A Systematic Evaluation of EM and Power Side-Channel Analysis Attacks on AES Implementations

Vishnuvardhan Iyer
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, USA
vishnuv.iyer@utexas.edu

Meizhi Wang
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, USA
wang.mz@utexas.edu

Jaydeep Kulkarni
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, USA
jaydeep@austin.utexas.edu

Ali E. Yilmaz
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, USA
ayilmaz@utexas.edu

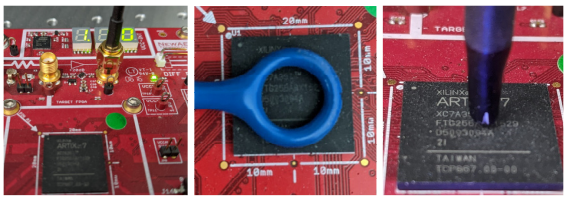
***Abstract*—The effectiveness of coarse- and fine-grained electromagnetic (EM) side-channel analysis (SCA) attacks, as well as power SCA attacks, are empirically evaluated on implementations of the Advanced Encryption Standard (AES) algorithm. Coarse-grained EM and power SCA attacks use a single sensor configuration to measure the aggregated EM emanation or power consumption for a large set of encryptions, and then analyze this set of signals to recover all encryption key bytes. In contrast, fine-grained EM SCA attacks first perform high-resolution scans with relatively small probes in multiple orientations to localize on-chip information leakage, and then use a specific probe configuration for each key byte to collect and analyze signals. The fine-grained EM SCA attacks are found to be up to >70× more effective than coarse-grained EM and power SCA attacks when extracting the key from 3 implementations of 128-bit AES. They are constrained, however, by the potentially prohibitive cost of the initial search to identify effective probe configurations. Search protocols, categorized according to the threat model, to reduce this one-time acquisition cost are presented and are found to require ~8-15× fewer measurements compared to an exhaustive search.**

Keywords—Side-channel attacks, electromagnetic measurements, measurement techniques, analysis of variance

I. Introduction

Side-channel analysis (SCA) attacks defeat cryptosystems by exploiting unintentional information leakage from physical realizations of cryptographic algorithms [1]-[6]; e.g., the Advanced Encryption Standard (AES) cipher key can be recovered from a chip by correlating side-channel signals to bit transitions in state registers [2]. This article focuses on SCA attacks that observe the power consumption or electromagnetic (EM) emanations during critical computations. In power SCA attacks, the observed power consumption is dictated by the aggregate current drawn by electronic logic blocks [1]. In EM SCA attacks, the observed fields depend on the size, location, and orientation of the probes as well as the physical layout of the device under test (DUT) (Fig. 1). In all SCA attack modalities, the information in the observed quantities may be obfuscated by measurement noise as well as by algorithmic noise from uncorrelated system processes [4].

Power SCA setups and coarse-grained EM SCA setups with relatively large probes (Fig. 1(a)-(b)) are commonly used to evaluate hardware security [1]-[3] in part because the setups are relatively easy to implement, requiring a single sensor configuration. In contrast, security evaluations using fine-grained EM SCA setups with relatively small probes (Fig. 1(c)) are rare, more elaborate, and potentially more potent. Because they can localize leakage sources [4]-[8], e.g., via high-resolution scans, these setups can circumvent some countermeasures that are effective against power and coarse-grained EM SCA attacks [1]. While fine-grained EM SCA attacks may require observing far fewer encryptions if they can place probes near relevant signal sources, they become even less effective than their coarse-grained counterparts when probes are away from the sources of interest because EM fields decay rapidly with distance. Thus, the attempt to localize vulnerabilities can accrue substantial measurement costs—also referred to as the "acquisition cost" of fine-grained EM SCA attacks [4]. If the search succeeds, the best configurations can be re-used; indeed, unlike other attacks that are memoryless, fine-grained EM SCA attacks reduce the marginal cost of future attacks on identical or substantially similar physical implementations.



(a) (b) (c)

Fig. 1. Setups for power, coarse-grained EM, and fine-grained EM SCA attacks: (a) A sensor monitoring the aggregate power use of the chip (via the top-right port).(b) A 10-mm diameter H-field probe aggregating fields emanated by sources distributed throughout the chip. (c) A 1-mm diameter H-field probe scanning the chip surface for vulnerabilities.

The acquisition cost of fine-grained EM SCA attacks can be high because the search space for effective configurations includes the transverse location, height, and orientation of a probe [4],[7]. In general, attacks using exhaustive search methods [8], which scan the entire chip/board at a very high resolution with multiple probe orientations, have to limit the number of measurements for each probe configuration in order to be feasible. Recently, various protocols have been proposed to accelerate the search and reduce the acquisition cost [4]-[6]. These search protocols for fine-grained EM SCA attacks should be contrasted carefully because they implicitly assume different restrictions on attackers; e.g., [5] repeated measurements and averaged the captured signals to improve the signal-to-noise ratio. This pre-supposes that attackers have at least partial control over the cryptosystem inputs. Moreover, if the threat model in fact permits attackers to repeat specific inputs, the repeatability of the measurements can be used instead as an indicator to discard configurations [6] and reduce the acquisition cost significantly. Indeed, the restrictions on attackers must be considered explicitly to rigorously analyze and meaningfully contrast the acquisition costs of different fine-grained EM SCA attacks.

This article systematically compares correlation-based power, coarse-grained EM, and fine-grained EM SCA attacks on both baseline and hardened implementations of AES-128. It first evaluates the marginal costs of attacks targeting the AES cipher key using typical measurement configurations for the three attacks. It then introduces a classification for restrictions placed on fine-grained EM SCA attacks that consists of three threat model categories. All of the threat models assume that the attackers have physical access to the DUT and can observe the output ciphertext, but

- the most restricted *black-box threat model* assumes attackers have no access to inputs;
- a less restricted *gray-box threat model* assumes attackers have partial control over inputs, i.e., they can repeat inputs but not observe them; and
- the least restricted *white-box threat model* assumes attackers have full access to inputs, i.e., they can repeat and observe them (the cipher key is unknown).

A different search protocol suitable is presented for each threat model and their acquisition costs are contrasted.

II. SCA Attacks on AES Implementations

A. Correlation Analysis and Its Measurement Costs

SCA attacks correlate switching in the final round of AES to quantities observed in the same time interval [2]-[4]. Specifically, for each byte b of the AES key, (i) the last-round key value is guessed as $0 \le g \le 255$, (ii) ciphertexts are observed and each one's value in the penultimate AES round corresponding to each guessed key is generated, (iii) the Hamming distance between each ciphertext and its penultimate value is computed, and (iv) the results are stored in the arrays $\mathbf{H}^{b,g}$. There are 16×256 such arrays, each storing N_e integers if N_e encryptions are observed.

In the power SCA or coarse-grained EM SCA attack, the aggregate power consumption or EM emanation is recorded during the final round of AES for each encryption; the observed signals are stored in the array $\mathbf{P}^t$ of size $N_\mathrm{t} \times N_\mathrm{e}$ for N_t time samples. Correlating $\mathbf{P}^t$ with the Hamming distances $\mathbf{H}^{b,g}$ yields the correlation coefficients $\rho_{P,H}^{b,g,t}$, for each key-byte b, guess key g, and time instant t. The correct guess key value g^* is identified by using thresholds derived from inverse t-distributions for desired confidence intervals [7] (Fig. 2). The minimum number of encryptions needed to disclose a key-byte b is defined as the "measurements to disclosure" MTD^b. These attacks require

$$Marg.Cost^{\mathrm{Pwr/cgEM}} = \min\left(N_\mathrm{e}^{\max}, \max_b MTD^{b,\mathrm{Pwr/cgEM}}\right) \quad (1)$$

encryptions to be observed, i.e., they observe more and more encryptions until either all bytes are disclosed or a limit on the number of observations, $N_\mathrm{e}^{\max}$, is reached.

In contrast, in the fine-grained EM SCA attack, fields are observed for each encryption and stored in the array $\mathbf{V}^{pc,t}$ of size $N_\mathrm{pc} \times N_\mathrm{t} \times N_\mathrm{e}$; here, N_pc denotes the number of different probe configurations pc —combination of transverse probe location l, height h, and orientation o—used to observe the emanated fields for each encryption. The array is correlated with $\mathbf{H}^{b,g}$ to find $\rho_{V,H}^{b,g,pc,t}$ and to identify $MTD^{b,pc}$, the minimum number of encryptions needed to disclose key-byte b with each probe configuration. As mentioned in the Introduction, some of the probe configurations will be ineffective and will reach the limit on the number of observations, i.e., $MTD_b^{pc} < N_\mathrm{e}^{\max}$ only for some pc. Let $pc^{b,\mathrm{opt}}$ denote the optimal probe configuration such that the key-byte b is disclosed with the minimum number of encryptions

$$mMTD^b = \min_{pc} MTD^{b,pc} \quad (2)$$

Then, the fine-grained EM attack requires observing (at least)

$$Marg.Cost^{\mathrm{fgEM}} = \sum_{b=1}^{16} mMTD^b \quad (3)$$

encryptions *if the optimal probe configurations are known.* Identifying these optimal configurations, however, can be expensive. A naïve approach to find these is to use a high-resolution scan and probe the fields for as many encryptions as feasible [4],[7],[8], i.e., performing the correlation analysis at N_l transverse locations, N_h probe heights, and N_o probe orientations would require [4]

$$Acq.Cost^{\mathrm{Exh}} = N_\mathrm{l} N_\mathrm{h} N_\mathrm{o} N_\mathrm{e}^{\max} \quad (4)$$

encryptions to be observed if $N_\mathrm{e}^{\max}$ encryptions are observed and correlated to $\mathbf{H}^{b,g}$ in each probe configuration. This exhaustive search for the optimal probe configurations can find them only if $N_\mathrm{e}^{\max} > \max_b mMTD^b$. The higher the scan/probe resolution, the smaller $N_\mathrm{e}^{\max}$ has to be, however, for the exhaustive search's acquisition cost to remain feasible.

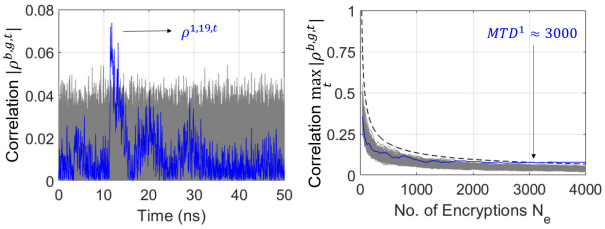


Fig. 2. The correlation coefficient (left) and its maximum value (right) for all 256 guesses (gray) for key byte 1 found by observing the power consumed during 4000 encryptions by an Artix-7 FPGA implementing AES-128. The correlation for the correct guess $g^* = 19$ (blue) crosses the null-hypothesis threshold (dashed) after $MTD^1 = 3000$ encryptions.

Alternative search protocols aim to minimize the acquisition cost, or equivalently, to rapidly isolate probe configurations least affected by noise. Correlation analysis is an effective and simple method to achieve this goal, especially if attackers have limited access to the DUT. If the threat model permits attackers to characterize noise, however, fewer observations will be needed compared to correlation analysis. To show this, let's decompose the fields in $\mathbf{V}^{pc,t}$ as a sum of independent hypothetical quantities representing target signals in $\mathbf{T}^{pc,t}$, measurement noise in $\mathbf{N}^{pc,t}$, and algorithmic noise from background operations in $\mathbf{B}^{pc,t}$ [6]. the correlation coefficients can be represented as [2],[6],[8],

$$\rho_{V,H}^{b,g,pc,t} = \underbrace{\frac{\mathrm{Cov}\left(\mathbf{T}^{pc,t},\mathbf{H}^{b,g}\right)}{\sqrt{\mathrm{Var}\left(\mathbf{T}^{pc,t}\right)\times\mathrm{Var}\left(\mathbf{H}^{b,g}\right)}}}_{\rho_{T,H}^{b,g,pc,t}} \times \frac{1}{\sqrt{1+\underbrace{\frac{\mathrm{Var}\left(\mathbf{B}^{pc,t}\right)}{\mathrm{Var}\left(\mathbf{T}^{pc,t}\right)}}_{\approx 1/F_B^{pc,t}}+\underbrace{\frac{\mathrm{Var}\left(\mathbf{N}^{pc,t}\right)}{\mathrm{Var}\left(\mathbf{T}^{pc,t}\right)}}_{\approx 1/F_N^{pc,t}}}} \quad (5)$$

Clearly, the variance terms degrade the noise-free correlation coefficient $\rho_{T,H}^{b,pc,t}$. The ratio of the variance of target signals to that of measurement (algorithmic) noise—also known as SNR in side-channel security literature [2], [8]—can be estimated from measured fields using the analysis of variance (ANOVA) F-statistic $F_N^{pc,t}$ ($F_B^{pc,t}$) [6], [8]. These F-statistics can be computed with relatively small datasets that characterize noise and thus can rapidly eliminate probe configurations least likely to disclose the key bytes, reducing the number of correlation-analysis measurements. The threat model dictates whether attackers can create such datasets.

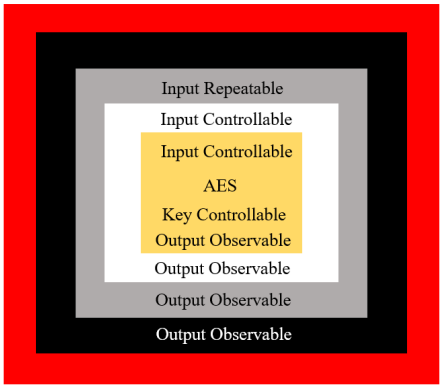


Fig. 3. SCA threat models for AES. Unrestricted attackers (gold box) control the key and have complete access to device peripherals. The access to the DUT is progressively restricted (white, gray, and black box) until attackers have no access to inputs and outputs (red box).

B. SCA Threat Models for AES

The search for the optimal probe configurations in fine-grained EM SCA attacks depends on the threat model. Naturally, all models assume that the cipher key is unknown and the attackers can access the side-channel signals. Starting from the most-restrictive "red box" model, where they have no additional information, and "black box" model, where they can observe the output ciphertext, three progressively less-restrictive threat models can be identified based on the attackers' access to the DUT's peripherals and key (Fig. 3). Search protocols suitable for each model are detailed next.

C. Attacking a Red Box: Pre-characterization Phase

An initial low-cost scan can discard ineffective probe configurations and reduce the search space in fine-grained EM SCA attacks. In this scan, $N_{\mathrm{e}}^{\mathrm{pre}}$ encryptions are observed with each probe configuration pc. The encryptions can potentially be all different; the only constraint is that the same encryption is not repeated $N_{\mathrm{e}}^{\mathrm{pre}}$ times for any pc. Once the observed fields are recorded, $\max_t \mathrm{STD}(\mathbf{V}^{pc,t})$ is computed for each pc. Probe configurations with the smallest standard deviations, close to the noise floor of measurement equipment, can be deemed insensitive to the sources of interest and discarded. This pre-characterization requires

$$Acq.Cost^{\mathrm{pre}} = N_{\mathrm{l}} N_{\mathrm{h}} N_{\mathrm{o}} N_{\mathrm{e}}^{\mathrm{pre}} \quad (6)$$

encryptions to be observed; here, $N_{\mathrm{e}}^{\mathrm{pre}} \ll N_{\mathrm{e}}^{\max}$.

As the AES input and output are not used, this phase can be considered a fine-grained EM SCA attack for the red box threat model. While configurations that give rise to the largest variations in $\mathbf{V}^{pc,t}$ are of interest, these variations can stem from not only the changes in targeted signal sources ($\mathbf{T}^{pc,t}$) but also measurement noise ($\mathbf{N}^{pc,t}$) and algorithmic noise ($\mathbf{B}^{pc,t}$). In general, attackers cannot use just the signals measured during the pre-characterization phase to perform correlation analysis. Instead, this phase enables attackers to rapidly judge if potentially exploitable signals exist, which can reduce the initial search space and acquisition costs of the following measurement protocols.

D. Attacking a Black Box

The black-box threat model, where attackers can observe the outputs but have no access to the inputs or the key, is commonly used for side-channel security evaluation. In this threat model, statistical methods that can rapidly identify probe configurations degraded by noise are unavailable because of the restrictions on the attackers. Search protocols based on correlation analysis [4],[5], including the exhaustive search, can be used; here, the method in [4] is implemented.

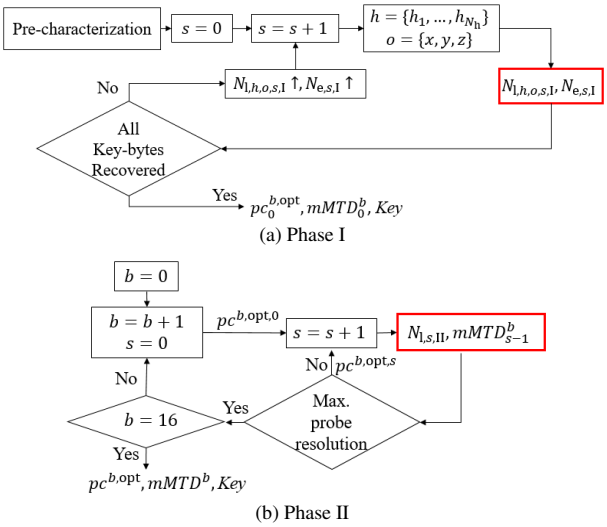


Fig. 4. The fine-grained EM SCA measurement protocol in the black-box threat model. Scans are marked with red and the number of locations and encryptions observed in each scan are specified. Phase I scans are performed with multiple probe orientations, becoming progressively more expensive, while Phase II scans become progressively cheaper.

The measurement protocol for the black-box threat model is an adaptive scan performed in 2 phases: In Phase I (Fig. 4(a)), $N_{\mathrm{scan,I}}$ progressively costlier low-resolution scans are performed to identify the probe configurations $pc_0^{b,\mathrm{opt}}$ that disclose the key-byte b with $mMTD_0^b$ measurements. In each scan s of Phase I, either the number of locations probed $N_{\mathrm{l},h,o,s,\mathrm{I}}$ or number of encryptions observed $N_{\mathrm{e},s,\mathrm{I}}$ is increased [7]. Then, for each key-byte, $N_{\mathrm{scan,II}}$ progressively cheaper scans are performed in Phase II (Fig. 4(b)) to optimize the configurations found in Phase I. Each scan in Phase II uses only the optimal orientations $o_0^{b,\mathrm{opt}}$ at height $h_0^{b,\mathrm{opt}}$, restricts the area of the scan near the optimal locations in the previous scan $l_{s-1}^{b,\mathrm{opt}}$, and observes only the minimum number of encryptions used to disclose the key byte in the previous scan. This requires

$$Acq.Cost^{\mathrm{Bbox}} = Acq.Cost^{\mathrm{pre}} + \sum_{s=1}^{N_{\mathrm{scan,I}}} \sum_{h=1}^{N_{\mathrm{h}}} \sum_{o=1}^{N_{\mathrm{o}}} N_{\mathrm{e},s,\mathrm{I}} N_{\mathrm{l},h,o,s,\mathrm{I}} + \sum_{b=1}^{16} \sum_{s=1}^{N_{\mathrm{scan,II}}} mMTD_{s-1}^b N_{\mathrm{l},s,\mathrm{II}} \quad (7)$$

measurements. In the black-box threat model, this search protocol may converge to local minima for MTDs and not identify the most optimal probe configurations [4].

E. Attacking a Gray Box

The gray-box threat model permits attackers partial control over the input: while they cannot modify or observe the plaintexts, attackers can repeat them. This enables signal averaging to improve the signal-to-noise ratio. It also enables the use of repeatability characterizations and ANOVA F-statistics to prune the search space because probe configurations showing low signal variance for repeated encryptions and high signal variance for changing encryptions are most likely to disclose the keys [6].

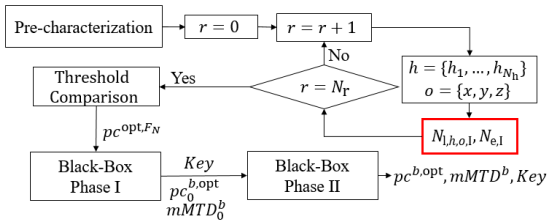


Fig. 5. The measurement protocol in Phase I of the gray-box threat model prunes the search space by repeating scans, computing $F_N^{pc,t}$, and comparing it to a threshold $F_{N,c}$. The reduced set of configurations are then evaluated with the black-box protocol.

The measurement protocol for the gray-box threat model is performed in 3 phases: In Phase I (Fig. 5), one scan per orientation is performed, where $N_{e,I}$ encryptions are repeated N_r times at $N_{l,h,o,I}$ locations. For each encryption e, the sample mean $\bar{x}_e^{pc,t}$ and variance $s_e^{pc,t}$ are computed across the repeated measurements and the F-statistic that quantifies the effect of measurement noise on signals is estimated as [6]

$$F_N^{pc,t} = \frac{N_{e,I} \times N_r \times \text{Var}\left(\bar{x}_1^{pc,t}, \bar{x}_2^{pc,t}, \ldots, \bar{x}_{N_{e,I}}^{pc,t}\right)}{\text{Mean}\left(s_1^{pc,t}, s_2^{pc,t}, \ldots, s_{N_{e,I}}^{pc,t}\right)} \quad (8)$$

The computed values are compared to a threshold $F_{N,c}$ derived from F-distributions for a selected confidence level. Configurations with F-values greater than the threshold are least affected by measurement noise. This model enables attackers to identify configurations significantly degraded by measurement noise (see (5)) and remove them from the search after Phase I. Typically, the resolution of the Phase I scan is higher than its black-box counterpart as it requires fewer encryptions to be observed. Once configurations pc^{opt,F_N} with high F-values are isolated, phases I and II of the measurement protocol for the black-box threat method are performed (Fig. 4). This requires

$$Acq.Cost^{\text{Gbox}} = Acq.Cost^{\text{pre}} + \sum_{h=1}^{N_h} \sum_{o=1}^{N_o} N_{l,h,o,I} N_r N_{e,I} + \sum_{s=1}^{N_{\text{scan,II}}} \sum_{h=1}^{N_h} \sum_{o=1}^{N_o} N_{e,s,II} N_{l,h,o,s,II} + \sum_{b=1}^{16} \sum_{s=1}^{N_{\text{scan,III}}} mMTD_{s-1}^b N_{l,s,III} \quad (9)$$

measurements.

F. Attacking a White Box

The white-box threat model permits attackers complete control over the inputs. The measurement protocol is performed in 4 phases (Fig. 6): Because the key is unknown, Phase I of the protocol for the gray-box threat model is implemented followed by Phase I of the protocol for the black-box threat model to recover the key. In these first two phases, the protocol prioritizes recovering the key over isolating optimal configurations; this allows low-resolution scans to first disclose the key and then further optimize the attack by computing the F-statistic $F_B^{pc,t}$. Because each byte of AES is targetted separately, the algorithmic noise is assumed to come from uncorrelated computations involving the remaining 15 bytes. Although each byte can potentially switch from 256 values in the penultimate round to 256 values in the final output, the Hamming distance (HD) of this transition reduces the number of combinations from 256×256 to 9 values, from HD_0 to HD_8. This simplification is consistent with the HD leakage model used in Section II.A for correlation analysis. For each HD_i of a target byte, $N_{e,III}$ encryptions are performed, where uncorrelated bytes are

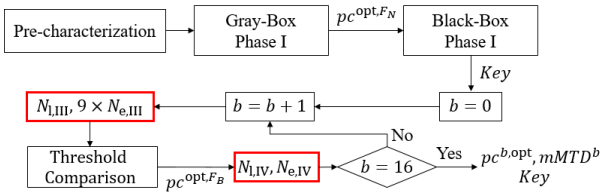


Fig. 6. The measurement protocol in the white-box threat model initially performs Phase I of the protocols used for gray- and black-box threat models. Once the key is disclosed, the search space is pruned by computing $F_B^{pc,t}$ statistic byte-wise and comparing it to a threshold $F_{B,c}$. The reduced set of configurations are then evaluated using correlation analysis.

chosen randomly to increase algorithmic noise. The mean $\bar{\bar{x}}_{\text{HD}_{i_1}}^{pc,t}$ and variance $\bar{s}_{\text{HD}_i}^{pc,t}$ are computed on the averaged signals across the changing encryptions, and the F-statistic $F_B^{pc,t}$ is estimated as

$$F_B^{pc,t} = \frac{9 \times N_{e,III} \times \text{Var}(\bar{\bar{x}}_{\text{HD}_0}^{pc,t}, \bar{\bar{x}}_{\text{HD}_1}^{pc,t}, \cdots, \bar{\bar{x}}_{\text{HD}_8}^{pc,t})}{\text{Mean}(\bar{s}_{\text{HD}_0}^{pc,t}, \bar{s}_{\text{HD}_1}^{pc,t}, \cdots, \bar{s}_{\text{HD}_8}^{pc,t})} \quad (10)$$

In Phase III, $F_B^{pc,t}$ is estimated in a single high-resolution byte-wise scan using configurations identified in Phase II. Comparing the computed values with a threshold $F_{B,c}$ derived from F-distributions enables attackers to remove configurations significantly degraded by algorithmic noise after Phase III. Phase IV subjects optimal configurations pc^{opt,F_B} to correlation analysis. This requires

$$Acq.Cost^{\text{Wbox}} = Acq.Cost^{\text{pre}} + \sum_{h=1}^{N_h} \sum_{o=1}^{N_o} N_{l,h,o,I} N_r N_{e,I} + \sum_{s=1}^{N_{\text{scan,II}}} \sum_{h=1}^{N_h} \sum_{o=1}^{N_o} N_{e,s,II} N_{l,h,o,,s,II} + \sum_{b=1}^{16} 9 N_{e,III} N_{l,III} + \sum_{b=1}^{16} N_{e,IV} N_{l,IV} \quad (11)$$

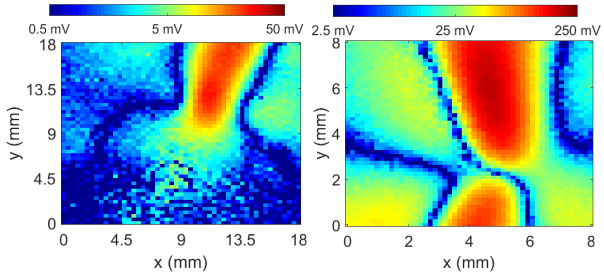
encryptions to be observed. Due to lack of space, the gold-box threat model is not addressed in this article, except to note that unrestricted attackers (controlling the key) can further reduce the acquisition cost by designing specific tests (plaintext-key combinations) that generate extreme variations in target signals to rapidly estimate the F-statistic $F_B^{pc,t}$.

III. Measurement Results

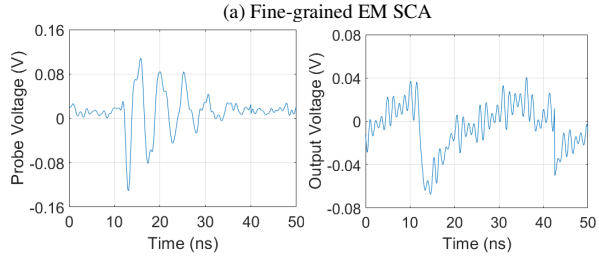
A. Setup

Fine-grained EM SCA attacks were implemented on AES-128 implementations using a 1-mm diameter H-field probe, at a fixed height $h_1 = 0.5$ mm, to scan an 8×8 mm^2 ASIC [1] and an 18×18 mm^2 Artix-7 FPGA [4]. Both chips operated at input clock frequency of 20 MHz and supply voltage of 1.1 V. A Keysight DSOS054A oscilloscope recorded the signals with a sampling rate of 10 GS/s. Analysis was performed locally on the oscilloscope, saving experiment time. The probe was positioned using Riscure's EM probe positioner. The setup allows scanning only in x- and y-orientation, i.e., $N_o = 2$. The search space included $N_l = 51 \times 51$ locations in both orientations. The spatial distributions of measured EM signals are shown in Fig. 7(a). Coarse-grained EM SCA attacks were performed using a 10-mm H-field probe while power attacks were performed using available supply pins on the test boards. Signals captured for power and coarse-grained EM SCA attack are shown in Fig. 7(b)-(c).

In addition to an unsecured AES implementation, the ASIC also used a module hardened against power and coarse-grained EM SCA attacks by using a power delivery mechanism based on the galvanic isolation principle [1]. Galvanic isolation is typically used in high-voltage power converters, where the secondary side of the converter is separated from the primary side to protect it from potentially



(a) Fine-grained EM SCA



(b) Coarse-grained EM SCA (c) Power SCA

Fig. 7. (a) Spatial map of the absolute value of the measured signals using an x-oriented 1-mm diameter H-field probe at ~12 ns during the last round for the FPGA (left) and the secured ASIC (right). N_l= 51×51 locations were probed in both cases. (b) EM signal measured by a z-oriented 10-mm diameter H-field probe positioned at the center of the FPGA. (c) Supply variation of FPGA during the last round of AES operations.

damaging transient voltages and currents [9]. Here, the AES core is isolated from the external power supply to protect the module from power SCA attacks. Reconfigurable capacitor banks are used to supply the necessary charge to perform AES computations. Therefore current signatures and ground bounce in the external supply have minimal data-dependent variance.

B. *Marginal Cost*

First, the marginal costs of EM and power SCA attacks are compared (correlation analysis was performed using the optimal probe configurations or the fine-grained EM SCA attack) to judge their effectiveness. The number of observations with each attack modality was limited to 2 million encryptions; in some cases, the AES key could not be extracted within this limit. The observed marginal costs for all the implementations are listed in Table I. Table I shows that the coarse-grained EM SCA attack was the least effective SCA modality against all the implementations. Surprisingly, the power SCA attack was the most effective against the FPGA (recovering the key with ~2.5× fewer encryptions than the best alternative); this may be because the FPGA and its test board are specifically designed and marketed to study power SCA attacks, i.e., they must have particularly low-noise outputs suitable for the power attack. The fine-grained EM SCA attack required ~3.7× fewer encryptions for the baseline ASIC and >70× times fewer encryptions for the secured ASIC compared to the power SCA attack.

TABLE I. MARGINAL COSTS OF SCA ATTACKS

Marginal Cost	DUT		
	FPGA	Baseline ASIC	Secured ASIC
Power	4.20×10^3	1.00×10^5	$>2.00\times10^6$
Coarse-Grained EM	4.58×10^4	1.48×10^5	$>2.00\times10^6$
Fine-Grained EM	1.05×10^4	2.65×10^4	2.80×10^4

Using the exhaustive search to isolate the optimal probe configurations for the fine-grained EM SCA attack would require ~10^8 measurements for both implementations, if $N_e^{max} = 20\,000$. Next, the results from the search protocols to reduce this cost are reported for the FPGA and the secured ASIC (similar acquisition costs were observed for both secured and unsecured implementations).

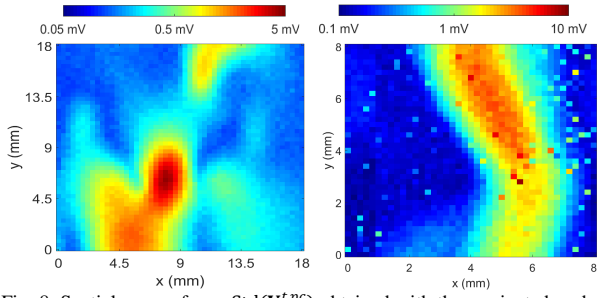
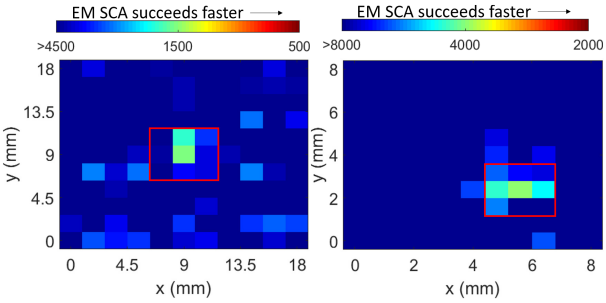
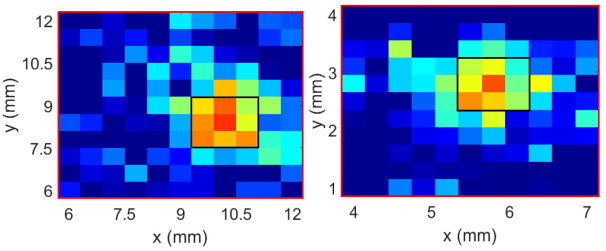


Fig. 8. Spatial maps of $\max_t \mathrm{Std}(\mathbf{V}^{t,pc})$ obtained with the x-oriented probe for the FPGA (left) and ASIC (right).



(a) Phase I scan 2



(b) Phase II scan 1

Fig. 9. MTD maps for byte 1 obtained from the black-box search protocol for the FPGA (left) and ASIC (right) implementations. Scans constrain area (red and black) and number of measurements progressively to reduce cost.

C. *Comparison of Fine-grained EM SCA Protocols*

The pre-characterization (Fig. 8) was performed using $N_e^{pre} = 50$ encryptions for the maximum number of observers on both chips. The signal's standard deviation across the chip was computed and configurations with low variance (< 0.1 mV) were discarded. The pre-characterization showed a significant reduction in the initial search space for the ASIC (~40%) compared to the FPGA (~15%). Before implementing measurement protocols, the configurations eliminated by the pre-characterization phase were noted. If a scan included such a configuration, that measurement was skipped and the probe was positioned at the next configuration.

The protocol for the black-box threat model (Fig. 9) [4], [7] required $N_{scan,I} = 2$ Phase I scans for the FPGA, with the second scan requiring $N_{e,2,II} = 6000$ encryptions per configuration and probed observers on an equally spaced grid of size 11 × 11 over the chip. It required $N_{scan,I} = 2$ Phase I scans for the ASIC, where $N_{e,2,II} = 8000$ encryptions per

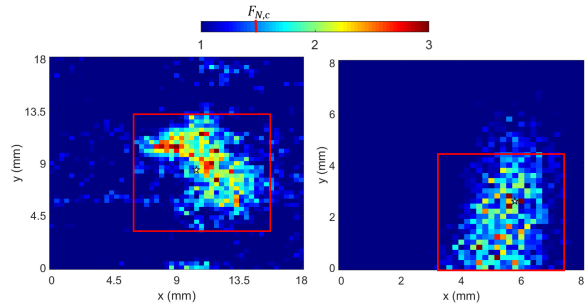


Fig. 10. Spatial map of $\max_t F_N^{pc,t}$ and the are used in subsequent analysis (red) with an *x*-oriented probe for the FPGA (left) and ASIC (right).

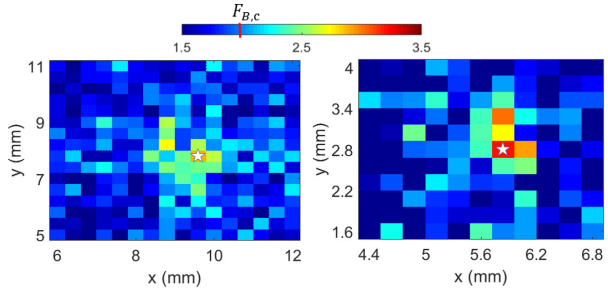


Fig. 11. Spatial map of $\max_t F_B^{pc,t}$ compared to optimal configurations (star) for the FPGA (left) and ASIC (right).

configuration were used in the second scan. Both implementations required $N_{scan,II} = 2$ scans to disclose all bytes of the key.

Attacks using the gray-box protocol first computed F-statistic $F_N^{pc,t}$ for configurations within the search space reduced by pre-characterization. To compute the F-statistic, $N_{e,I}$ =20 encryptions were repeated $N_r = 50$ times [4]. As shown in Fig. 10, comparing the values with the critical threshold 1.6 (confidence level 95%), several non-optimal configurations were discarded. Phases II and III implemented the black-box search protocol over a reduced area, using $N_{scan,II} = 1$ and $N_{scan,III}$=2 scans.

Attacks using the white-box protocol started with the pre-characterization and Phase I for the gray-box model. Phase II performed a low-resolution scan with $N_{l,1,II} = 6 \times 6$, in the region marked in Fig. 10. Once the final round keys were identified, inputs were provided to the chip such that for each variation of Hamming distance switching of an output byte, $N_{e,III} = 20$ encryptions were generated to compute the $F_B^{pc,t}$ statistic (Fig. 11) in Phase III. The statistic was computed at a comparatively finer resolution for the FPGA since a larger region was observed to leak information in previous phases.

D. Acquisition Cost Comparison

The pre-characterization stage required ~2.6×10^5 encryptions for both AES implementations. The acquisition costs were ~9.9×10^6, ~7.3×10^6, and ~6.9×10^6 (~1.27×10^7, ~9.8×10^6, and ~6.8×10^6) measurements for the FPGA (ASIC) when the black-, gray-, and white-box threat model was used. The number of probe configurations and the accumulation of the acquisition cost at each phase of the search protocols are shown in Figs. 12(a)-(c). The final acquisition costs are compared to that of the exhaustive approach in Fig. 12(d). Compared to the exhaustive search, the search protocols for the black-, gray-, and white-box threat models showed ~8-10×, ~10-13×, and ~14-15× cost reduction. The search

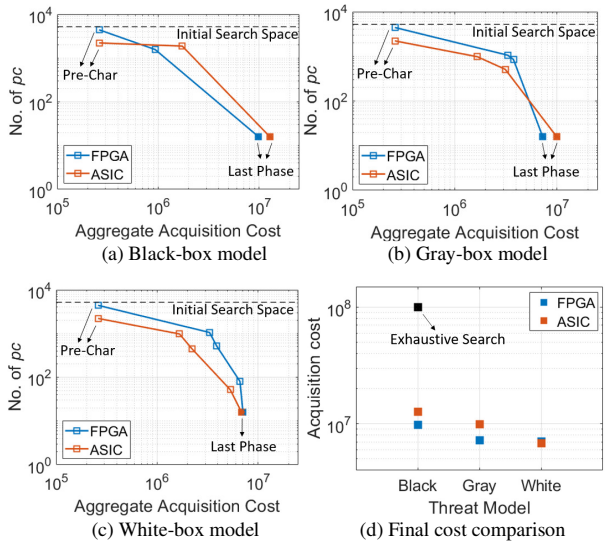


Fig. 12. Reduction of the search space for optimal probe configurations. The optimal configurations were more rapidly isolated for the less restrictive threat models.

protocols for the gray- and white-box threat models required ~1.3-1.35× and ~1.5-2× fewer measurements compared to that for the black-box one, respectively.

IV. Conclusion

Fine-grained EM SCA attacks were systematically compared to coarse-grained EM and power SCA attacks. Though fine-grained EM SCA attacks were found to be more than 70× effective compared to the alternatives on AES-128, they are constrained by the potentially infeasible acquisition cost of the measurements. Various threat models were introduced to categorize search protocols that can rapidly isolate optimal probe configurations in fine-grained EM SCA attacks. Experiments showed that different search protocols can reduce the acquisition cost compared to an exhaustive search by ~8-15×. These protocols enable designers to rapidly evaluate the security o cryptographic modules that implement EM and power SCA countermeasures.

References

[1] M. Wang, *et al.*, "Galvanically isolated, power and electromagnetic side-channel attack resilient secure AES core with integrated charge pump based power management," in *Proc. IEEE CICC*, Apr. 2021.

[2] S. Mangard, E. Oswald, and T. Popp, *Power analysis attacks*, Springer Science & Business Media, 2008.

[3] A. Singh *et al.*, "Improved power/EM side-channel attack resistance of 128-bit AES engines with random fast voltage dithering," *IEEE J. Solid-State Circ.*, vol. 54, pp. 569–583, Feb. 2019.

[4] V. V. Iyer and A. E. Yılmaz, "An adaptive acquisition approach to localize electromagnetic information leakage from cryptographic modules," in *Proc. IEEE Texas Wireless Symp.*, Mar. 2019.

[5] 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. 2020.

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

[7] V. V. Iyer, "An adaptive measurement protocol for fine-grained electromagnetic side-channel analysis of cryptographic modules," M.S. thesis, Univ. of Texas, Austin, Aug. 2019.

[8] F. Unterstein, *et al.* "Dissecting leakage resilient prfs with multivariate localized em attacks," in *Proc. COSADE*, Jul. 2017.

[9] N. Mohan, T. M. Undeland, and W. P. Robbins, *Power Electronics: Converters, Applications, and Design,* 3rd Edition, Wiley, 2002.