Side Channel Attack on GPUs

Saoni Mukherjee, Chao Luo, Colleen Finnegan, Yunsi Fei, David Kaeli
Dept. of Electrical and Computer Engineering
Northeastern University
Boston, MA

I. Introduction

Graphic Processing Units (GPUs) have evolved from accelerators for processing graphics and generating high quality games to a platform for general purpose computing. GPUs can accelerate a range of applications including security, pattern matching, business analytics and medical diagnostics. The growing demands of computationally-intensive applications such as cryptography, and the availability of inexpensive many-core architectures, has prompted the security community to consider how best to leverage these accelerators.

A side channel attack (SCA) exploits the physical implementation of a cryptographic system, rather than the inherent theoretical weaknesses of the algorithm itself, they attempt to identify the secret key and crack the cryptosystem. The most widely known SCA techniques include Differential Power Analysis (DPA) and Correlational Power Analysis (CPA). They both target the correlation between intermediate results produced by a cryptographic algorithm and power consumption values.

In this paper we discuss a way to measure power on a GPU and map the result with simulated power measurements while running AES-128 on the GPU. We also discuss our plan on attacking the system through the power metrics to produce the secret information that includes the plain text and the encryption key.

II. Advanced Encryption Standard

The Advanced Encryption Standard (AES) has been adopted by the US government to encrypt data in applications ranging from personal to highly confidential domains [**?**]. In this work we consider execution of the AES algorithm developed in NVIDIA's CUDA, and run on a NVIDIA Kepler GPU [**?**], [**?**], [**?**].

AES has a fixed block size of 128 bits. The key size can vary between 128, 192 and 256 bits. A block is the unit of plain text that the algorithm takes as input and uses to produce the corresponding n-bit cipher text. Text that is longer than a block are divided into multiple blocks, padding the last chunk, and encrypting each block separately. The AES algorithm is comprised of many rounds that ultimately turn plain text into cipher text. Each round has multiple processing steps that include AddRoundKey, SubBytes, ShiftRows and MixColumns. Key bits must be expanded using a precise key expansion schedule.

III. Differential Power Analysis

Differential power analysis (DPA) is an advanced side channel attack, which allows an attacker to extract the secret key of a cryptographic system by statistically analyzing power traces collected from multiple cryptographic operations [**?**]. DPA exploits the correlation of power consumption and the intermediate values of the cryptographic operation, which is dependent on the secret key. The power traces are divided into two sets, based on the intermediate values computed from a secret key candidate. Then the average traces for both sets are computed, and the difference of means (DOM) is computed by subtracting one average from another. The correct key candidate will show the maximum value DOM. In an AES attack, the last round's SubBytes operation is targeted, since this operation is a byte independent operation. Using a long secret key (128, 192 or 256 bits) will involve a brute-force byte-by-byte attack with key candidates. We have previously analyzed power traces using DPA collected on a GPU [**?**].

IV. Measurements

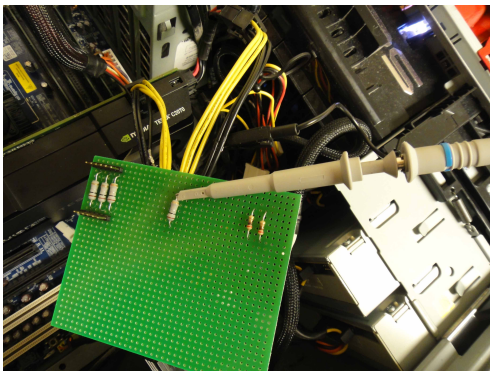

Fig. 1: Setup for collecting traces from a GPU.

Our implementation of AES-128 developed in CUDA has been tested both on an Nvidia GTX 480 and Tesla C2070, running with Ubuntu 14.04.1 on a AMD-64 desktop. We us a LeCory waverunner 640zi oscillascope to obtain power measurements. Power traces are captured by inserting a small resistor (0.1 ohm) in series with the GPU card's external power supply line (multiple power lines are merged into one). The setup shown is shown in Figure 1. The voltage drop across the resistor is recorded when the AES operation is being performed on the GPU, as shown in Figure 2. We have also simulated the results of running AES-128 with GPUWattch

using the same GTX 480 model [**?**]. Figure 3 presents power traces of AES-128 while using GPUWattch.

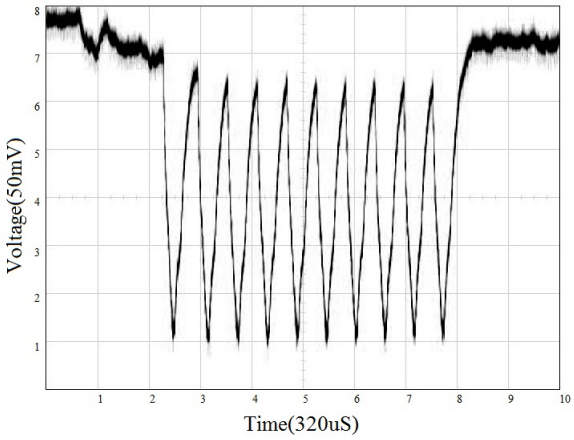

Fig. 2: Traces while running AES 128 on C2070 GPU

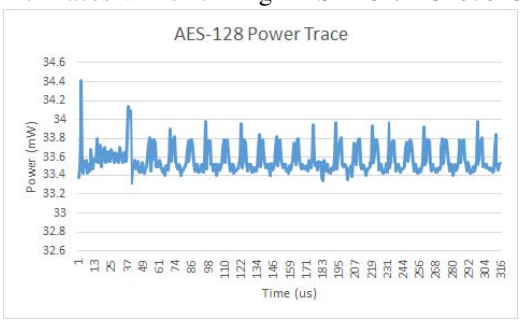

Fig. 3: Traces while running AES 128 on the GPUWattch simulator

The traces are dependent on the data, to some extent. Figure 4 shows the differences computed between the average of all 0 and all 1 data inputs. Here, at time = 20μs, the difference reaches the peak value, which implies that power is highly dependent on the data at this point of the AES execution. The width of the peak is primarily due to the misalignment of each trace when acquired by the oscilloscope. If power was truly independent of the data, Figure 4 would look like random noise.

V. Ongoing Work

Given their performance capabilities, GPUs have become an attractive platform for performing encryption/decryption. We are interested in hardening GPU execution to side-channel analysis. The goals of our work are to acquire simulated power measurements for GPUs, and correlate these results with actual measurements obtained on physical devices. Our final aim is to perform SCA securely on the GPU while using it as the co-processor for encryption.

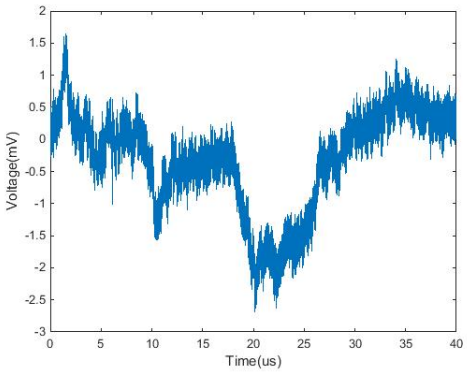

Fig. 4: Difference in traces while changing the input between all 0 and all 1.

References

[1] AMD corporation. Bulk encryption on gpus. In *http://developer.amd.com/resourcxes/documentation-articles/articleswhitepapers/bulk-encryption-on-gpus/*.

[2] Osvaldo Gervasi, Diego Russo, and Flavio Vella. The aes implantation based on opencl for multi/many core architecture. In *Computational Science and Its Applications (ICCSA), 2010 International Conference on*, pages 129–134. IEEE, 2010.

[3] Paul Kocher, Joshua Jaffe, and Benjamin Jun. Differential power analysis. In *Advances in CryptologyCRYPTO99*, pages 388–397. Springer, 1999.

[4] Jingwen Leng, Tayler Hetherington, Ahmed ElTantawy, Syed Gilani, Nam Sung Kim, Tor M Aamodt, and Vijay Janapa Reddi. Gpuwattch: Enabling energy optimizations in gpgpus. In *Proceedings of the 40th Annual International Symposium on Computer Architecture*, pages 487–498. ACM, 2013.

[5] Svetlin A Manavski. Cuda compatible gpu as an efficient hardware accelerator for aes cryptography. In *Signal Processing and Communications, 2007. ICSPC 2007. IEEE International Conference on*, pages 65–68. IEEE, 2007.

[6] National Institute of Standards and Technology (NIST). Fips 197: Advanced encryption standard (aes). 2001.

[7] Tushar Swamy, Neel Shah, Pei Luo, Yunsi Fei, and David Kaeli. Scalable and efficient implementation of correlation power analysis using graphics processing units (gpus). In *Proceedings of the Third Workshop on Hardware and Architectural Support for Security and Privacy*, page 10. ACM, 2014.