 대해 0.556의 정확도 달성

0x	Number (10-round)				Lowercase (2-round)			
	Training	Validation	Test	Reliability	Training	Validation	Test	Reliability
01	0.732	0.741	0.733	0.233	0.500	0.500	0.500	0.000
02	0.741	0.752	0.743	0.243	0.510	0.512	0.510	0.010
03	0.711	0.712	0.711	0.211	0.522	0.520	0.522	0.022
04	0.751	0.752	0.752	0.252	0.511	0.512	0.510	0.010
05	0.752	0.751	0.752	0.252	0.511	0.512	0.511	0.011
06	0.751	0.752	0.752	0.252	0.511	0.512	0.511	0.011
07	0.751	0.751	0.752	0.252	0.511	0.511	0.511	0.011
08	0.801	0.802	0.802	0.302	0.511	0.511	0.511	0.011
09	0.841	0.842	0.841	0.341	0.522	0.521	0.522	0.022
0A	0.842	0.841	0.841	0.341	0.500	0.510	0.510	0.010
0B	0.822	0.821	0.822	0.322	0.511	0.511	0.511	0.011
0C	0.855	0.854	0.855	0.355	0.500	0.500	0.500	0.000
0D	0.788	0.788	0.788	0.288	0.511	0.511	0.511	0.011
0E	0.811	0.812	0.811	0.311	0.522	0.521	0.522	0.022
0F	0.855	0.854	0.855	0.355	0.522	0.522	0.522	0.022

0x	Number (8-round)				Lowercase (2-round)			
	Training	Validation	Test	Reliability	Training	Validation	Test	Reliability
01	0.629	0.624	0.623	0.123	0.545	0.544	0.543	0.043
02	0.829	0.825	0.825	0.325	0.552	0.548	0.545	0.045
03	0.783	0.769	0.771	0.271	0.52	0.514	0.513	0.013
04	0.761	0.756	0.757	0.257	0.523	0.52	0.517	0.017
05	0.773	0.752	0.747	0.247	0.539	0.538	0.537	0.037
06	0.758	0.748	0.75	0.25	0.523	0.519	0.523	0.023
07	0.756	0.739	0.74	0.24	0.532	0.529	0.529	0.029
08	0.987	0.976	0.977	0.477	0.556	0.554	0.554	0.054
09	0.962	0.942	0.941	0.441	0.547	0.543	0.549	0.049
0A	0.969	0.953	0.951	0.451	0.538	0.534	0.532	0.032
0B	0.97	0.965	0.966	0.466	0.53	0.526	0.522	0.022
0C	0.97	0.959	0.959	0.459	0.538	0.536	0.539	0.039
0D	0.968	0.965	0.966	0.466	0.532	0.524	0.518	0.018
0E	0.964	0.963	0.963	0.463	0.549	0.549	0.551	0.051
0F	0.965	0.939	0.941	0.441	0.528	0.524	0.524	0.024

< FF1, FF3 ModelOne 데이터셋 >

실험 결과(ModelMul)

< ModelMul의 입력 차분 데이터 셋 정보 >

- ModelMul에서 입력 차분을 0x0||K로 설정하여, 각 데이터셋은 사용된 입력 차분 쌍에 따라 구성된다.(각 클래스에 약$2^{18.6097}$개 데이터 포함)
- 0x08이 가장 좋은 차분으로 간주되어 고정된 입력 차분으로 설정되며, 이를 기준으로 다른 입력 차분 데이터를 확장하여 데이터셋을 생성
- 예를 들어, 입력 차분 세 개를 사용하는 경우, 0.3333(= $\frac{1}{3}$) 이상의 정확도를 달성해야 해당 모델이 유효하다고 할 수 있다.

Dataset	Data Size	Input Difference Pair	Valid Accuracy
I1	$2^{18.6097}$ per class	01, 08	>0.500
I2		01, 02, 08	>0.333
I3		01∼03, 08	>0.250
I4		01∼04, 08	>0.200
I5		01∼05, 08	>0.166
I6		01∼06, 08	>0.142
I7		01∼08	>0.125
I8		01∼09	>0.111
I9		01∼0A	>0.100
I10		01∼0B	>0.090
I11		01∼0C	>0.083
I12		01∼0D	>0.076
I13		01∼0E	>0.071
I14		01∼0F	>0.066

< ModelMul의 입력 차분 데이터 셋 정보 >

실험 결과(ModelMul)

< FF1, FF3-1의 딥러닝 기반 ModelMul 구별자 실험결과 >

- Number, Lowercase 모든 도메인에서 정확도가 유효 정확도를 상회하였으며
- 특히 I2 데이터 셋을 활용했을 때 가장 높은 신뢰도를 보임

Dataset	Number (8 Rounds)				Lowercase (2 Rounds)			
	Training	Validation	Test	Reliability	Training	Validation	Test	Reliability
I1	0.520	0.520	0.520	0.020	0.520	0.520	0.520	0.020
I2	0.340	0.339	0.340	0.007	0.360	0.360	0.360	0.027
I3	0.260	0.260	0.260	0.010	0.270	0.270	0.270	0.020
I4	0.210	0.210	0.210	0.010	0.200	0.200	0.200	0.010
I5	0.170	0.170	0.170	0.004	0.180	0.180	0.180	0.004
I6	0.150	0.150	0.150	0.008	0.150	0.150	0.150	0.008
I7	0.130	0.130	0.130	0.005	0.130	0.130	0.130	0.005
I8	0.120	0.120	0.120	0.009	0.120	0.120	0.120	0.009
I9	0.120	0.110	0.120	0.020	0.100	0.100	0.110	0.010
I10	0.100	0.100	0.100	0.010	0.100	0.100	0.100	0.010
I11	0.090	0.090	0.090	0.007	0.090	0.090	0.090	0.007
I12	0.080	0.080	0.080	0.004	0.080	0.080	0.080	0.004
I13	0.080	0.080	0.080	0.009	0.080	0.080	0.080	0.009
I14	0.070	0.070	0.070	0.004	0.070	0.070	0.070	0.004

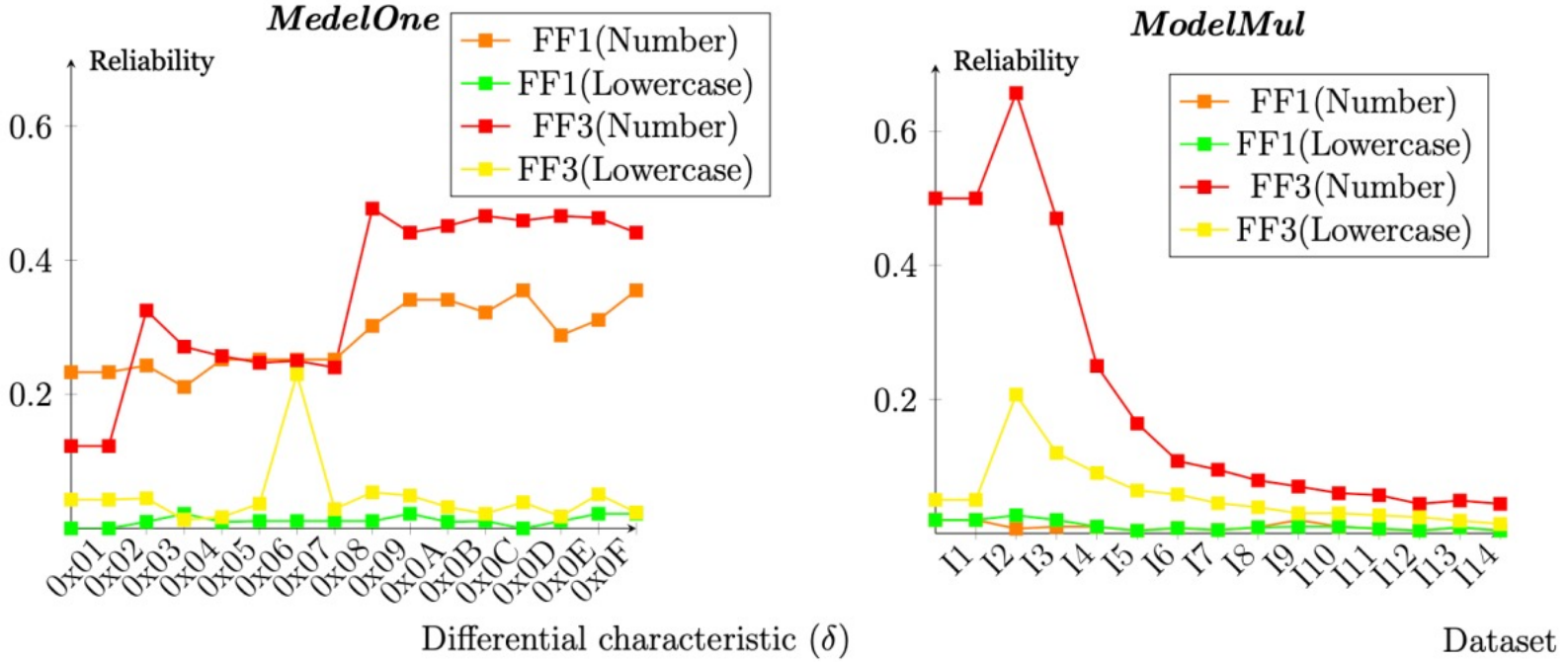
Dataset	Number (8 Rounds)				Lowercase (2 Rounds)			
	Training	Validation	Test	Reliability	Training	Validation	Test	Reliability
I1	1.00	1.00	1.00	0.500	0.55	0.55	0.55	0.050
I2	0.99	1.00	0.99	0.657	0.54	0.54	0.54	0.207
I3	0.72	0.72	0.72	0.470	0.38	0.37	0.37	0.120
I4	0.46	0.45	0.45	0.250	0.29	0.29	0.29	0.090
I5	0.33	0.33	0.33	0.164	0.24	0.23	0.23	0.064
I6	0.25	0.25	0.25	0.108	0.20	0.20	0.20	0.058
I7	0.22	0.22	0.22	0.095	0.17	0.17	0.17	0.045
I8	0.19	0.19	0.19	0.079	0.15	0.15	0.15	0.039
I9	0.17	0.17	0.17	0.070	0.13	0.13	0.13	0.030
I10	0.16	0.15	0.15	0.06	0.12	0.12	0.12	0.030
I11	0.14	0.14	0.14	0.057	0.11	0.11	0.11	0.027
I12	0.13	0.12	0.12	0.044	0.10	0.10	0.10	0.024
I13	0.12	0.11	0.12	0.049	0.09	0.09	0.09	0.019
I14	0.11	0.11	0.11	0.044	0.08	0.08	0.08	0.014

< FF1, FF3-1 ModelMul 데이터셋 >

실험 결과(데이터셋에 따른 신뢰도 표)

< 각 차분 특성과 데이터셋에 따른 신뢰도 표 >

- 아래 그림은 각 차분 특성과 데이터셋에 따른 신뢰도를 보여줌
- ModelOne) FF1에서 입력이 0x0F일 때 가장 높은 신뢰도를 FF3-1에서 0x08에서 가장 높은 신뢰도를 보임
- ModelMul) I2 데이터셋을 사용 할 경우 두 도메인 모두에서 가장 높은 신뢰도를 나타냄



<차분 특성과 데이터셋에 따른 신뢰도>

결론

< FPE(FF1, FF3)의 딥러닝 기반의 구별자 실험 최종 결론>

- 결론
  - 형태보존암호인 FF1, FF3에 대한 최초의 딥러닝 기반 신경망 구별자를 제안
  - ModelOne) FF1에 대해 입력차분으로 0x0F, FF3에 대해 0x08(입력차분)을 사용
    - 숫자 도메인에서 10, 8라운드까지 0.855 / 0.98 이상의 정확도를 달성
    - 소문자 도메인에서는 2라운드까지 0.522 / 0.554의 정확도로 구별 가능
  - ModelMul) FF1, FF3-1 모두에서 I2 데이터 셋을 사용
    - 두 도메인 모두에서 가장 높은 정확도 달성
- 향후 연구
  - 차분을 더 늘려서 다중 차분을 구별하는 모델 구현 예정
  - 더 큰 도메인 및 높은 라운드에서도 동작 가능한 FF1, FF3-1 신경망 구별자를 구현 예정

Q & A