코드기반암호에 대한 ISD 공격 알고리즘 연구 동향

송경주*, 강예준*, 장경배*, 서화정*
*한성대학교 IT융합공학부
thdrudwn98@gmail.com, etus1211@gmail.com, starj1023@gmail.com,
hwajeong84@gmail.com

ISD attack algorithm research trend for code-based cryptography

Gyeong-Ju Song*, Ye-june Kang*, Kyung-Bae Jang*, Hwa-Jeong Seo*
*Dept. of IT Convergence Engineering, Han-Sung University

요 약

현재, 주요 선진국들을 포함하여 Google과 IBM과 같은 국제 대기업들을 필두로 양자 컴퓨터 개발에 전폭적인 투자들을 하고 있다. 양자 컴퓨터는 특정 분야에 있어 월등한 계산 능력을 보여주며, 기존 컴퓨터에서는 해결할 수 없던 몇몇 문제들을 빠른 시간 내에 해결한다. 이러한 양자 컴퓨터의 등장은 기존 컴퓨터에서는 사실상 풀 수 없는 암호 알고리즘들을 빠른 시간 내에 해결하여 암호학계에 큰 위협이 되고 있다. 현재 사용하고 있는 대부분의 공개키 암호 알고리즘인 RSA와 ECC(Elliptic Curve Cryptography) 또한 공격 대상이다. NIST에서는 다가오는 양자 컴퓨터 시대에 대비하여 양자내성암호 공모전을 주최하였으며 현재 라운드 3에 도입하였다. 본 논문에서는 라운드 3의 후보 알고리즘인 코드기반암호를 공격하는 ISD(Information Set Decoding) 알고리즘에 관한 동향을 조사하였다.

1. 서론

2019년, Google에서는 53-큐비트 양자 프로세서 Sycamore를 통해 기존 컴퓨터에서는 1만년의 시간이 소요되는 난수 증명 문제를 3분 20초 만에 해결하였다[1]. 이는 양자 컴퓨터로 슈퍼 컴퓨터의 계산 능력을 뛰어넘은 첫 번째 양자 우월성을 달성한 사례이다. 2020년, 경쟁 기업인 IBM은 진취적인 양자 로드맵을 발표하였으며 향후 3년 안에 1,000 큐비트 이상의 양자 프로세서, 고성능 양자 컴퓨팅 클라우드 서비스 개발을 목표로 하고 있다.

불가능해 보였던 양자 컴퓨터는 실현되었고 고성능의 양자 컴퓨터 또한 현실화 되어가고 있다. 양자 컴퓨터는 기존 비트가 아닌 0과 1이 확률로서 동시에 존재하는 중첩 성질의 큐비트를 사용하는데, 큐비트가 늘어날수록 다양한 경우의 수를 모두 표현할 수 있어 병렬 계산의 이점을 가진다. 양자 컴퓨터의 계산 능력은 많은 암호 시스템의 안전성을 위협하고 있다. 대표적으로 가장 많이 쓰이고 있는 공개키 암호 시스템인 RSA와 ECC는 소인수 분해와 이산대수 문제의 어려움에 안전성을 기반하고 있다. 하지만 1994년 Shor는 양자 컴퓨터를 활용하여 해당 문제들을 다항시간 내에 해결할 수 있음을 증명하였다[2]. 3072-bit 키를 사용하는 RSA를 공격하기 위해서는 6,145개의 큐비트[3, 4], 같은 보안레벨의 Prime curve를 사용하는 ECC의 경우 2,124개의 논리 큐비트가 필요하다[5]. 다행히도 아직 양자 컴퓨터는 개발 단계이기 때문에 대규모 큐비트는 사용할 수 없다. 하지만 RSA와 ECC를 대체할 수 있는 양자 내성 암호가 필요한 상황이며, NIST(National Institute of Standards and Technology)에서는 양자 내성 암호 표준화 공모전을 진행하고 있다. 현재 라운드 3에 도입함에 따라 Finalist로 7개의 암호 알고리즘, Alternate로 8개의 후보 알고리즘들이 추려진 상태이다. Finalist에는 격자기반, 코드기반, 그리고 다변수 다항식기반의 암호로 구성되어 있다.

코드기반암호는 공개키가 너무 크다는 단점을 가지고 있어 효율성 문제로 거의 사용되지 않았지만 양자 컴퓨터가 등장하고 이에 대한 내성을 가질 것으로 기대되어 현재 주목받고 있다. 양자내성암호는 양자 컴퓨터의 공격으로부터 안전해야 되는 것은 물론이며, 기존 컴퓨터의 공격에서도 안전해야 한다. 본 논문에서는 코드기반암호에 대한 가장 효과적인 공격 알고리즘인 ISD(Information Set Decoding)의 연구 동향에 대해 살펴본다.

2. 관련연구

2.1 코드기반암호

1978년, Robert J. McEliece는 통신 채널에서 사용되던 코딩이론을 활용하여 새로운 공개키 암호 시스템 McEliece를 제안하였다[6]. 동작 구조는 다음과 같다. 오류 수정이 가능한 생성 행렬을 공개키로 사용한다. 송신자는 공개키와 메시지를 조합하여 인코딩 한 뒤, 의도적으로 오류를 추가한 암호문을 전송한다. 올바른 수신자는 생성 행렬에 대한 패리티 체크 행렬을 통해 오류를 제거하여 원본 메시지를 획득할 수 있다. 코드기반암호는 양자 컴퓨터의 공격에 대해 내성을 가질 것으로 기대되며, 현재 코드기반암호인 Classic McEliece는 NIST 양자내성암호 공모전의 Finalist에 포함되어 있다.

2.2 신드롬 디코딩

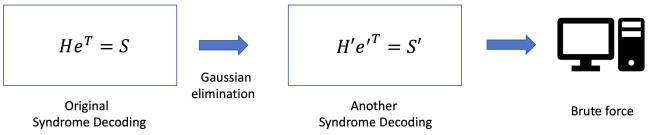
신드롬 값인 S는 패리티체크 행렬 H와 특정 무게조건의 벡터 e를 곱한 결과 값이며 수식 1과 같다.

$$He^T = S \tag{1}$$

코드기반암호에서는 H를 공개키로 사용하며 벡터 e는 비밀정보, 신드롬 값 S를 암호문으로 사용한다. 이때 공개키인 H와 암호문인 S를 알고 있다 하더라도 특정 무게조건의 벡터 e를 찾는 것은 NP-hard 로 분류된다. H를 생성할 때 사용 된 값들은 개인키가 되고, 올바른 수신자는 S로부터 신드롬 디코딩을 수행한다. 이를 통해 원본 벡터 e를 복구하고 비밀 정보로 사용하여 상호간에 키 교환을 완료한다.

2.3 Information Set Decoding

ISD는 코드기반암호를 공격하는 가장 효과적인 알고리즘이다. 신드롬 디코딩 문제에서 H와 S가 주어졌을 때, ISD는 개인키를 찾아내는 것이 아닌 바로 원본 벡터 e를 복구한다. ISD는 크게 두 가지 단계로 구성되며 그림 1과 같다.



(그림 1) Information Set Decoding.

첫 번째, $He^T = S$의 신드롬 디코딩 문제에 대하여 Gaussian elimination을 수행함으로써 다른 문제 $H'e'^T = S'$ 로 변환한다. 두 번째, 새로운 문제 $H'e'^T = S'$에 대해 무게 조건을 만족하는 벡터 e'을 찾기 위한 전수조사를 수행한다. 만약 찾지 못했다면 처음으로 돌아가, Guassian을 다시 수행한다.

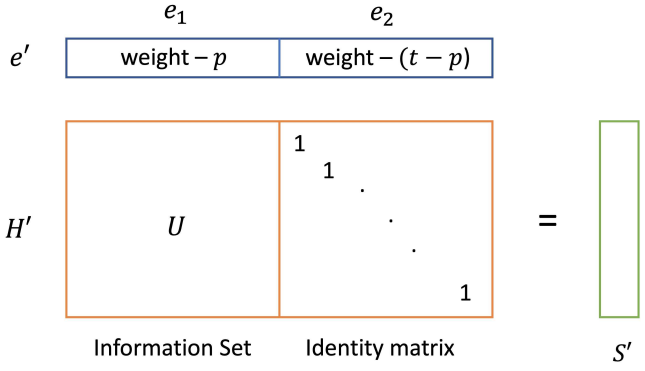
3. Infomation Set Decoding 연구 동향

3.1 Prange ISD

1962년, Prange는 가장 기본적인 ISD를 제안하였으며[7] 오늘날 다양한 버전의 ISD는 모두 Prange의 ISD로부터 개선된 알고리즘들이다. Prange는 앞서 ISD의 두 가지 단계 중, 1단계인 Gaussian elimination을 수행하지 않는다. 때문에 신드롬 값 S가 0인 경우에는 공격이 불가능하다는 단점이 있다.

3.2 Lee-Brickell ISD

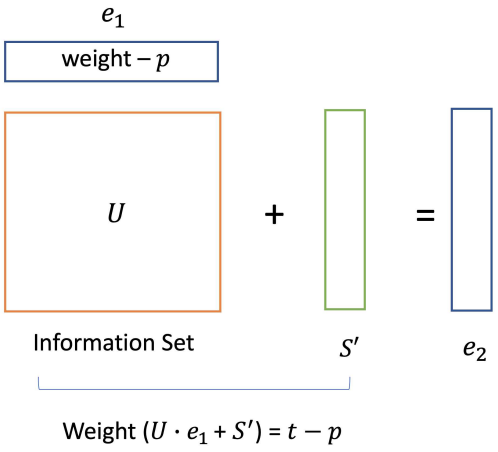
Lee-Brickell은 Prange의 알고리즘을 개선하였다[8]. 공개키 역할을 하는 패리티체크 행렬 H에 Guassian elimination을 적용하여 왼쪽 또는 오른쪽 부분이 Identity 행렬인 Systematic 형식으로 재구성한다. 이때, Identity가 아닌 행렬을 Information Set으로 정의하며, Information Set과 곱해지는 벡터 부분은 e_1, Identity matrix와 곱해지는 벡터 부분은 e_2로 정의한다. 신드롬 디코딩 문제에서, 벡터 e의 무게조건이 t일 때, 이 때, e_1 벡터의 무게조건은 작은 수 p, e_2의 무게조건은 $t-p$가 된다. 무게 p는 행렬 H의 크기마다 다르며 최적의 p가 사용된다. Lee-Brickell의 ISD에서 1단계, Gaussian elimination을 통한 신드롬 디코딩의 재구성 및 벡터 e의 무게분포 설정은 그림 2와 같다.



(그림 2) Lee-Brickell Information Set Decoding.

1 단계인 신드롬 디코딩 문제의 변형을 완료했다면, 무게조건 p의 벡터 e_1을 찾기 위해 다음 수식 2를 만족하는 전수조사를 수행한다. 이에 대한 적체적인 구조는 그림 3과 같다.

$$\text{weight-}(U \bullet e_1 + S') = t - p \quad (2)$$



(그림 3) Brute force for Lee-Brickell ISD

수식 2를 만족하는 벡터 e_1을 찾으면 ISD 공격을 통한 원본벡터 e를 복구할 수 있다. 벡터 e_2는 Identity matrix와 곱해지기 때문에 만약 수식 2를 만족할 경우, $U \bullet e_1 + S'$의 값과 동일하게 벡터 e_2가 구성되고 무게조건인 $t-p$를 달성한다. 이와 동시에 신드롬 값 S'를 만족하게 되고, e_1과 e_2로 구성된 무게 t의 벡터 e'을 찾을 수 있다. 따라서 e_1을 변경해가며 전수조사를 수행하고 수식 2를 만족하는 벡터 e_1을 찾지 못했다면 다시 돌아가 Gaussian elimination을 수행한다.

3.3 Quantum Information Set Decoding

대표적인 양자 알고리즘인 Grover 알고리즘은 최대 N번의 쿼리가 필요한 전수조사에 대해, 최대 $\sqrt{N}$번의 쿼리로 감소시키는 양자 탐색 알고리즘이다. Grover 알고리즘은 크게 Oracle과 Diffusion operator 두 단계로 구성된다. Oracle에서는 탐색 대상인 답을 반환하며, Diffusion operator는 Oracle에서 반환한 답의 관측 확률을 증폭시킨다. Information Set Decoding에서 오랜 시간을 차지하는 부분은 2 단계인 전수조사이다. 따라서 Grover 알고리즘을 ISD의 전수조사 단계에 적용하여 복잡도를 반으로 줄이는 양자 버전의 ISD 또한 연구되고 있다[9-11].

3.4 Information Set Decoding with Hints

해당 논문[12]에서는 신드롬 디코딩 문제에 대하여 ISD를 수행 시 획득한 부가적인 정보들을 힌트로 표현하며, 힌트들을 조합하여 ISD의 복잡도를 줄였다. 원본 벡터에서 확실히 1인 곳과 확실히 0인 위치를 아는 것은 힌트가 된다. 그리고 이러한 힌트를 조합한 부채널 분석을 통해 특정한 부분만 검색함으로써 공격 복잡도를 감소 시켰다. 원본 벡터에서 블록 단위의 무게를 알고 있는 것 또한 힌트에 해당한다. 이때는 템플릿 공격을 활용한 무게 템플릿을 적용, 신드롬 디코딩 문제에 대해 다수의 작은 신드롬 디코딩 문제로 변환하여 복잡도를 감소시켰다.

4. 평가 분석 및 결론

본 논문에서는 코드기반암호에 대한 공격 알고리즘인 ISD의 연구 동향에 대해 살펴보았다. 고전적인 Prange의 ISD부터 시작하여, 다양한 버전의 ISD가 연구되고 있다. 이는 기존 Prange ISD의 복잡도와 비교해 보았을 때 획기적인 개선이 이루어지지는 않는다. 하지만 앞서 살펴보았던 Grover 탐색 알고리즘을 활용하거나 부가적인 정보를 활용한 부채널 분석을 적용한 연구 결과들은 ISD의 복잡도를 크게 개선시킬 수 있었다. ISD에 대한 연구 동향들을 종합해 보았을 때, Lee-Brickell의 ISD 또는 본 논문에서 언급되지는 않았지만 Stern의 ISD와 같이 기존 방법에서 수학적인 접근을 통해 개선한 ISD는 복잡도를 미세하게 감소시킨다. 하지만 양자 알고리즘을 활용하거나 부채널 분석 즉, 다른 기술을 기존 ISD에 적용하는 연구들은 복잡도를 크게 개선시키는 결과를 보여준다. 코드기반암호는 다가오는 양자 컴퓨터 시대에 대비하여 양자내성암호의 역할을 할 수 있을 것이라 기대되고 있다. 그리고 이 안전성이 계속 보장되는지에 대해서는 공격 알고리즘인 ISD의 최신 연구 동향들을 주시할 필요가 있다.

5. Acknowledgment

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (<Q|Crypton>, No.2019-0-00033, 미래컴퓨팅 환경에 대비한 계산 복잡도 기반 암호 안전성 검

증 기술개발)

참고문헌

[1] Arute, F. et al. “Quantum supremacy using a programmable superconducting processor” Nature 574, 505 - 510. 2019

[2] Peter W. Shor “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer” SIAM review, Vol. 41. No. 2. 303-332. 1999

[3] Thomas Häner, Martin Roetteler, Krysta M. Svore “Factoring using 2n+2 qubits with Toffoli based modular multiplication” Quantum Information and Computation. Vol. 17. No. 7. pp 673-684. 2017

[4] C. Gidney, “Factoring with n+2 clean qubits and n-1 dirty qubits” eprint. 2018.

[5] Martin Roetteler, Michael Naehrig, Krysta M. Svore, Kristin Lauter. “Quantum resource estimates for computing elliptic curve discrete logarithms” ASIACRYPT. pp 241-270. 2017

[6] Robert J. McEliece. “A public-key cryptosystem based on algebraic coding theory” Technical report. NASA. 1978.

[7] Prange. “The use of information sets in decoding cyclic codes” IRE Transactions. IT-8. S5-S9. 1962

[8] Lee, P., Brickell. “An observation on the security of McEliece’s public-key cryptosystem” In G¨unther, C., Advances in Cryptology - EUROCRYPT ’88. Vol. 330 pp 275 - 280. 1988

[9] Kachigar, Ghazal, Jean-Pierre Tillich “Quantum information set decoding algorithms” International Workshop on Post-Quantum Cryptography. Springer. pp 66-89. 2017

[10] Kirshanova, Elena. “Improved quantum information set decoding” International Conference on Post-Quantum Cryptography. Springer. pp 507-527. 2108

[11] Canto-Toress. R, Sendrier. N, “Analysis of information set decoding for a sub-linear error weight” In Post-Quantum Cryptography. pp 144-161. 2016

[12] AL. Horlemann, S. Puchinger, J. Renner, T Schamberge, A. Wachter-Zeh “Information-Set Decoding with Hints” Cryptology ePrint Archive. Report 2021/279, 2021.