A Novel GPU Overdrive Fault Attack 논문 리뷰

https://youtu.be/YP6GYGMf1KU

배경

부채널 공격

- GPU의 부채널 공격
  전력 소비, 전자기 방출 및 마이크로아키텍처 기능(예: GPU의 병합 장치, 뱅크 캐시)의 타이밍 누출

- CPU, FPGA 및 ASIC을 포함한 다른 플랫폼에서 많은 현실적인 결함 공격이 시연
  광학 빔[10], 클록 글리치, 동적 전압 및 주파수 스케일링(DVFS) 활용,
  DRAM에서 Rowhammer를 활용한 액세스 동작

GPU에서 오류 주입 몇 가지 문제

- GPU 장치에 있는 서로 다른 공간 영역은 제조 공정 변동에 따른 결함.

- GPU의 계산 장치 및 저장소에서 실행되는 스레드/명령의 스케줄링 및 매핑에 상당한 비결정성

- 상용 GPU의 명령어 세트와 마이크로아키텍처 세부 사항에 대한 공개 문서는 제한적

배경

A Novel GPU Overdrive Fault Attack

- 상용 GPU에서 VFS 저전력 기능의 안정성 및 보안 영향을 분석
- GPU 오버드라이브 공격과 관련된 문제를 연구하고
  이를 극복하여 최초의 비침습적 GPU 기반 SDC 오류 주입을 시작할 수 있음을 보임
- 이 공격 등급에 대한 서로 다른 GPU 명령의 결함 민감도를 평가
- 키 복구를 목표로 하는 AES GPU 구현에 대한 성공적인 종단 간 오류 공격을 시연

Timing Constraints and Violations

- 대부분의 디지털 디자인에서 플립플롭(FFS)은 사이클 사이의 상태를 유지하는 데 사용되며
  GPU의 SIMD 레인에 상태를 저장하는 데 많이 사용
- FF의 두 등급 사이에 여러 수준의 조합 논리가 있는 표준 순차 회로의 경우
  입력 및 출력의 경우 회로는 다음과 같이 엄격한 타이밍 제약 조건을 충족

$$T_{clk} \geq T_{ff} + T_{prop} + T_{setup} \qquad (1)$$

- Tclk : 클록 주기, Tff: 입력 FF의 지연, Tprop: 조합 논리의 전파 지연, Tsetup: 출력 FF의 설정 시간

오버드라이브

- (1)의 타이밍 제약은 Tclk가 감소되는(즉, 클럭 주파수가 증가하는) 오버클럭킹을 사용하거나
  감소된 공급 전압으로 인해 Tff 및 Tprop가 증가하는 언더볼팅을 사용할 때 위반

- 타이밍 위반으로 인해 출력 FF에서 게이트가 되는 임의의 오류가 발생할 수 있으며, 이를 오버 드라이브라함
  여기서 정확한 오류 값을 알 수 없고 제어할 수 없음

Differential Fault Analysis

- DFA는 비밀을 추론하기 위해 결함 주입 전후에 대상 알고리즘의 출력 상태를 비교하는 일종의 암호 분석
- 피해자 커널로 AES를 선택, Tunstall et al.이 제안한 DFA 모델
- 이 모델은 논문의 오버드라이브 방법론과 일치하는 무작위 결함을 가정

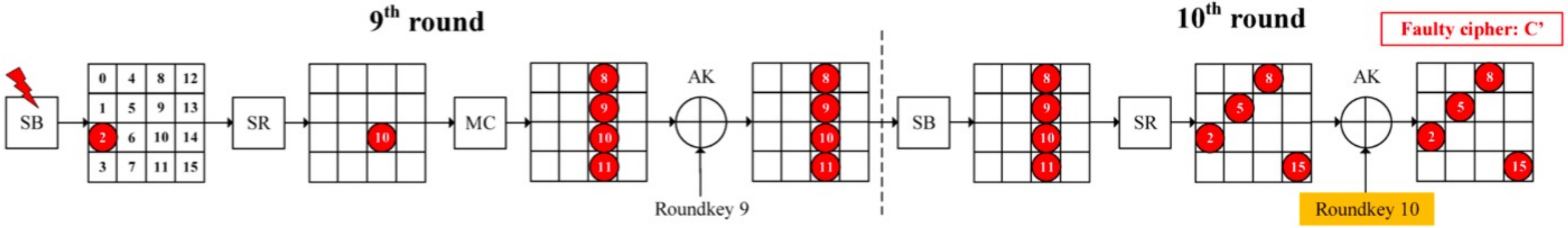

Fig. 3: Our single-byte random fault model and the associated fault propagation to the output.

Differential Fault Analysis

- 오류는 마지막 두 라운드를 통해 전파되고 4개의 잘못된 암호문 바이트가 발생합니다. DFA는 올바른 암호문과 잘못된 암호문 쌍을 사용하여 마지막 라운드 키 바이트 4개를 복구
- DFA에 대한 이전 작업은 주입된 오류가 다른 평문을 사용하는 두 번의 실행에서 동일할 경우 4개의 키 바이트를 복구하는 데 두 쌍만 필요
- 그러나 AES 커널을 실행하는 GPU의 경우 서로 다른 CU에서 실행되는 서로 다른 스레드에서 생성된 결함은 상당히 다름
- 따라서 분산 결함을 주입할 때 작동하도록 기존 DFA 방법론을 조정
  DFA는 키 바이트를 복구하기 위해 더 많은 쌍이 필요

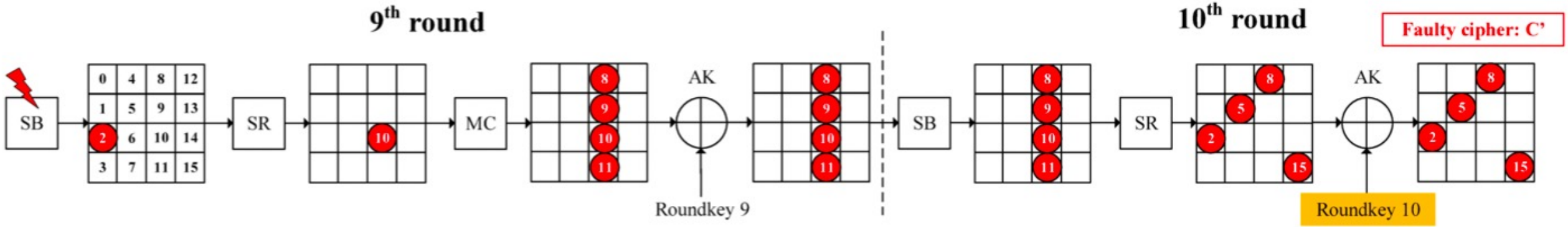

Fig. 3: Our single-byte random fault model and the associated fault propagation to the output.

GPU 전압 및 주파수 확장

- 전력 관리 프레임워크는 에너지를 절약하거나 과열을 방지하기 위해 장치의 사용 패턴 및 작동 조건에 따라 전압과 주파수를 동적으로 조정
- 일반적인 전원 관리 방법에는 동적 전압 및 주파수 스케일링(DVFS) 및 적응형 전압 및 주파수 스케일링(AVFS)이 포함
- DVFS는 공급업체가 개별 DPM(동적 전원 관리) 상태(칩 전압 및 주파수 구성)를 미리 설정하는 개방 루프 스케일링을 사용. 가드밴드(안전 마진)가 있음.
- 환경 및 프로세스 변동을 모두 수용하기 위해 동일한 주파수에 대한 실리콘 한계 근처의 전압과 DPM 상태 사이에서 설정.
- 대조적으로, AVFS는 온다이 하드웨어 메커니즘이 GPU의 여러 위치에서 접합 온도와 전압을 실시간으로 측정하여 전압과 주파수의 저전력 고성능 조합을 적응적으로 선택하는 폐쇄 루프 시스템을 사용
- 이 방법은 기존의 가드 밴드를 제거하여 전력 낭비를 제거합니다. AMD Polaris GPU는 AVFS를 사용
- 장점에도 불구하고 이 전력 관리 기술은 상당한 신뢰성과 보안에 영향을 미치며, 하드웨어가 실리콘 한계 근처의 더 낮은 전압에서 더 높은 주파수에서 작동하도록 과도하게 구동하여 소프트웨어 제어 오버드라이브 기반 데이터 손상을 유도하는 데 악용될 수 있음

적응 확장의 한계

AVFS는 워크로드에 의해 유발된 동적 변화를 적응적으로 처리하도록 설계

- 예를 들어, 컴퓨팅 및 메모리 집약적인 커널이 실행되기 시작하면 GPU의 많은 부분이 활성화되어 갑자기 전원이 소모되어 전압 조정기가 응답할 때까지 일시적인 전압 강하가 발생
- 전원 공급 장치 모니터는 1-2 사이클에서 전압 강하를 감지하고, 클록 스트레칭 회로는 전력 수요의 급격한 변화를 보상하기 위해 주파수를 감소시켜 회로가 드룹 동안 안전하게 작동하도록 도움
- 이 기술에는 두 가지 문제

언더볼팅 데이터 손상

- 클록 스트레처는 2.5%보다 큰 전압 강하에만 응답
- 따라서 무거운 커널이 특정 하드웨어 및 환경에서 2.5% 미만의 전압 드룹을 유발하도록 설계된 경우 응답 회로가 활성화되지 않아 후속 작업이 데이터 손상에 취약

오버클러킹 데이터 손상

- 주파수를 최대 20%까지 줄일 수 있음.
- 다시 말하지만, GPU가 매우 높은 주파수에서 작동하는 경우(예: 오버클러킹을 통해) 클록 스트레처 20% 응답은 GPU의 안전하고 결함 없는 작동을 보장하기에 충분하지 않을 수 있음

오버드라이브 결함 공격

- 언더볼팅과 오버클럭킹 효과를 결합하여 "오버드라이브" 데이터 손상을 일으키는 방식으로 구성
- 공격자가 GPU에서 이 취약점을 악용하여 피해자 커널에서 비밀 데이터를 추출하거나 데이터 무결성에 영향을 미치거나 DoS를 유발할 수 있음을 발견

SDC susceptibility

- 커널 실행 중에 SDC를 주입할 가능성을 평가하기 위해 공급업체에서 규정한 OPP(Operationing Performance Points)이외의 OPP를 선택하고 GPU에서 Sobel 에지 감지 커널을 시작
- OPP 설정은 1245MHz의 클록 주파수와 800mV의 전압을 사용
- 그림 6d에서 정상 출력과 불량 출력의 픽셀별 차이를 보임
- 이 오버드라이브 설정에서 관찰된 총 11개의 결함 픽셀

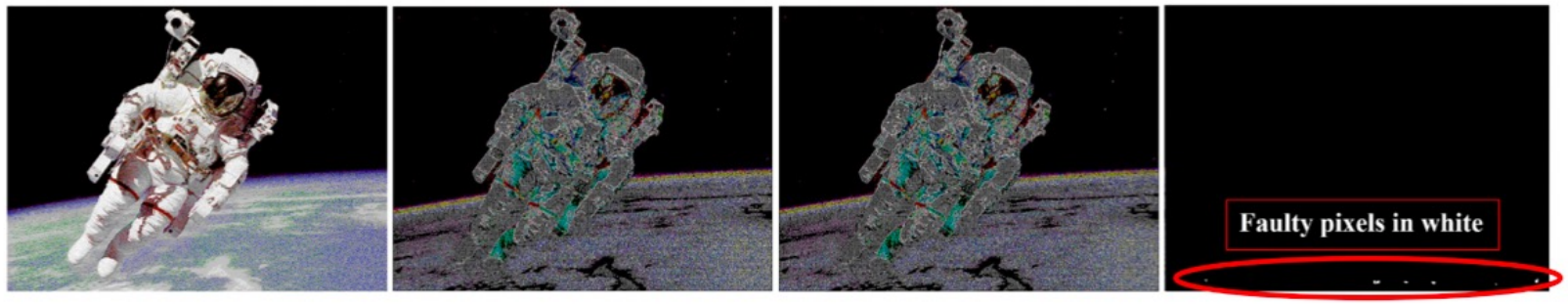

Fig. 6: (a) the input image, (b) the normal Sobel edge output, (c) the faulty Sobel edge output, and (d) the pixel-wise difference of normal and faulty outputs.

Vulnerable Instructions

- 오버드라이빙을 통해 가능한 3가지 일시적인 오류를 모두 고려하고 GPU에서 실행되는 다양한 명령의 취약성을 평가
- (i) 산술/논리, (ii) 제어/분기, (iii) 로드, (iv) 명령어 저장의 4가지 명령어 범주
- 실험과 분석에 따르면 메모리 로드는 명령 범주 중 SDC(silent data corruption) 오류에 가장 취약한 명령
- 동일한 오버드라이브 설정에서 커널에 더 많은 메모리 로드가 포함될 때 더 많은 수의 SDC를 지속적으로 관찰

TABLE I: Different types of GCN instructions and the most common kernel outcome under overdrive fault injection.

Category	Examples	Most common faulty outcome
Arithmetic/logical	S_MUL_I32, S_LSHR_B64, S_XOR_B32, V_MAC_F32	Hang/Crash
Control/branch	S_NOP, S_GETPC, S_BRANCH	Hang/Crash
Store	S_STORE_DWORD, S_BUFFER_STORE_DWORD, FLAT_STORE_UBYTE	Hang/Crash
Load	S_LOAD_DWORD, FLAT_LOAD_UBYTE	SDC/Hang/Crash

AES GPU 구현에 대한 DFA

- 오픈 소스 AES GPU 구현에서 소프트웨어 제어 오버드라이브 기반 DFA
- AES 128비트 ECB 모드, 각 스레드는 일반 텍스트의 블록(16바이트)을 독립적으로 암호화
- SDC 결함이 올바른 위치(즉, 9번째 라운드 SB 출력 상태)에만 주입되도록 하기 위해 희생자 커널에 여러 체크포인트를 삽입하여 비준수 결함이 있는 스레드를 거부
  (예: 8번째 라운드보다 일찍 주입된 결함 MC 출력) 또는 단일 스레드 실행 중에 발생하는 여러 오류. AES 커널은 결함을 대상 AES 상태로 제한하기 위해 일련의 NOP 및 메모리 배터리 명령으로 계측
- 결함이 있는 OPP를 활성화하기 위한 적절한 타이밍 프로파일 495MHz SCLK와 1000mV VDC

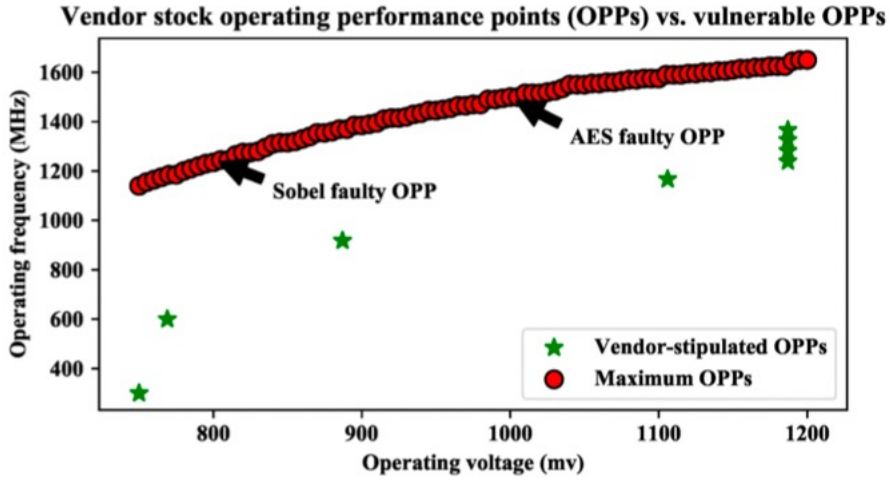

Fig. 5: The AMD Readon RX 580 GPU overdrive characterization. The OPPs that cause faulty outputs for the Sobel edge detection kernel and the AES kernel are also marked.

AES GPU 구현에 대한 DFA

- 공격에서 마지막 라운드 키 바이트(를 암호 바이트 인덱스 A(0,7,10,13), B(1,4,11,14), C(2, 5,8,15) 및 D(3,6,8,12).

- 세트 A의 경우 오류는 MC9 입력의 첫 번째 열로, B는 두 번째 열로, C는 세 번째 열로, D는 네 번째 열로 전파

- 허용 가능한 O 및 O' 쌍을 캡처하기 위해 악의적인 오버드라이브 설정에서 15분 이내에 최대 40회 8192 스레드(128KB 일반 텍스트의 8K 16바이트 블록 AES 암호화에 해당)로 AES 커널을 시작

- 4번의 실행 허용 가능한 SDC(성공적인 실행 최소 3개의 허용 가능한 O)를 생성한 반면 다른 실행은 시스템 중단/충돌 또는 원치 않는 오류로 이어짐

- 허용되는 SDC를 사용하여 14쌍의 O 및 O'(세트 A 및 세트 B의 경우 6개 쌍, 2x3, 세트 C 및 세트 D의 경우 8개 쌍, 2x4)로 16개의 AES 키 바이트를 모두 성공적으로 추출

Limitations and Future Work

- 이 작업의 모든 실험은 AMD GPU에서 수행,
  이 공격의 일반성과 이식성을 입증하기 위해 동일한 결함 주입 방법론을 다른 AMD GPU 제품군과 Nvidia 및 Intel GPU에 적용 예정

- 둘째, 우리는 은밀한 오버드라이브 결함 공격이 중요한 애플리케이션에 심각한 안정성 문제를 제기할 것으로 예상
  딥 러닝, 암호화 해싱 및 커널 서명을 포함한 다양한 유형의 애플리케이션을 평가할 계획

- 마지막으로, 결함 위치와 주입 시간의 무작위성을 줄이기 위해
  피해자 커널을 모니터링하고 공격자가 공격을 시작할 시기를 알려주는
  온디바이스 모니터링을 제공하는 방법을 살펴 볼 예정

Q & A