Rapid Pre-Characterization of Fine-Grained EM Side-Channel (In)Vulnerability of AES Modules

Vishnuvardhan V. Iyer, Ali E. Yilmaz
The University of Texas at Austin, Austin, TX 78751, USA, (vishnuv.iyer@utexas.edu)

***Abstract*— A statistical method that rapidly identifies near-field probe configurations that are ineffective for leaking information from advanced encryption standard (AES) implementations is proposed. The method can be used as a pre-characterization stage to accelerate a recently introduced multi-stage analysis of variance (ANOVA)-based measurement protocol for evaluating crypto-systems' vulnerability to fine-grained EM side-channel analysis attacks.**

I. Introduction

Sensitive information about cryptographic modules, such as encryption keys, can be recovered by statistically processing the fields they radiate during critical computations [1]-[7]. Indeed, EM side-channel analysis (SCA) attacks that correlate measured signals to hypothesized keys are commonly used in experiments to demonstrate that attackers can recover the actual keys from advanced encryption standard (AES) [8] implementations, even ones hardened against such attacks [1],[5]. Because correlation analysis requires observing a multitude of encryptions for each (and potentially every) possible probe configuration, it is generally infeasible to use only correlation analysis to evaluate AES implementations [5] and verify their (in)vulnerability against fine-grained EM SCA attacks, which use relatively small probes to scan the near fields at a high resolution. Instead, security evaluators—assuming they have more access to/control over the implementation than actual attackers—can use statistical indicators to narrow down the search space, needing potentially far fewer measurements [1].

Methods that accelerate (in)vulnerability detection via fine-grained EM SCA attacks using statistical indicators [1],[6], quantify the exploitable signals and noise for various probe configurations using a few carefully-chosen test cases; e.g., a 3-stage protocol was introduced in [1], where stage I (II) identifies configurations sensitive to measurement (algorithmic) noise by using the analysis of variance (ANOVA) F-statistic with 17 (15) specially designed encryptions. Each stage of the protocol in [1] progressively eliminates ineffective probe configurations, culminating in correlation analyses for a significantly condensed set of probe configurations in Stage III. AES keys are typically analyzed byte-wise to (i) reduce the complexity of key search significantly during correlation analysis ($2^{116}\times$ for AES-128 [8]) and (ii) minimize the marginal cost of future evaluations [5]. Using the same leakage model [1]-[7] as the conventional byte-wise attack, each stage of the protocol in [1] analyzes AES byte-by-byte, evaluating leakage from each sub-block separately.

This article introduces a method that can rapidly identify ineffective probe configurations by using select inputs that exercise all sub-blocks of AES to generate maximum variation in observed fields. These specific encryptions are performed repeatedly, the near fields are measured, and probe configurations sensitive to measurement noise are identified by computing the ANOVA F-statistic.

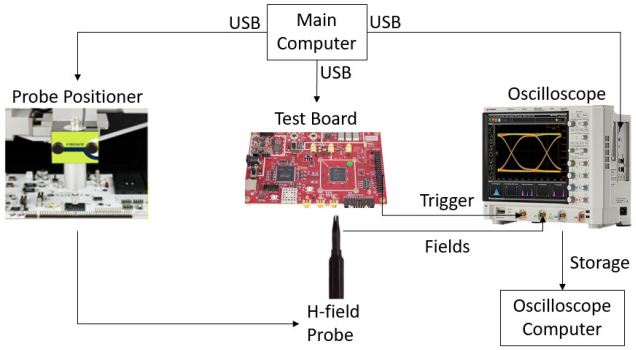

Fig. 1. Fine-grained EM SCA attack setup. Measurements are controlled by the main computer, while analysis is performed in the oscilloscope's computer.

II. Method

A. Generation of Test Cases

The most commonly exploited vulnerability of AES is the dependency of observed fields to transformations in the last round of the algorithm [1], [2]. For an encryption e, the final round of AES-128 performs the following operations on the penultimate round output $\mathbf{oc}_e^9$ to generate the ciphertext $\mathbf{oc}_e^{10}$: First, the *Sbox* operation performs a non-linear one-to-one byte mapping; then, *Shiftrows* operation is linear and re-orders the bytes in a fixed pattern; finally, the final round key $\mathbf{k}^{10}$ is XORed with the intermediate result after the first two steps. All operations are invertible. To perform SCA attacks, it is hypothesized that observed fields depend on the number of bit transitions between the input and output of the final round, i.e., the Hamming Distance (HD) between the two values $\mathbf{oc}_e^9$ and $\mathbf{oc}_e^{10}$. While the analysis is most commonly performed byte-wise, in this article, the entire AES block is considered as a single unit, and therefore the HD is found for all 16 bytes together. It is assumed that the two cases with HD 0 and HD 128 will result in maximum variation in the observed fields. Unlike works where evaluators are assumed to have similar constraints as attackers [5], here the evaluator follows a gold-box threat model [1]. Consequently, by controlling both the input plaintext, as well as the encryption key, the encryptions with HD 0 and HD 128 between the penultimate and final round outputs can be generated easily by using available documentation of AES operations [8]. A method to generate these test cases is summarized next.

For simplicity, all bytes in the penultimate round output are set to 0x00, i.e., $\mathbf{oc}_e^9 = [0x00, \cdots, 0x00]$. Therefore, for the HD 0 case, all bytes in the ciphertext must also be 0, i.e., $\mathbf{oc}_e^{10} =$

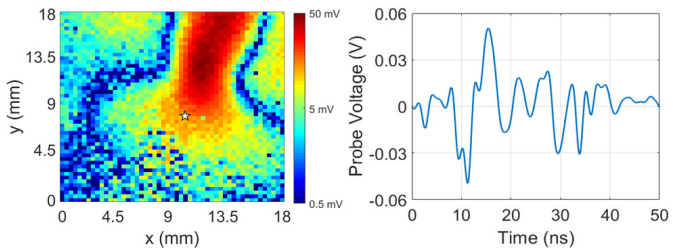

Fig. 2. Spatio-temporal distribution of measured fields using an *x*-oriented probe, for an encryption with HD 0 in the last round of AES. The spatial map is plotted at $t = 12$ ns and time plot is shown for an optimal configuration (star) recovering the first key byte $pc^{0,\text{opt}}$ [1], [2].

$[0\text{x}00, \cdots, 0\text{x}00]$. On the other hand, for the HD 128 case, every single bit must transition from 0 to 1, resulting in each ciphertext byte being set to 0xFF i.e., $\mathbf{oc}_e^{10} = [0\text{xFF}, \cdots, 0\text{xFF}]$.The final round key can be generated for both cases using,

$$k_e^{10,b} = 0\text{x}63 \oplus oc_e^{10,b}, \tag{1}$$

since 0x00 is always mapped to 0x63 by the *Sbox* operation.

B. Measurement Method

Once the 2 encryptions are generated, they are repeated N_r times at each probe configuration pc—combination of location, orientation, and height. For each encryption, the mean observed field $\bar{x}_{\text{HD}_{0/128}}^{pc,t}$ and the variance of observed fields $s_{\text{HD}_{0/128}}^{pc,t}$ are computed across the repetitions at each probe configuration pc and time instant t. If the fields are observed at N_l locations, N_o orientations, and N_h heights, $2 \times N_\text{r} N_\text{l} N_\text{o} N_\text{h}$ encryptions are measured. The ANOVA F-statistic is computed as [1],

$$F_N^{pc,t} = \frac{2N_\text{r} \times \text{Var}(\bar{x}_{\text{HD}_0}^{pc,t}, \bar{x}_{\text{HD}_{128}}^{pc,t})}{\text{Mean}(s_{\text{HD}_0}^{pc,t}, s_{\text{HD}_{128}}^{pc,t})} \tag{2}$$

A high value for F-statistic $F_N^{pc,t}$ represents that the configuration and time instant are highly sensitive to change in the data of interest, while being relatively insensitive to measurement noise. The computed F-values are subjected to null hypothesis testing with a critical threshold $F_{N,\text{c}}$, derived using a confidence interval of 99.9%. The null hypothesis testing is used to generate an indicator [1],[2],

$$Indicator^{pc} = \begin{cases} 0 & \text{if } \max\limits_t F_N^{pc,t} < F_{N,c} \\ 1 & \text{if } \max\limits_t F_N^{pc,t} \geq F_{N,c} \end{cases} \tag{3}$$

Configurations with 0 indicator value are deemed ineffective.

III. Measurement Results

The setup used for measurements in this article is shown in Fig. 1 and detailed in [1], [2]. The proposed method was evaluated on a baseline FPGA implementation of the 128-bit AES algorithm [1]. Observed fields measured with the setup are shown in Fig. 2. High-resolution scans were performed at $N_\text{l} = 51 \times 51$ locations for 2 orientations and 1 height (~0.5 mm above the chip surface). The measurements were repeated $N_\text{r} = 20$ times. For the chosen value of N_r repetitions and 2

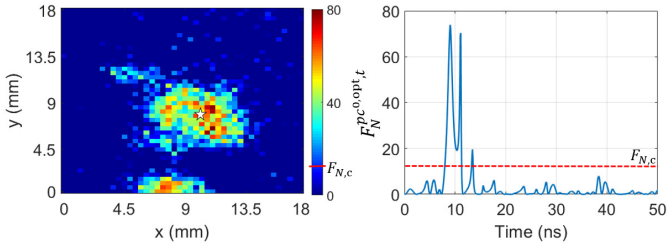

Fig. 3. Spatial map at time 8 ns and time plots of $F_N^{pc,t}$ at an optimal configuration (star).

encryptions, the threshold $F_{N,\text{c}} \sim 13$ and $\sim 2.1 \times 10^5$ encryptions are measured. The computed F-statistic are plotted in Fig. 3, which shows that most of the potential vulnerabilities are close to the center of the chip. Further, no obvious relation between the amplitude of observed fields (Fig. 2) and the computed F-statistics was observed. Overall, the search space of probe configurations was reduced by ~70% using this method.

IV. Conclusion

An ANOVA F-static method, which selects the AES key and inputs to generate the 2 extreme transitions in the ultimate round, is proposed to rapidly identify probe configurations that are insensitive to the computations of interest. The proposed method can be adopted as a precursor to byte-wise AES analysis and is particularly well suited for evaluating AES implementations hardened with physical design strategies [1] that attenuate exploitable signals. For example, if performed as a pre-characterization stage for the protocol in [1], it can make the measurements required to evaluate the shielding-based countermeasure in [1] up to ~10× cheaper. The low cost of the method makes it suitable for performing even higher-resolution measurements with more sensitive probes [5]. The method may also be useful in identifying AES blocks integrated within larger systems, such as processors.

References

[1] V.V. Iyer and A.E. Yilmaz, "An ANOVA method to rapidly assess information leakage near cryptographic modules," to appear in *IEEE Trans. Electromagn. Compat.*

[2] V. V. Iyer and A. E. Yilmaz, "Using the ANOVA F-statistic to rapidly identify near-field vulnerabilities of cryptographic modules," in *Proc. IEEE Int. Microw. Symp.,* June 2021.

[3] L. Sauvage, S. Guilley, and Y. Mathieu, "Electromagnetic radiations of fpgas: High spatial resolution cartography and attack on a cryptographic module," *ACM Trans. Reconfigurable Technol. Syst.*, 2009.

[4] J. Heyszl *et. al*, "Strengths and limitations of high-resolution electromagnetic field measurements for side-channel analysis,'' in Proc. *Int. Conf. Smart Card Res. Adv. Appl.*, Nov. 2012.

[5] V. Iyer, M. Wang, J. Kulkarni, and A. Yilmaz, "A systematic evaluation of EM and power side-channel analysis attacks on AES implementations," in *Proc. IEEE ISI,* Nov. 2021.

[6] J. Danial, D. Das, S. Ghosh, A. Raychowdhury, and S. Sen, "SCNIFFER: low-cost, automated, efficient electromagnetic side-channel sniffing," *IEEE Access*, vol. 8, pp. 173414-173427, Sep. 2021.

[7] A. Kumar, C. Scarborough, A. E. Yilmaz, and M. Orshansky, "Efficient simulation of EM side-channel attack resilience," *in Proc. ICCAD*, pp. 123 – 130, Nov. 2017.

[8] NIST FIPS Pub. "197: Advanced encryption standard (aes)". *Federal information processing standards publication*, 197(441):0311, 2001.