Transformer를 이용한 LWE 기반 암호 공격 기법 분석

김영준, 신동준

한양대학교 융합전자공학과

june0888@hanyang.ac.kr, djshin@hanyang.ac.kr

Analysis of Attacks on LWE-based Cryptography using Transformers

Youngjun Kim, Dong-Joon Shin

Dept. of Electronic Engineering, Hanyang Univ.

요 약

LWE 문제는 가장 중요한 양자내성암호인 격자 기반 암호와 암호문 상에서 다양한 연산을 수행할 수 있는 완전동형암호 등을 생성하는 기반 난제로 사용된다. 따라서 이 암호들의 안전성을 검증하기 위해서는 LWE 문제를 풀 수 있는 알고리즘들에 대한 분석이 중요하다. 본 논문에서는 기존 제안된 Transformer를 이용하여 LWE 문제를 푸는 기법을 통해 LWE와 RLWE를 기반으로 하는 암호에 대한 안전성을 분석하고, LWE 기반 암호 공격에 필요한 샘플 수를 줄일 수 있는 새로운 기법을 제안한다.

Ⅰ. 서 론

양자컴퓨터의 발달로 인해 RSA와 ECC 등의 기존 암호체계가 위협받고 있는 상황에서 양자컴퓨터의 공격에도 안전하다고 알려진 다양한 양자내성암호(Post-Quantum Cryptography) 기법들이 제안되고 있다. 격자 기반 암호는 수학적으로 어려운 격자 문제인 LWE(Learning with Errors) 문제 등을 기반으로 생성되는 암호이며 양자내성을 가진다는 점에서 주목받고 있다. 또한, 암호화된 상태에서도 다양한 연산을 수행할 수 있어 정보를 더 안전하게 처리할 수 있는 동형암호에서도 기반 난제로써 주로 LWE 문제가 사용된다.

한편, n차 다항식의 쌍으로 표현되는 RLWE(Ring LWE) 문제에서 하나의 RLWE 샘플은 (구조화된 형태의) n개의 n차원 LWE 샘플로 변환되기 때문에 효율적 측면에서 LWE보다 RLWE 문제를 기반으로 암호를 설계하는 것이 선호된다. 이러한 구조가 있는 형태에도 불구하고 RLWE를 효과적으로 공격하는 방법에 대해서는 현재까지 잘 알려지지 않았다.

2022년 Lauter 등은 딥 러닝 모델인 Transformer를 사용한 SALSA 알고리즘을 통해 128 차원 이하의 LWE에 대한 공격에 성공했다는 실험 결과를 발표하였다[1]. 이로부터 SALSA와 같은 학습을 기반으로 한 딥 러닝 모델 공격에서는 RLWE가 LWE에 비해 하나의 RLWE 샘플이 n배만큼의 LWE 학습 샘플을 제공하는 효과를 낼 수 있다는 추측에 의해 RLWE가 더욱 취약하다고 예상하였다.

본 논문에서는 SALSA 알고리즘을 이용한 실험을 통해 LWE와 RLWE를 기반으로 하는 암호에 대한 공격 성능을 비교하고, SALSA 기반 공격에 필요한 LWE 샘플 수를 줄이는 새로운 기법을 제안한다.

Ⅱ. 본론

우선 LWE 문제[2]와 RLWE 문제에 대한 정의를 설명하고 SALSA 알고리즘을 소개한다.

2.1. LWE와 RLWE

주어진 비밀키 $\boldsymbol{s}\in Z_q^n$에 대해 LWE 분포 $A_{m,n,q,\chi}^{LWE}$는 $(\boldsymbol{a}_i, b_i = \boldsymbol{a}_i \cdot \boldsymbol{s} + e \bmod q)\,(i=1,2,...,m)$의 쌍들로 구성된다. 이때, 벡터 $\boldsymbol{a}_i$의 각 원소값은 Z_q에서 무작위로 추출된 정수이며 e는 에러 분포 χ로부터 추출된 정수이다. 본 논문에서는 LWE 분포 $A_{m,n,q,\chi}^{LWE}$의 한 LWE 쌍$(\boldsymbol{a}, b)$을 LWE 샘플로 정의하였다.

탐색 LWE(search-LWE) 문제는 $A_{m,n,q,\chi}(\boldsymbol{s})$에 의해 생성된 $(\boldsymbol{a}_i, b_i)$ $(i=1,2,...,m)$가 주어졌을 때, 비밀키 벡터 $\boldsymbol{s}$를 찾는 문제이다. 결정 LWE(decision-LWE) 문제는 $(\boldsymbol{a}_i, b_i)\,(i=1,2,...,m)$쌍들이 주어졌을 때, 이 쌍들이 $A_{m,n,q,\chi}(\boldsymbol{s})$로부터 생성된 샘플들인지, 무작위로 생성된 샘플들인지 결정하는 문제이다.

RLWE 문제는 다항식 환 $R_q = Z_q[x]/(f(x))$위에서 정의되는 LWE 문제로 비밀키를 나타내는 다항식 $s(x)$에 대해 RLWE 분포 $A_{n,q,\chi}^{RLWE}$는 $(a(x), b(x) = a(x)\cdot s(x) + e(x))$의 쌍들로 구성된다. 이 때, $a(x)$의 각 계수들은 Z_q에서 무작위로 추출된 정수이며 $e(x)$의 각 계수들은 에러 분포 χ로부터 추출된 정수들이다. 본 논문에서는 RLWE 분포의 한 RLWE 쌍을 RLWE 샘플 $(a(x), b(x))$로 정의하였다.

탐색 RLWE(search-RLWE) 문제는 $A_{n,q,\chi}^{RLWE}(s(x))$에 의해 생성된 $(a(x), b(x))$가 주어졌을 때, $s(x)$를 찾는 문제이다. 결정 RLWE(decision-RLWE) 문제는 $(a(x), b(x))$가 주어졌을 때, $A_{n,q,\chi}^{RLWE}(s(x))$에 의해 생성된 샘플인지 무작위로 생성된 샘플인지 결정하는 문제이다.

2.2. SALSA 알고리즘

본 논문의 실험은 기존 논문의 알고리즘인 SALSA 알고리즘[1]을 사용하였으며, 이 알고리즘은 LWE 샘플 $(\boldsymbol{a}, b)$의 $\boldsymbol{a}$를 입력값으로 받아 b를 예측하는 Transformer M에 대해 다음과 같은 순서에 의해 작동한다.

1. LWE 샘플 중 중복을 허용하여 무작위로 $n \times 10{,}000$ 개를 선택해 M을 학습시킨다. 모델학습은 b의 예측값 b'와 실제 b값 간의 크로스 엔트로피를 최소화하는 방향으로 이루어진다.
2. 학습된 M으로부터 비밀키 복원 알고리즘을 사용하여 비밀키의 예측값 $\boldsymbol{s}'$를 얻는다.
3. 비밀키 검증 알고리즘을 사용했을 때, $\boldsymbol{s} \neq \boldsymbol{s}'$라면 1단계로 돌아가서 반복하고, $\boldsymbol{s} = \boldsymbol{s}'$라면 알고리즘을 종료한다.

- 비밀키 복원 알고리즘

학습이 종료되면 아래의 두 가지 방법을 모두 사용하여 비밀키의 복원을 시도한다.

1) 직접 복원

$i=1,2,...,n$에 대해 $\boldsymbol{a}=K\boldsymbol{u}_i$($\boldsymbol{u}_i$는 i번째 원소값이 1인 단위 벡터, K는 임의의 큰 정수)로 설정한다. Transformer M의 학습이 잘 이루어졌을 때, $\boldsymbol{s}$의 i번째 원소값이 0이라면 $b'=M(\boldsymbol{a})=e$이므로 작은 값을 가지게 되고, i번째 원소값이 1이라면 $b'=M(\boldsymbol{a})=K+e$이므로 큰 값을 가지게 되므로, 이 연산의 모든 i에 대한 반복 수행을 통해 $\boldsymbol{s}$를 복원할 수 있다.

2) Distinguisher 복원

주어진 LWE 샘플 $(\boldsymbol{a}, b)$와 무작위 정수로 이루어진 쌍 $(\boldsymbol{a}_r, b_r)$에 대해 $\boldsymbol{a}$의 i번째 원소값에 임의의 정수 c를 더한 벡터를 $\boldsymbol{a}'$로 설정한다. Transformer M의 학습이 잘 이루어졌을 때, $\boldsymbol{s}$의 i번째 원소값이 0이라면 $M(\boldsymbol{a}')=\boldsymbol{a}'\cdot\boldsymbol{s}+e \bmod q=b$이므로 $M(\boldsymbol{a}')$가 $M(\boldsymbol{a}_r)$보다 b에 가까운 값을 가질 확률이 높다. 모든 i에 대해 이 방식의 반복 수행을 통해 $\boldsymbol{s}$를 복원할 수 있다.

- 비밀키 검증 알고리즘

비밀키 복원 알고리즘을 통해 예측된 $\boldsymbol{s}'$에 대해 주어진 LWE 샘플들로 $r=b-\boldsymbol{a}\cdot\boldsymbol{s}' \bmod q$의 분포를 구한다. 만약 $\boldsymbol{s}=\boldsymbol{s}'$라면 r은 에러 e와 같은 분포를 가지고 $\boldsymbol{s}\neq\boldsymbol{s}'$라면 r은 Z_q상에서의 무작위 값과 같은 분포를 가지므로 이를 통해 예측된 비밀키가 맞는지 여부를 검증할 수 있다.

Ⅲ. 실험 결과 및 결론

LWE와 RLWE 문제를 공격하는 실험을 수행하여, 표 1과 같이 LWE가 RLWE에 비해서 공격에 성공하는데 필요한 샘플이 약 n배 가까이 필요한 것으로 나타나 딥 러닝 기반 공격에서 RLWE 기반의 암호가 LWE 기반의 암호보다 필요한 샘플 수 측면에서 취약하다는 결론을 내렸다.

Type \ n	20	30
LWE	44,000	105,000
RLWE	2,100	3,000

표 1. LWE/RLWE의 차원 변화에 따라 공격에 필요한 최소 샘플 수 비교

한편, 기존 논문에서는 공격자가 얻은 LWE 샘플들 중 K개를 골라 $\{1, 0, -1\}$중 하나의 계수를 곱한 뒤 선형 결합을 하여 새로운 학습 샘플을 생성하는 방법을 통해 공격에 필요한 독립적으로 수집한 샘플 개수 N을 줄이는 방법을 제시하였다. 표 2는 $n=30$인 RLWE에 대해 이 방법을 사용해 실험한 결과이다. 이 실험에서 공격에 사용된 총 학습 샘플 수는 $\binom{n\cdot N}{K}\cdot(3^K-1)$이며, 표 1의 결과와 비교했을 때 $K=2,3,4$에 대해 총 학습 샘플 수는 각각 약 200배, 162000배, 2300배만큼 증가하였다.

K	2	3	4
N	70	50	3
총 학습 샘플 수 $\left(\binom{n\cdot N}{K}\cdot(3^K-1)\right)$	1.76×10^7	1.46×10^{10}	2.04×10^8

표 2. RLWE 문제에 대해 선형 결합 개수에 따른 공격에 필요한 독립적인 최소 RLWE 샘플 수 ($n=30$)

본 논문에서는 공격에 필요한 샘플 수를 줄이기 위해 LWE 샘플 $(\boldsymbol{a}, b)$의 b에 이산균등분포 $[-r, r]$로부터 추출된 작은 노이즈를 더해서 새로운 학습 샘플을 생성하는 방법을 제안한다. 표 3은 이 방법을 사용해 실험한 결과이며 표 1과 표 3의 비교를 통해 $n=20$의 RLWE 문제에 대해 공격에 성공하는 데 필요한 RLWE 샘플 수를 최대 약 0.286배, $n=30$에 대해 최대 약 0.267배만큼 줄이는 데 성공한 것을 확인하였다.

r \ n	20	30
2	1100	1500
3	750	1000
4	700	800
5	600	800
6	800	1300

표 3. RLWE 차원과 노이즈의 크기 변화에 따른 공격에 필요한 독립적인 최소 RLWE 샘플 수

표 4는 새로 제시한 방법과 기존 논문의 선형결합 방법을 결합한 기법을 실험한 결과이다. 표 2와 표 4의 비교를 통해 새 기법으로 공격에 필요한 독립적인 학습 샘플 수를 최대 0.5배 더 줄인 것을 확인하였다.

r \ K	2	3
1	40	30
2	35	25
3	60	25

표 4. RLWE 문제에 대해 선형 결합 개수와 노이즈의 크기에 따른 공격에 필요한 독립적인 최소 RLWE 샘플 수 ($n=30$)

본 논문의 실험 결과는 실제 암호에서 사용하는 LWE 문제의 차원(일반적으로 512, 1024 등)과는 많은 차이가 있다. SALSA 논문[1]에서도 이를 언급하고 있으며 이 논문의 후속 논문[3], [4]에서는 BKZ 알고리즘을 활용하여 최대 512차원의 LWE 문제에 대해 $4n$개의 LWE 샘플만으로 비밀키를 복원한 결과를 제시하였다. 이에 따라 BKZ 알고리즘을 사용하여 높은 차원의 LWE에 대해서도 유사한 실험 결과를 나타내는지, 새로 제시한 방법으로 공격에 필요한 LWE 샘플 수를 $4n$개에서 더 줄일 수 있는지에 대한 연구를 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임.(No. 2021-0-00400-002, 저사양 디바이스 대상 고효율 PQC 안전성 및 성능 검증 기술 개발)

참 고 문 헌

[1] Wenger, E., Chen, M., Charton, F., Lauter, K., “SALSA: Attacking Lattice Cryptography with transformers,” Proc. of NeurIPS, 2022.
[2] Regev, O., "On lattices, learning with errors, random linear codes, and cryptography," Journal of the ACM (JACM), Vol. 56, No. 6, pp. 1-40, 2009.
[3] Li, C., Sotáková, J., Wenger, E., Malhou, M., Garcelon, E., Charton, F., Lauter, K., “SALSA PICANTE: a machine learning attack on LWE with binary secrets,” arXiv preprint arXiv:2303.04178, 2023.
[4] Li, C., Sotáková, J., Wenger, E., Allen-Zhu, Z., Charton, F., Lauter, K. et al., “SALSA VERDE: a machine learning attack on Learning With Errors with sparse small secrets,” arXiv preprint arXiv: 2306.11641, 2023.