인공지능기반 오류주입 매개변수 탐색 기술 동향

이진용, 나중찬, 김명현
한국전자통신연구원 차세대시스템보안연구실
jinyonglee@etri.re.kr, njc@etri.re.kr, kimmh12@etri.re.kr

A Study on AI-Based Fault Injection Parameter Search Algorithm

Jinyong Lee, Joongchan Na, Myunghyun Kim
Electronics and Telecommunications Researcch Institute
System Security Research Division
jinyonglee@etri.re.kr, njc@etri.re.kr, kimmh12@etri.re.kr

요 약

오류 주입 공격(Fault Injection Attack)은 시스템에 의도적으로 오류를 주입하여 취약점을 드러내는 강력한 공격 기법으로, 공격에 사용되는 매개변수 탐색 최적화는 제한된 시간 내 성공적인 공격을 위해 필수적인 기술이다. 본 논문에서는 인공지능(Artificial Intelligence) 기반 알고리즘, 특히 유전 알고리즘(Genetic Algorithm, GA)과 미메틱 알고리즘(Memetic Algorithm, MA)이 오류 주입 매개변수 탐색 최적화에 어떻게 적용되었는지 사례를 통해 살펴보고, 이러한 기법들이 기존의 무작위 탐색(Random Search, RS) 방법에 비해 높은 탐색 효율성과 성공률을 어떻게 달성했는지 분석한다. 또한, 각 알고리즘의 장단점을 비교하여 평가하고, 오류 주입 공격의 정밀도와 효율성을 더욱 향상시키기 위한 추가 연구 방향을 제안한다.

1. 서론

오류 주입 공격은 하드웨어에서 작동하는 소프트웨어나 알고리즘의 취약점을 의도적으로 노출시키기 위해 결함을 유도하는 공격 기법으로, 임베디드 시스템에 대한 강력한 위협으로 부상하고 있다. 이 공격은 전압 글리치, 레이저, 전자기파 등 다양한 물리적 오류원을 타겟 칩에 직접 주입하는 방식으로 진행되며, 성공적인 공격을 위해 오류의 강도, 시점 등 매개변수를 세심하게 조정해야 한다. 그러나 이러한 매개변수의 탐색 공간은 매우 고차원적이고 복잡하여, 효율적인 탐색 방법이 필수적이다.

최근 인공지능 기반의 알고리즘, 특히 유전 알고리즘과 미메틱 알고리즘은 오류 주입 공격에서 매개변수 탐색의 효율성을 크게 향상시킬 수 있는 방법으로 주목받고 있다. 유전 알고리즘은 생물학적 진화 과정을 모방하여 최적의 해를 탐색하며, 미메틱 알고리즘은 여기에 국부 탐색(Local search)을 추가하여 더 정교한 최적화가 가능하다.

본 논문에서는 오류 주입 공격의 매개변수 탐색 문제를 해결하기 위한 AI 기반 알고리즘을 살펴보고, 기존 알고리즘과의 비교를 통해 각 방식의 장단점을 분석한다. 마지막으로 향후 연구 방향에 대해 제안한다.

2. 오류 주입 공격

오류 주입 공격은 하드웨어 시스템의 보안 메커니즘을 우회하거나 소프트웨어 혹은 알고리즘의 오동작을 유발하기 위해 인위적으로 오류를 발생시키는 능동적인 공격 기법으로, 시스템의 물리적 환경에 외부 충격을 가하여 정상적인 동작을 방해하고, 이를 통해 민감한 정보를 추출하거나 시스템을 비정상적으로 작동하게 만드는 것을 목표로 한다. 오류 주입 공격은 다양한 방식으로 실행될 수 있으며, 일반적으로 전압 글리칭 (Voltage Glitching) [1-3], 전자기파 주입 (Electromagnetic Fault Injection, EMFI) [4-5], 레이저 주입 (Laser Fault Injection) [6] 과 같은 물리적 공격 기법들이 주로 사용된다.

전압 글리칭은 전자 기기에 공급되는 전압을 순간적으로 변동시켜 오류를 유발하는 방법으로, 주요 매개변수로는 글리치의 타이밍, 전압 크기, 그리고

글리치 지속 시간이 있다. 이 방법은 비교적 저비용으로 공격을 수행할 수 있으며, 소형 임베디드 장치에서 매우 효과적인 공격 기법으로 평가받고 있다.

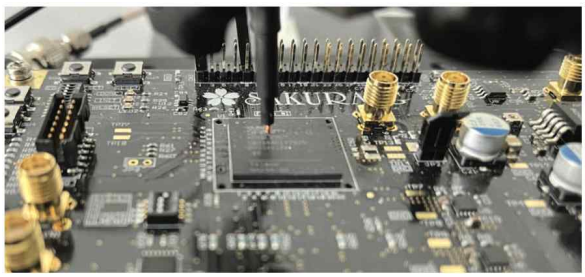

그림 1. FPGA에 대한 전자기파 오류 주입 예 [5]

전자기파 주입은 고출력 전자기파를 일시적으로 기기에 주입하여 기기 내의 저장장치에 있는 데이터를 왜곡시키거나, 명령어 파이프라인 내 명령어를 잘못 처리하도록 유도해 다른 명령어로 인식시키거나 건너뛰게 만드는 공격 기법이다. 이 기법은 다른 방법에 비해 무접촉 방식으로 대상 기기에 대한 파손 우려가 적어 매력적인 공격 기법으로 각광받고 있다. 주요 매개변수로는 전자기파의 강도, 지속 시간, 대상 기기상의 물리적 주입 위치 등이 있다.

레이저 주입은 가장 고가의 장비를 요구하지만 매우 높은 정밀도로 반도체 소자나 마이크로컨트롤러의 특정 위치에 레이저를 발사하여 특정 영역에 오류를 발생시키는 방식으로 매우 정밀한 공격이 가능하다는 점에서 높은 파괴력을 가진다. 때문에 전압, 전자기파 오류 주입이 원활하지 않은 경우에 사용되는 경우가 많으며, 주요 매개변수는 레이저의 강도, 펄스 지속 시간, 그리고 주입 위치 등이다.

오류 주입 공격의 성공적 수행을 위해서는 정확한 매개변수를 설정이 중요하며, 이를 통해 기기에 보안상 치명적인 오류를 일으킬 수 있다. 그러나 매개변수의 개수가 많아질수록 조합이 복잡해지고, 무작위 탐색(Random Search) 또는 몬테카를로 탐색(Monte Carlo Search) 방법은 탐색 공간이 클 경우 매우 비효율적이며, 공격 성공 확률이 낮다는 단점이 있다 [2]. 이러한 문제를 해결하기 위해 최근에는 인공지능 기반 탐색 기법들이 제안되고 있으며, 그 중 유전 알고리즘과 미메틱 알고리즘이 오류 주입 공격의 매개변수 최적화 문제를 해결하는 데 효과적임이 입증되었다 [3].

3. AI 기반 오류 주입 매개변수 탐색 기술

오류 주입 매개변수의 효과적인 탐색을 위해서는 오류 주입의 성공 여부를 평가하고, 이를 바탕으로 더 나은 매개변수를 탐색하는 과정이 필요하다. 기존의 AI 기반 최적화 기술, 특히 유전 알고리즘은 주어진 매개변수에 따라 오류 주입을 수행한 후, 타겟 장치의 응답(예: 리셋, 오류 응답, 무응답 등)을 기반으로 적합도 점수(Fitness Value)를 계산해 이후 탐색 방향을 결정하는 방식을 사용한다. 본 장에서는 유전 알고리즘과 미메틱 알고리즘을 중심으로 AI 기반 오류 주입 매개변수 탐색 기법을 다룬다.

3-1. 유전 알고리즘 (Genetic Algorithms, GA)

유전 알고리즘은 생물학적 진화 과정을 모방한 탐색 기법으로, 자연 선택, 교차(Crossover), 돌연변이(Mutation)와 같은 진화 연산을 통해 최적의 해를 탐색한다. 유전 알고리즘은 오류 주입 공격에서 성공 가능성이 높은 매개변수 조합을 우선적으로 빠르게 탐색하기 위해 도입되었으며, 기존의 무작위 탐색보다 훨씬 빠르고 정확하게 공격 매개변수를 찾을 수 있다는 점에서 많은 연구가 이루어졌다 [1-6].

Picek 등은 스마트카드에 대한 전압 글리칭 공격에서 유전 알고리즘을 적용하여 전통적인 랜덤 탐색보다 성공률이 크게 향상된 것을 보여주었다 [1-2]. 특히, 유전 알고리즘은 매개변수 공간을 무작위로 탐색하는 것이 아니라, 이전 세대의 결과를 바탕으로 더 유망한 후보를 선택하는 방식으로 진행되므로 탐색 효율을 크게 높일 수 있다. 또한, 유전 알고리즘은 블랙박스 상황에서도 작동할 수 있기 때문에, 공격 대상 시스템에 대한 사전 정보가 부족할 때에도 유용하다.

3-2. 미메틱 알고리즘 (Memetic Algorithms, MA)

미메틱 알고리즘은 유전 알고리즘에 국부 탐색(Local Search)을 결합한 하이브리드 최적화 방법이다. 유전 알고리즘이 전역 최적화를 수행하는 동안, 국부 탐색은 특정 영역 내에서 더 세밀한 탐색을 진행하여 최적해를 보다 빠르게 찾는 데 기여한다. 미메틱 알고리즘은 오류 주입 공격에서 단순한 유전 알고리즘보다 더 정교한 매개변수 최적화를 가능하게 하며, 공격 성공률을 극대화하는 데 효과적이다 [3].

특히, 최근 연구에서는 레이저 오류 주입 공격에서 미메틱 알고리즘이 매우 유용하게 적용되고 있다 [6]. Krček 등은 레이저 오류 주입에서 미메틱 알고리즘을 활용하여, 넓은 범위의 매개변수 공간을 효율적으로 탐색하면서도 다양한 취약 지점을 찾아냈다고 보고하였다. 이 연구에서는 국부 탐색을 위해 Hooke-Jeeves 알고리즘을 사용하였으며, 이를 통해 레이저 주입 시 다양한 위치에서 오류를 유발하는 최적 매개변수를 찾는 것을 보였다 [6].

4. 알고리즘별 성능 비교 및 장단점

4-1. 유전 알고리즘과 무작위 탐색의 비교

전통적인 무작위 탐색은 탐색 공간 내에서 가능한 매개변수 조합을 무작위로 선택하여 테스트하는 방식으로, 일반적으로 탐색 공간이 매우 큰 경우에는 효율성이 떨어진다. 특히, 오류 주입 공격에서는 수많은 매개변수 조합 중 성공적인 공격을 유발하는 소수의 최적 조합을 찾아내야 하는데, 무작위 탐색은 이러한 최적화를 거의 고려하지 않는다. Picek 등의 연구에 따르면, 몬테 카를로 기반 탐색은 매우 낮은 성공률을 보였으며, 약 3,072개의 측정을 수행했으나 성공적인 오류 주입을 단 한 번도 찾아내지 못했다 [1-2].

반면, 유전 알고리즘은 매 세대마다 적합도가 높은 매개변수를 선택하고, 교차 및 돌연변이를 통해 탐색을 진행하기 때문에 무작위 탐색보다 훨씬 높은 효율성을 보인다. Boix Carpi 등의 연구에서는 약 1,560번의 측정만으로 유전 알고리즘이 8번의 성공적인 오류 주입을 찾아냈으며, 이는 무작위 탐색보다 훨씬 우수한 성과였다. 이러한 결과는 유전 알고리즘이 오류 주입 공격에서 매우 유용한 기법임을 시사한다 [1, 2].

4-2. 유전 알고리즘과 미메틱 알고리즘의 비교

유전 알고리즘은 전체 탐색 공간을 효율적으로 탐색할 수 있지만 특정 상황에서는 국부 최적화(Local Optimum)에 빠지는 단점이 있다. 이에 비해 미메틱 알고리즘은 유전 알고리즘에 국부 탐색을 추가하여 탐색 효율성 및 속도를 높이는 한편, 유전 알고리즘이 가지고 있는 전역 탐색 기능을 극대화시킬 수 있었다.

Krček 등은 레이저 오류 주입 실험에서 미메틱 알고리즘을 사용한 결과, 유전 알고리즘보다 더 많은 취약점을 빠르게 발견했을 뿐만 아니라 탐색 공간의 넓이(Location Coverage)가 크게 향상되었다고 보고하였다. 이는 미메틱 알고리즘의 특성, 즉 유전 알고리즘을 이용한 전역 탐색 후 지역 최적해를 빠르게 탐지하는 국부 탐색을, 이후 유전 알고리즘을 이용한 전역 탐색이 이어지는 최적화 기법의 효과가 반영된 결과이다. 이로 인해 더 적은 시간 내에 다양한 오류 주입 경로를 찾아낼 수 있었으며 이는 레이저 오류 주입과 같이 매개변수 공간이 매우 넓은 공격에서 특히 필요한 특성이다 [6].

탐색 방법	장점	단점
무작위 선택	간단하고 구현이 용이	매우 낮은 탐색 효율 및 낮은 성공률
유전 알고리즘	빠른 탐색 속도와 높은 성공률	국부 최적해에 빠질 가능성이 있음
미메틱 알고리즘	국부 정밀 탐색을 통한 높은 정밀도, 높은 위치 커버리지	유전 알고리즘에 비해 높은 계산 자원 소모

표 1. 알고리즘별 장단점

5. AI 기반 오류주입 매개변수 최적화 연구 방향

오류 주입 공격의 매개변수 최적화 문제는 매우 복잡하고 다차원적인 문제로, AI 기반 탐색 기법은 오류 주입 공격에서 중요한 성과를 거두고 있지만, 아직 해결해야 할 과제와 향후 발전 가능성이 존재한다. 다음은 이러한 연구 분야에서 주요한 미래 방향과 미해결 과제이다.

5-1. 정교한 탐색 기법의 개발

유전 알고리즘과 미메틱 알고리즘은 매개변수 탐색에서 뛰어난 성과를 보여주고 있지만, 여전히 국부 최적화에 빠지거나 탐색 공간 내 특정 영역을 놓치는 한계가 있다. 이러한 문제를 해결하기 위해 강화학습(Reinforcement Learning, RL)과 같은 자가학습(Self-learning) 기반의 방법론이 결합될 가능성이 있다. 강화학습은 시스템의 환경 변화에 적응하여 실시간으로 탐색 전략을 수정할 수 있으며, 더 동적인 탐색 기법을 개발하는 데 기여할 수 있다.

5-2. 다양한 데이터를 활용한 탐색 기법 개발

현재의 AI 기반 탐색 기법은 의사결정에 사용되는 데이터가 제한적으로, 이를 개선하기 위해서는 더 다양한 데이터를 통합한 탐색 기법 개발이 필요하다. 기존 알고리즘들은 대체로 주어진 매개변수에

따라 오류 주입을 실시하였을 때의 타겟 장치의 응답 (예: 타겟 리셋, 오류 응답, 무응답, 혼재된 응답)을 기반으로 적합도 점수를 계산하고, 이를 바탕으로 다음 탐색 방향을 결정하는 방식으로 동작해왔다.

그러나, 보다 정교한 탐색을 위해서는 타겟 장치의 파형 응답(Waveform Response)과 같은 더 복합적이고 세밀한 데이터를 활용할 필요가 있다. 이러한 추가적인 데이터를 통합함으로써, 알고리즘이 매개변수 조합의 결과를 더 정확하게 평가할 수 있으며, 공격의 정밀도와 탐색 효율성을 크게 향상시킬 수 있을 것으로 기대된다.

5-3. 다양한 공격 유형에 대한 탐색 기법 확장

현재 AI 기반 탐색 기법은 주로 전압, 전자기파, 레이저 오류 주입과 같은 특정 물리적 공격 유형에 최적화되어 개발되어 왔다. 그러나 매개변수 탐색 기법의 발전으로 탐색 가능 영역이 넓어지게 되면 향후에는 복합적인 오류 주입 기법(예: 두 가지 이상의 결합된 공격)이나 온도 변화 등 더 다양한 오류 주입 방식에 대해서도 효율적인 매개변수 탐색 기법이 가능해질 것이다. 이에 따라 복합적 오류 주입에 최적화된 탐색 기법 개발이 요구될 것이다.

이를 위해서는 각 공격 유형별 매개변수 설정의 특성을 체계적으로 분석하고, 이를 기반으로 특성화(Characterization) 하는 작업이 필요하다. 이러한 분석을 바탕으로 공격 유형의 특성에 맞춘 AI 기반 탐색 전략을 개발하는 연구가 필요하다. 특히, 복합적인 오류 주입 기법의 경우 매개변수 간의 상호작용이 복잡하게 얽혀 있을 수 있으므로, 다차원적인 탐색 공간에서의 효율적인 탐색을 지원하는 알고리즘 개발이 필요할 것이다. 이러한 접근은 공격의 정밀도와 성공률을 크게 향상시키는 데 기여할 것이다.

6. 결론

오류 주입 공격에서 효율적인 매개변수 탐색은 공격 성공의 핵심 요소로, 전통적인 방법보다 AI 기반 알고리즘이 훨씬 뛰어난 성능을 발휘한다. 유전 알고리즘과 미메틱 알고리즘은 복잡한 탐색 공간에서 신속하고 정확한 탐색을 가능하게 하며, 특히 오류 주입 공격의 성공률을 크게 높였다.

AI 기반의 오류 주입 공격 매개변수 최적화 연구는 현재 중요한 성과를 거두고 있지만, 여전히 극복해야 할 과제와 발전 가능성이 남아 있다. 유전 알고리즘과 미메틱 알고리즘을 넘어, 강화학습과 같은 자가 학습 기법의 도입은 보다 동적이고 적응적인 탐색을 가능하게 할 것이다. 또한, 다양한 데이터를 통합한 정교한 탐색 기법과 복합적인 오류 주입 기법에 대응하는 다차원적 탐색 방법론의 개발은 향후 연구에서 중요한 도전 과제가 될 것이다. 이를 통해 오류 주입 공격의 정밀도와 효율성을 극대화함으로써, 보다 복잡한 시스템을 효과적으로 분석하고 보안을 강화하는 데 기여할 수 있을 것이다.

ACKNOWLEDGEMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구결과임. (RS-2024-00396269, IC Chip에 대한 오류 주입 보안 검증 및 취약성 분석 시스템 개발)

참고문헌

[1] Boix Carpi, R., et al. "Glitch It If You Can: Novel Parameter Search Strategies for Successful Fault Injection." CARDIS 2013, Berlin, Germany, 2013.

[2] Picek, Stjepan, et al. "Evolving genetic algorithms for fault injection attacks." MIPRO 2014, Opatija, Croatia, 2014, pp. 1105-1111.

[3] Picek, S., et al. "Fault injection with a new flavor: Memetic algorithms make a difference." COSADE 2015, Berlin, Germany, 2015, pp. 159-173.

[4] Maldini, A., Samwel, N., Picek, S., & Batina, L. (2019). Optimizing electromagnetic fault injection with genetic algorithms. Automated Methods in Cryptographic Fault Analysis, 281-300.

[5] Rais-Ali, I., et al. "Quantifying the speed-up offered by genetic algorithms during fault injection cartographies." FDTC 2022, Seoul, Korea, 2022, pp. 61-72.

[6] Krček, M., & Ordas, T. "Diversity Algorithms for Laser Fault Injection." ACNS 2024, Cham, Switzerland, 2024, pp. 121-138.