DDR4 DRAM에 대한 EM Fault Injection 기반 Persistent Fault Analysis 기법에 관한 연구

허재원*, 김연재, 장우현
엘아이지넥스원
*jaewon.huh@lignex1.com, yeonjae.kim@lignex1.com, woohyun.jang2@lignex1.com

A Study on Persistent Fault Analysis Method for DDR4 DRAM Based on Electromagnetic Fault Injection

Huh Jae Won*, Kim Yeon Jae, Jang Woo Hyun
LIG Nex1

요 약

최근 고속 통신 및 고성능 연산을 요구하는 스마트폰, 자율주행 차량, 인공지능 등에서 DRAM은 필수적인 구성 요소로 자리잡고 있다. 기술의 발전으로 DRAM의 cell 집적도가 올라가면서 외부 물리적 요인에 대한 취약성이 증가하고 있으며, 이는 기존의 소프트웨어 기반 공격 외에도 하드웨어 차원의 취약성이 발생할 수 있다. 본 논문에서는 전자기파(Electromagnetic, EM) 오류 주입을 통해 DDR4 DRAM의 데이터를 변조할 수 있음을 실험적으로 확인하고 이를 활용한 Persistent Fault Analysis(PFA) 기법의 적용 가능성을 DDR4 DRAM 수준에서 제시한다.
실험 결과, 전자파 기반 오류 주입은 DRAM 내부의 데이터에 대한 영구적인 오류를 발생시킬 수 있으며, 이로 인한 암호키가 유출될 수 있는 취약점으로 작용할 수 있음을 확인하였다. 기존의 DRAM에서 다뤄지는 Rowhammer 공격에 대한 여러 논의와 달리, 본 연구는 직접적인 물리 채널 공격의 현실적 위협성과 그에 따른 가능성을 다룬다는 점에서 의미를 갖는다. 이러한 결과는 향후 안전한 통신 시스템을 설계하기 위해 논리적인 채널뿐만 아니라 물리 채널에 대한 보안 대응 기법 역시 반드시 고려되어야 함을 시사한다.

Ⅰ. 서 론

최근 스마트폰, 자율주행 차량, 인공지능 등의 분야에서는 대규모 데이터를 고속으로 처리해야 할 필요성이 증가하고 있다. 이러한 환경에서 DRAM은 시스템을 구성하는 프로세서의 성능과 직결되는 핵심 메모리 소자로서 그 중요성이 높아지고 있으며, DRAM의 성능과 보안성에 관한 다양한 연구가 활발히 진행되고 있다.

DRAM은 인공지능 분야에서 학습된 모델의 가중치나 신경망의 주요 파라미터와 같은 민감한 정보를 저장하는 저장 매체로 활용된다. 이에 따라 DRAM 내 데이터의 무결성을 보장하는 것은 중요한 과제로 대두되고 있다.

기술의 발전으로 DRAM의 집적도가 증가함에 따라 처리할 수 있는 데이터의 양과 속도는 향상되었지만, 그에 비례하여 다양한 공격이 이루어지면서 여러 가지 취약성이 발견되고 있다. 기존의 DRAM 무결성을 위협할 수 있는 공격인 Rowhammer[1]에 대응하기 위해 Target Row Refresh(TRR)와 같은 방어 기법이 도입되는 추세이며, 이 기법은 행(row)에 접근하는 패턴을 감지하여 특정 행에 과도한 접근이 감지되면 인접한 행을 자동으로 갱신함으로써 비트 플립을 억제한다. 그러나 이러한 기법은 논리적 접근 패턴 분석에 기반한 방어에 국한되며, 전자기파(Electromagnetic, EM) 기반의 물리적 오류 주입(Fault Injection)과 같은 공격에는 대응이 어렵다는 한계를 가진다[2].

EM 오류 주입은 회로에 직접적인 전자기적 간섭을 유발하여, TRR과 같은 반복적인 접근을 감지하는 보호 메커니즘을 우회하여 특정 메모리 영역에 오류를 발생시킬 수 있다. 특히 이러한 오류가 암호 알고리즘의 핵심 연산에 사용되는 AES S-box와 같은 테이블에 발생할 경우, 암호화 자체에 문제가 발생할 수 있으며 이는 곧 통신 과정에서 메시지 무결성 손상, 키 유출, 또는 인증 실패와 같은 치명적인 보안 문제로 이어질 수 있다. 따라서 기존의 DRAM 보안 대책만으로는 물리적 채널을 통한 공격을 방지하기 어려우며, AES와 같은 암호 모듈을 포함한 시스템 전반의 신뢰성을 보장하기 위해서는 오류 주입 공격에 대한 분석 및 대응 기법에 대한 체계적인 연구가 요구된다.

Ⅱ. 본론

1. 배경지식

1.1 DRAM 구조

DRAM은 여러 개의 DRAM 칩이 탑재된 DIMM(Dual inline Memory Module) 형태의 모듈 구조로 구성된다. 하나의 DIMM에는 여러 DRAM 칩이 탑재되며, 각 DRAM 칩은 내부적으로 다수의 bank로 구성되어 병렬적인 데이터 접근을 가능하게 한다. DDR4 규격에서는 각 칩은 16개의 bank로 구성되며 각 bank는 다수의 row로 구성되어 있고, 각 row는 다시 여러 개의 메모리 cell로 이루어져 있다. DRAM의 각 cell은 데이터를 저장하는 커패시터(capacitor)와 이를 제어하는 트랜지스터로 구성되며, 저장된 전하는 시간이 지남에 따라 자연스럽게 누설된다. 이러한 전하 누출로 인해, 모든 cell은 주기적으로 데이터를 읽고 다시 쓰는 '리프레시(refresh)' 작업이 필요하며 JEDEC 표준에 따르면, 모든 cell은 64ms 이내에 최소 한 번 이상 리프레시되어야 한다.

1.2 전자파기반 오류주입 공격

오류주입 공격은 공격 대상 장치가 동작할 때 장치 내부 회로 등에 인위적인 오류를 주입하여 오동작을 유발한다. 오류를 주입하면 내부 데이터를 변조하거나 코드의 흐름 인위적으로 변경할 수 있으며, 공격 장비의 성능과 주입하는 오류 원 등에 따라서 다양한 연산이나 저장된 데이터에 대한 공격이 가능하다. 이를 활용하여 암호화 동작 과정에서 일시적인 오류를 생성하여 비정상적인 출력을 생성하여 암호키를 분석하거나[3], 장치가 재시작되는 경우가 아니라면 장치 내 데이터에 오류가 남아 변조된 데이터로 인하여 지속적인 오류가 유발되도록 하는 연구가 진행되고 있다[4].

1.3 AES T-table

Advanced Encryption Standard(AES)는 블록 단위의 대칭키 암호화 알고리즘으로, 비선형 연산(SubByte)과 선형 연산들(ShiftRow, MixColumn, Addroundkey)으로 이루어진 라운드가 반복되면서 암호화를 진행한다. T-table 방식은 이러한 반복 연산을 미리 계산된 T-table을 참조하는 것으로 대체하여 연산 속도를 높이는 기법이다[5]. 이 테이블은 SubByte, ShiftRow, Mixcolumn의 조합을 사전에 계산하여 32비트 단위로 저장하며, 각 바이트 입력값에 대한 결과 값을 테이블에서 바로 조회할 수 있는 장점이 있다.

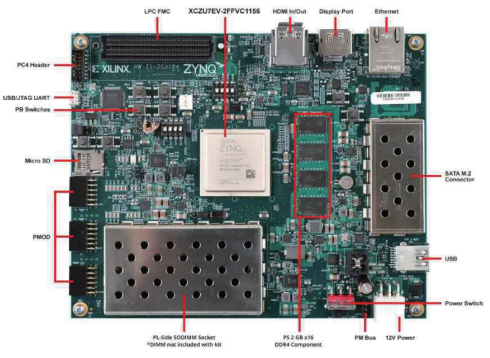
1.4 Persistent Fault Analysis

암호 알고리즘 내의 테이블(예: S-box)에 영구적인 오류가 주입된다면, 오류가 주입된 이후 시스템이 재부팅되어 메모리가 초기화되기 전까지 여러 번의 암호 연산에 지속적으로 영향을 미치게 된다. 이러한 환경에서는 공격자가 여러 개의 오류가 반영된 암호문을 수집할 수 있으며, 이를 바탕으로 통계적인 방법을 활용해 키 정보를 추출할 수 있다[4].

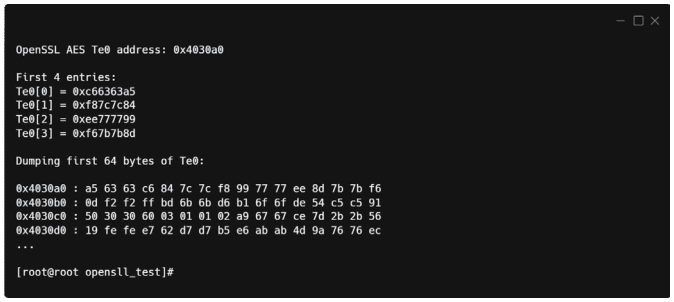
2. 실험환경 및 결과

2.1 실험 환경

DRAM은 DDR4 4GB SO-DIMM 상용 모듈을 Xilinx 사의 FPGA ZCu104 보드([그림 1])에 장착하여 진행했다. OpenSource 기반 DRAM 컨트롤러를 FPGA에 구현하여 DRAM의 row에 대한 직접적인 접근을 가능하게 하였으며 비트 플립이 발생한 row, column의 값과 함께 저장된 데이터를 확인할 수 있도록 구성했다. OpenSSL의 crypto/aes/aes_core.c에서 사용할 때 선언된 T-table Te0[256]의 메모리를 덤프한 결과는 [그림 2]와 같고 리틀 엔디언 방식으로 64바이트씩 정보가 저장됨을 알 수 있다. 이 정보를 참고하여 DRAM에 비슷한 방식으로 T-table 데이터를 저장한 다음 공격에 사용했다.



〔그림 1〕 ZCU 104 보드 구성



〔그림 2〕 OpenSSL aes_core.c의 T-table 메모리 덤프 결과

3.2 실험 결과

전자기파 오류를 주입한 다음 비트 플립이 발생한 row와 변경된 데이터를 확인한 결과 한 개의 row 전체가 0xff로 변경이 되거나 T-table의 데이터의 여러 바이트 또는 한 바이트가 비트 플립으로 인해 변조되는 등의 결과를 확인할 수 있었다. T-table에 비트 플립이 지속적으로 남아있는 것을 확인한 뒤 이후 다량의 암호화 연산을 통해서 오류 암호문을 확보하여 AES 마지막 라운드 키를 복구할 수 있었다.

Ⅲ. 결론

본 논문에서는 오픈소스 암호 라이브러리에서 사용되는 AES T-table 기반 구현이 DRAM에 정적으로 저장될 수 있는다는 점에 주목하고, 실제 데이터가 저장되는 형태와 유사한 환경을 구성한 다음 전자기파 기반 오류 주입을 수행하여 T-table 값이 변조될 수 있음을 실험적으로 확인하였다. 이를 통해 DRAM에 저장되는 암호 테이블에 물리적으로 오류를 주입하였을 때 그 값이 변조되어 AES의 비밀키 값을 유출할 수 있음을 보였다. 향후에는 DRAM에 데이터가 정적 배열로 선언된 상황에서 전자기파 오류주입 등을 활용하여 저장된 영역에 대한 물리 주소 혹은 가상 주소 매핑을 역으로 식별할 수 있는 방법을 연구하고자 한다.

참 고 문 헌

[1] Kim, Yoongu, et al. "Flipping bits in memory without accessing them: An experimental study of DRAM disturbance errors."ACM SIGARCH Computer Architecture News42.3 (2014): 361-372.

[2] 허재원, 박형동, 여인국, 김다연, 권건우, 한동국, “DDR4 DRAM의 소프트웨어 기반 Rowhammer 기법과 전자파 오류주입 기반 Ehammer 기법에 관한 연구,” 2023 한국정보보호학회 하계학술대회, 2023.

[3] Biham, Eli, and Adi Shamir. "Differential fault analysis of secret key cryptosystems." Advances in Cryptology—CRYPTO'97: 17th Annual International Cryptology Conference Santa Barbara, California, USA August 17 - 21, 1997 Proceedings 17. Springer Berlin Heidelberg, 1997.

[4] Zhang, Fan, et al. "Persistent fault analysis on block ciphers." IACR Transactions on Cryptographic Hardware and Embedded Systems (2018): 150-172

[5] Chow, S., Eisen, P., Johnson, H., & van Oorschot, P. (2002). "A White-Box AES Implementation for DRM Applications." In Proceedings of the 4th International Workshop on Information Security (ISW 2002), LNCS 2575, pp. 250 - 270.