# Entropy-based Power Attack

Houssem Maghrebi, Sylvain Guilley, Jean-Luc Danger, Florent Flament

# Entropy-based Power Attack

Houssem Maghrebi, Sylvain Guilley, Jean-Luc Danger, Florent Flament
Département COMELEC
Institut TELECOM, TELECOM ParisTech, CNRS LTCI (UMR 5141), 46 rue Barrault
75 634 Paris Cedex, France
⟨ **houssem.maghrebi@telecom-paristech.fr** ⟩

***Abstract*—Recent works have shown that the mutual information is a generic side-channel distinguisher, since it detects any kind of statistical dependency between leakage observations and hypotheses on the secret. In this study the mutual information analysis (MIA) is tested in a noisy real world design. It indeed appears to be a powerful approach to break unprotected implementations. However, we observe that the MIA fails when applied on a DES cryptoprocessor with masked substitution boxes (Sboxes) in ROM. Nevertheless, this masking implementation remains sensitive to Higher-Order Differential Power Analysis (HO-DPA). For instance, an attack based on a variance analysis clearly shows the vulnerabilities of a first order masking countermeasure. We propose a novel approach to information-theoretic HO attacks, called the Entropy-based Power Analysis (EPA). This new attack gives a greatest importance to highly informative partitions and in the meantime better distinguishes between the key hypotheses. A thorough empirical evaluation of the proposed attack confirms the overwhelming advantage of this new approach when compared with MIA.**

***Index Terms*—Side-channel attack, masking countermeasure, Mutual Information Analysis (MIA), High-Order Differential Power Analysis (HO-DPA), Variance-based Power Attack (VPA), Entropy-based Power Analysis (EPA), FPGA.**

## I. Introduction

Side Channel Analysis (SCA) is a cryptanalytic technique that consists in analyzing the physical leakage produced during the execution of a cryptographic algorithm embedded on a physical device.

The most classical distinguishers used in SCA are Kocher *et al.*'s original DPA [11] and correlation attacks using Pearson's correlation coefficient, introduced by Brier *et al.* [2].

Another important distinguisher is used in the so-called Template Attacks (TA) [4]. TAs use maximum-likelihood as similarity measure, that can capture any type of dependency between its predictions and the leakage measurements (if the probabilistic model is found to be adequate), whereas, for example correlation analysis only captures linear dependencies.

In 2008, another interesting side-channel distinguisher has been proposed, denoted as Mutual Information Analysis (MIA) [8]. It is an attractive alternative to the aforementioned attacks since some assumptions about the adversary can be relaxed. In particular it does not require a linear dependency between the leakage and the predicted data, as it is the case for DPA and CPA, and so it is able to exploit any kind of dependency but also without needing to profile the leakage as it is the case for TA.

The MIA has been largely studied and tested on unprotected implementations [8], [16], [22]. The evaluations performed to prove the efficiency of MIA in protected implementations are incomplete as they are based on simulations or on a limited implementation of the algorithm (*e.g.* only table look-ups in [7]). This motivates the study of the MIA in the context of protected implementation based on the masking countermeasure. The idea of masking is to conceal intermediate values through arithmetic or Boolean operations with random values, which makes it extremely chancy to correctly predict the intermediate sensitive variables [1], [3], [9].

The proposed attack studied here is called the Entropy Power Analysis (EPA) using a weighted sum of conditional entropies as a distinguisher. It is designed to promote partitions of high informative content and to ease the distinguisability between hypotheses on candidate keys. It is carried out on a DES coprocessor which is part of a SoC programmed in an FPGA. This attack is compared with the MIA attack and another successful attack so-called Variance Power Analysis (VPA) based on the variance distinguisher proposed in [13], [20].

The rest of the paper is organized as follows. Sec. II presents the state-of-the-art of MIA and the existing estimation methods used to optimize the MIA. The robustness evaluation of the unprotected DES against the MIA is provided in Sec. III. This section includes the description of the ROM DES masked architecture and the results of the MIA attack on it. Sec. IV presents the proposed entropy test. It provides experimental results against the masked ROM implementation and deals with the empirical evaluation of our EPA proposal. Finally, Sec. V concludes the paper and opens some perspectives.

## II. Mutual information

### A. Background

Our EPA is information-theoretic distinguisher. Therefore, we begin with a brief review of information theory.

**The Shannon entropy**

Consider a system $A$ with $n$ possible states. That is, a measurement performed on $A$ will yield one of the possible values $a_1,\ldots,\ a_n$, each with its corresponding probability $p(a_i)$. The average amount of information gained from a measurement that specifies one particular value $a_i$ is given by the entropy $H(A)$ of the system [18]:

$$H(A) \doteq -\sum_{i=1}^{n} p(a_i) \cdot \log p(a_i)\,.$$

As stated by Faser and Swinney [6], the entropy $H(A)$ could be described as the "quantity of surprise you should feel upon

reading the result of a measurement".

The joint entropy $H(A, B)$ of two discrete systems $A$ and $B$ is defined analogously:

$$H(A, B) \doteq -\sum_{i=0}^{n}\sum_{j=0}^{m} p(a_i, b_j) \cdot \log p(a_i, b_j).$$

Here $p(a_i, b_j)$ denotes the joint probability that $A$ is in state $a_i$ and $B$ is in state $b_j$. The number of possible states $n$ and $m$ may be different. If the systems $A$ and $B$ are statistically independent the joint probabilities factorize and the joint entropy $H(A, B)$ becomes:

$$H(A, B) = H(A|B) + H(B), \tag{1}$$

with $H(A|B)$ being called the conditional entropy, defined as:

$$H(A|B) \doteq -\sum_{i=0}^{n}\sum_{j=0}^{m} p(a_i, b_j) \cdot \log p(a_i|b_j),$$

instead of Equation (1). The mutual information $I(A; B)$ between the systems $A$ and $B$ is then defined as [12], [18]:

$$I(A; B) \doteq H(A) + H(B) - H(A, B) \geq 0.$$

**The differential entropy**

We now introduce the concept of differential entropy, which is the entropy of a continuous random variable [15]. Differential entropy is similar in many ways to the entropy of a discrete random variable (The Shannon entropy). But there are some important differences, and there is need for some care in using the concept.

The differential entropy $H(X)$ of a continuous random variable $X$ with a density $f(x)$ is defined as:

$$H(X) \doteq -\int_S f(x) \cdot \log f(x)\,\mathrm{d}x,$$

where $S$ is the support set of the random variable. Unlike discrete entropy, differential entropy can be *negative*.

We now extend the definition of the mutual information $I(X, Y)$, to probability densities.

The mutual information $I(X; Y)$ between two random variables with joint density $f(x, y)$ is defined as:

$$I(X; Y) \doteq \iint f(x, y) \cdot \log \frac{f(x, y)}{f(x)f(y)}\,\mathrm{d}x\mathrm{d}y\,.$$

The properties of $I(X; Y)$ are the same as in the discrete case.

### *B. Probability density function estimation*

In parametric density estimations, we choose some distribution (such as the normal distribution or the extreme value distribution) and estimate the values of the parameters appearing in these functions from the observed data. However, often the functional form of the true density function is not known. In this case, the probability density function can be estimated non-parametrically, by using histogram or kernel density estimation, or parametrically by using the Expectation Maximization (EM) algorithm.

TABLE I
SOME KERNEL FUNCTIONS FOR PDF ESTIMATION.

| **Kernel name** | **Function** $k(t)$ | **Optimal bandwidth** $h$ |
|---|---|---|
| **Uniform** | $\frac{1}{2}i(t)$ | $\sigma(\frac{12\sqrt{\pi}}{n})^{1/5}$ |
| **Triangle** | $(1-\lvert t\rvert)i(t)$ | $\sigma(\frac{64\sqrt{\pi}}{n})^{1/5}$ |
| **Epanechnikov** | $\frac{3}{4}(1-t^2)i(t)$ | $\sigma(\frac{40\sqrt{\pi}}{n})^{1/5}$ |
| **Triweight** | $\frac{35}{32}(1-t^2)^3 i(t)$ | $\sigma(\frac{25200\sqrt{143\pi}}{n})^{1/5}$ |
| **Gaussian** | $\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}t^2)$ | $\sigma(\frac{4}{3n})^{1/5}$ |

**Histogram method**

Histograms are commonly used to represent a statistical distribution. To calculate a histogram, we divide the data into bins of size $h$, and count the number of data in each bin. For relatively simple distribution, reasonable choices of $h$ are Scott's rule ($h = 3.49 \times \hat{\sigma}(x) \times n^{-1/3}$, where $\hat{\sigma}$ is the empirical standard deviation and $n$ is the number of bins) and Freedman-Diaconis' rule [23]. More generally, by varying $x$ we can estimate the probability density function $f(x)$ as a function of $x$.

**Kernel density estimation**

The probability is estimated as:

$$f(x) = \frac{1}{nh}\sum_{i=0}^{n} k\left(\frac{x - x_i}{h}\right),$$

where the function $k$ is the kernel function and $h$ is called the bandwidth or smoothing parameter.

Some commonly used kernel functions are listed in the table I (all refer to [19]), where $i$ is a step function defined as $i(t) = 1$ if $|t| \leq 1$, 0 otherwise.

**Parametric estimation**

If we considered the observations $x_i$ to be a mixture of Gaussians, the parametric method models the probability density function as:

$$f(x) = \sum_{i=0}^{n-1} w_i \cdot N(x, \mu_i, \sigma_i)\,,$$

where the $w_i$, $\mu_i$ and $\sigma_i$ are respectively the weight, the mean and the standard deviations of each component. An efficient algorithm called the Expectation Maximization algorithm [5] allows one to give good approximation of a probability density function in the form of a finite mixture.

## III. PRACTICAL MIA ATTACK

To test the MIA in a real-life context, we performed it against two DES hardware implementations. The first one is an unprotected DES and the second one is a full-fledged masked DES using a ROM in an Altera Stratix II FPGA on the SASEBO-B evaluation board provided by the RCIS [17].

### *A. Attack on unprotected DES*

We expressed the Hamming weight (HW) as $HW(\Delta(x, k)) = HW(x \oplus S(x \oplus k))$, where $k$ is the key, $x$ is the initial data and $S(x \oplus k)$ is the final register value. Considering 4-bit registers, there are five possible distributions depending on the $HW(\Delta(x, k))$ values.

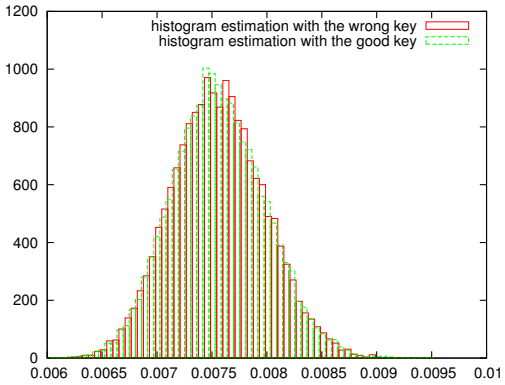


Fig. 1. Histogram method used to distinguish two key hypotheses.

TABLE II
CONDITIONAL ENTROPY ESTIMATION.

| Conditional entropy | Good key | Bad key |
|---|---|---|
| Histogram method | $-9.16542 \pm 2 \times 10^{-5}$ | $-9.16367 \pm 2 \times 10^{-5}$ |

Before performing the MIA attack, we tried to estimate the probability density function (PDF) of the first DES S-box when $HW(\Delta(x,k)) = 0$. We use two hypothesis of the key, the first is right and the second is bad. We carried out the PDF estimation for both cases, using real power consumption measurements of our circuit.

Fig. 1 plots the estimations of the PDF using the good and wrong key prediction when applying the histogram method. One Gaussian PDF seems to be estimated when using the good key prediction, whereas a mixture of two Gaussian distributions seems to be estimated when using the wrong key prediction. Moreover, we summarize the estimation of the conditional entropy in each case in the table II and we validated that the estimated conditional entropy is minimum for the good key, so the MIA attack can be carried out on the unprotected DES. We have estimated the accuracy of $2 \times 10^{-5}$ bit as the quadratic error with respect to the theoretical value $\log_2(\hat{\sigma}\sqrt{2\pi e})$, where $\hat{\sigma}$ is the empirical standard deviation.

We performed the MIA attack with the histogram estimation method with the following procedure:

1) Apply $n$ plaintext messages ($x_i$, $i \in [1,n]$) and collect $n$ observations of power consumption (traces $O_i$).
2) Compute the entropy of the observations $H(O)$.
3) For each S-Box, make assumptions about the key $k \in [0,63]$:
   - Sort the traces $O_i$ to get five activity partitions $set_l$, $l \in [0,4]$, corresponding to the five $HW(\Delta(x,k)) = l$ possible values.
   - Compute the conditional entropy $H(O|HW(\Delta(x,k)) = l)$ for each $set_l$.
   - Compute the mutual information $MIA_k$, as the difference between the observations entropy and the sum of the conditional entropy weighted with the probability $p_l$: $MIA_k = H(O) - \sum_{l=0}^{4} p_l \times H(O|HW(\Delta(x,k)) = l)$.
4) The correct guess of the key $k$ corresponds to $\underset{k}{\operatorname{argmax}}\, MIA_k$.

The MIA is tested on 50,000 traces of an unprotected DES implementation. Fig. 2 shows the mutual information values according to each key predicted for the first DES S-Box. Consequently, the round subkey guessed by the MIA attack is the key corresponding to the highest mutual information.

The eight DES S-Boxes subkeys used during the first round of our DES implementation have been guessed by the MIA attack. The number of traces to break the key is given in Tab. III. In the next subsections we try to answer the question: Is the masked DES sensitive to the MIA attack?

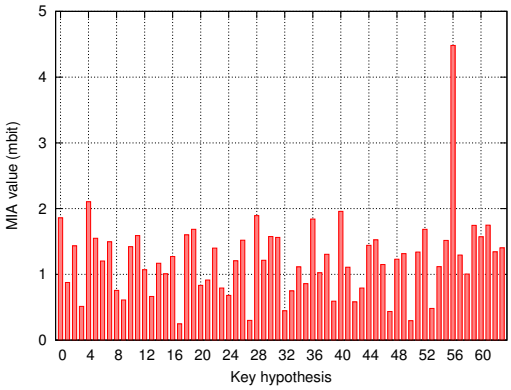


Fig. 2. MIA results on 50,000 power consumption traces of an unprotected DES implementation. The correct key is $k = 56$.

TABLE III
NUMBER OF MEASUREMENTS TO DISCLOSE (MTD) THE SUB-KEY FOR EACH S-BOX OF THE UNPROTECTED DES MODULE ON THE FIRST ROUND.

| Sbox # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| MTD | 12,133 | 10,827 | 12,974 | 12,317 | 11,034 | 13,578 | 10,651 | 11,635 |

## B. First order masking

The masking technique relies upon the concealment of internal sensitive variables $x$ by a mask $m$ which takes random values, in order to avoid the correlation between the cryptographic device's power consumption and the data being processed [3], [9]. The internal variable $x$ does not exist as a net in the cryptosystem but can be reconstructed by a pair of signals ($m$, $x_m = x\ \theta\ m$) where $x_m$ is the masked variable and $\theta$ is an operation which can be Boolean or arithmetic. In the sequel, we consider the masked DES studied at UCL [21]; our variable $x$ represents the right half of the LR register.

At each round an intermediate mask $ML_i$, $MR_i$ is calculated in parallel with the intermediate cipher word $L_i$, $R_i$

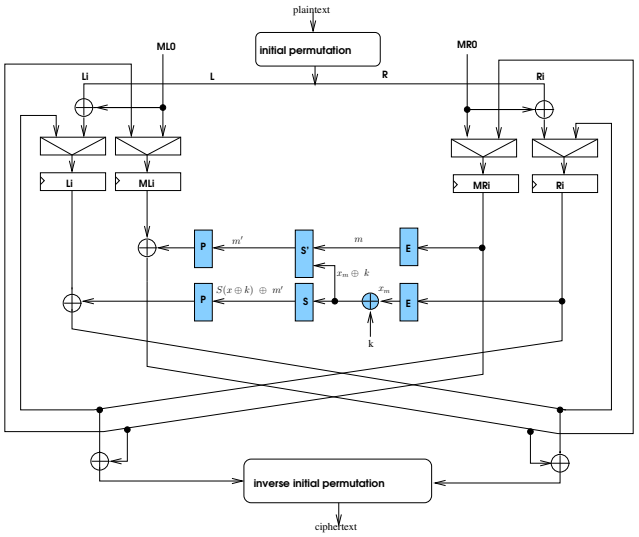


Fig. 3. Masked DES with S-boxes $S$ and $S'$ in ROM.

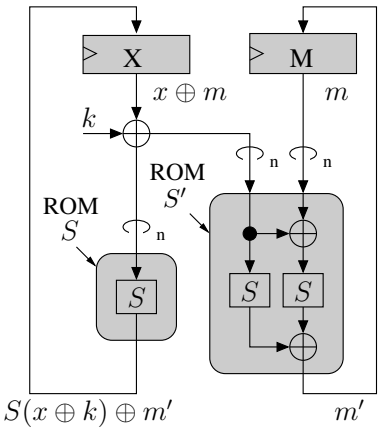


Fig. 4. Masked DES using two paths, implemented with ROMs.

as shown in Fig. 3. If we ignore the expansion $E$ and the permutation $P$, the DES round function $f$ is implemented in a masked way by using a set of functions $S$ and a set of functions $S'$:

$$S(x_m \oplus k) = S(x \oplus m \oplus k) = S(x \oplus k) \oplus m' \,,$$
$$m' = S'(x_m \oplus k, m)) = S'(x \oplus m \oplus k, m) \,. \qquad (2)$$

The variable $m'$ is a new mask reusable for the next round.

The set of functions $S$ contains the traditional S-boxes applied on masked intermediate words. The size of each $S$ is 64 words of 4 bits when implemented with a ROM. $S'$ is a new table which has a much greater ROM size of $4K$ words of 4 bits, as there are two input words of 6 bits. The classical Correlation Power Analysis, as well as the Differential Power Analysis, did not allow us to extract a single S-Box subkey used by the cryptoprocessor using up to 100,000 traces. This is because the transient demasking observed in [14] occurs only in combinational logic, as opposed to ROM. Do we have the same results when performing the MIA attack on this ROM masked DES?

### C. Attack on masked DES

In software implementations, the masked data $x_m = x \oplus m$ and the mask $m$ are manipulated sequentially. Therefore, combined attacks can be carried out. Of special interest is the multivariate MIA (MMIA) introduced recently in [7]. It uses $I(O; \Delta(x,k) \oplus \Delta(m), \Delta(m))$ as a distinguisher, where $O$ denotes each observation, $\Delta$ expresses the distance of a register output, *i.e.* $\Delta(x,k) \doteq x \oplus S(x \oplus k)$, $\Delta(m) \doteq m \oplus m'$. However, in a hardware implementation (our study), $x_m$ and $m$ are used simultaneously. The countermeasure is thus referred to as "zero-offset". In this case, the MMIA cannot be applied.

So, we considered the PMF (Probability Mass Function) of the activity corresponding to those of the combined X and M registers of Fig. 4. The activity of these two registers is expressed by:

$$A = HW(\Delta(x,k) \oplus \Delta(m)) + HW(\Delta(m)) \,. \qquad (3)$$

Considering 4-bit registers, there are five possible PMFs depending on the $HW(\Delta(x,k))$ values, when the key is correct, as shown in the top of Fig. 5.

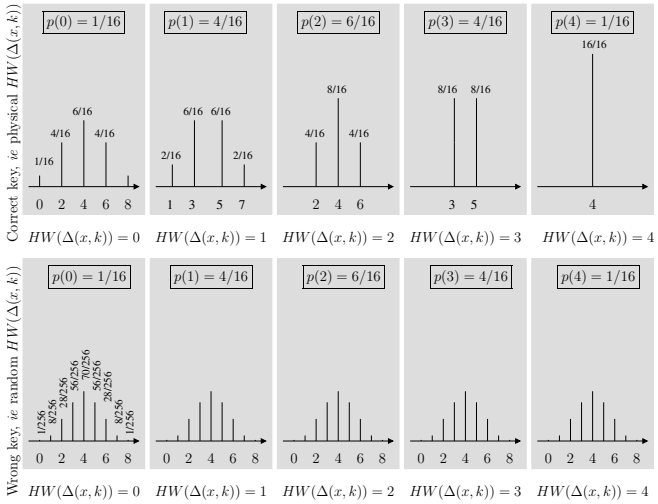


Fig. 5. PMFs corresponding to the five possible values of $HW(\Delta(x,k))$.

TABLE IV
THEORETICAL CONDITIONAL ENTROPY OF THE ROM MASKED DES.

| Theoretical entropies | The correct key | Any wrong key |
|---|---|---|
| $H(O\|HW(\Delta(x,k)) = 0)$ | 2.0306 bit | 2.5442 bit |
| $H(O\|HW(\Delta(x,k)) = 1)$ | 1.8113 bit | 2.5442 bit |
| $H(O\|HW(\Delta(x,k)) = 2)$ | 1.5000 bit | 2.5442 bit |
| $H(O\|HW(\Delta(x,k)) = 3)$ | 1.0000 bit | 2.5442 bit |
| $H(O\|HW(\Delta(x,k)) = 4)$ | 0.0000 bit | 2.5442 bit |
| $H(O\|HW(\Delta(x,k)))$ | **1.3992 bit** | **2.5442 bit** |

When the key is incorrect, the leakage corresponds to that of the function $A$ described in Equation (3) where:

- $x$ is uniformly distributed in $[\texttt{0x0}, \texttt{0xf}]$, because the guessed key is wrong,
- $m$ is uniformly distributed in $[\texttt{0x0}, \texttt{0xf}]$, because the mask is random and unknown by the attacker.

Table IV summarizes the theoretical values of the conditional entropy of each values of $HW(\Delta(x,k))$ in the two cases. Therefore, the entropy, in the case of the bad key, is equal to: 2.5442 bit. Hence a contrast in mutual information of $I(O; HW(\Delta(x, k_{\text{correct}})) - I(O; HW(\Delta(x, k_{\text{incorrect}}))) = \cancel{H(O)} - H(O|HW(\Delta(x, k_{\text{correct}}))) - \cancel{H(O)} + H(O|HW(\Delta(x, k_{\text{correct}}))) = -1.3922 + 2.5442 = 1.1520$ bit. So, theoretically, with ideal S-Boxes, the MIA attack can succeed on a zero-offset implementation: it does distinguish the correct key guess from the wrong ones. However, it clearly appears that the partitions do not contribute equally to disambiguating the correct key from the incorrect ones: the smaller the value $H(O|HW(\Delta(x,k)))$ the better the entropy difference. This noting is the first motivation to devise an improved version of the MIA.

Additionally, we simulated the computation of the conditional entropy $H(O|HW(\Delta(x,k)))$ for all key hypotheses. Figure 6 describes the result for the 4th S-Box using good key equal to 38. We observe that the conditional entropy $H(O|HW(\Delta(x,k)))$ for some keys is under the theoretic value 2.5442 bit. We explain this result by the fact that the activity of the S-box itself leaks the information and decreases

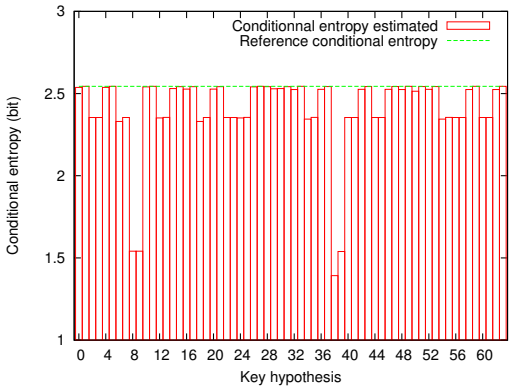


Fig. 6. Comparison between the reference and the estimated conditional entropy for each key in sbox #4. The correct key is $k = 38$.

TABLE V
CONDITIONAL ENTROPY ESTIMATION OF THE MASKED DES.

| Conditional entropy | Good key |
|---|---|
| $HW(\Delta(x,k)) = 0$ | 23.2842549755 |
| $HW(\Delta(x,k)) = 1$ | 23.2460651542 |
| $HW(\Delta(x,k)) = 2$ | 23.2185655678 |
| $HW(\Delta(x,k)) = 3$ | 23.189564065 |
| $HW(\Delta(x,k)) = 4$ | 23.1286079923 |

the entropy. This phenomenon had already been observed in [2] and referred to as "ghost peaks", characterized in [10]. It is a second compelling reason to define an upgraded version of the MIA. So, in practice, it is harder than expected to discriminate the right key used by the MIA attack on the context of zero-offset implementation.

In a real application, the noise coming from other computing blocks and the environment shapes the PMF shown in Fig. 5 as a sum of Gaussian distributions. We reproduce the MIA attack described in Sec. III-A on the masked DES with sboxes in ROM; the attack failed even with up to 200,000 power consumption traces.

## IV. EVALUATION OF A NEW ENTROPY-BASED ATTACK

### A. Proposed Entropy-based Power Analysis (EPA)

By choosing a fixed and appropriate (key, message) couple in regard to a specific S-Box, the distribution of power consumption has the same mean, but different conditional entropy as shown in table V. For instance the conditional entropy difference between the PDF for $HW(\Delta(x,k)) = 0$ and $HW(\Delta(x,k)) = 4$ should be enough discriminating, since it is maximum when using the good key. This leads us to define accordingly the *Entropy-based Power Analysis* or *EPA* which is an improved partition distinguisher. The EPA is a combination of conditional entropies of the power consumption traces computed during the first DES round, while ciphering random messages. The *EPA* algorithm is made explicit below:

1) Apply $n$ plaintext messages ($x_i$, $i \in [1,n]$) and collect $n$ observations of power consumption (traces $O_i$).
2) For each S-Box, make assumptions about the key $k \in [0,63]$:
   - Sort the traces $O_i$ to get five activity partitions $set_l$, $l \in [0,4]$, corresponding to the five $HW(\Delta(x,k)) = l$ possible values.
   - Compute the conditional entropy $H(O|HW(\Delta(x,k)) = l)$ for each set $set_l$.
   - Compute a EPA indicator $H_k$ being a linear combination of the conditional entropy with weights $w_l$: $H_k = \sum_{l=0}^{4} w_l \times H(O|HW(\Delta(x,k)) = l)$.
3) The correct guess of the key $k$ corresponds to $\underset{k}{\operatorname{argmax}}\, H_k$.

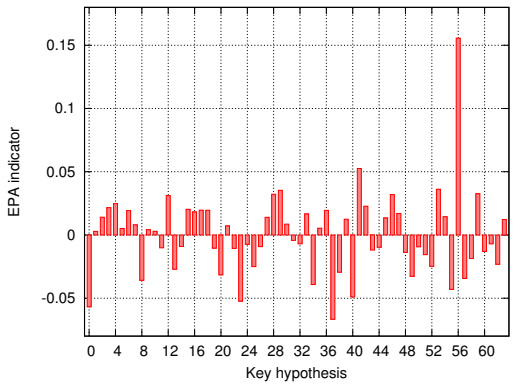


Fig. 7. EPA results on 200,000 power consumption traces of a ROM masked DES implementation. The correct key is $k = 56$.

### B. Experimental results

The *EPA* is carried out on a ROM masked DES implementation. It is tested on 200,000 traces of a masked DES implementation with different weights $(w_0, w_1, w_2, w_3, w_4)$ values. The weights of the $H$ function producing the best results are $(0.25, 1, 0, -1, -0.25)$.

Fig. 7 shows the EPA indicator values according to each key predicted for the first DES S-Box. Then for each S-Box, the round subkey guessed by the EPA attack is the key corresponding to the highest indicator value. The eight DES S-Boxes subkeys used during the first round of our masked DES have been guessed by the EPA attack, using the same weights $w_{i,i\in[0,4]}$.

### C. EPA Vs VPA Vs MIA

In this subsection, we compare the EPA attack (this article) to the MIA [8] and VPA [13] attacks. Following the recent advances concerning the comparison of univariate side-channel distinguishers [20], we apply the first-order success rate to assess the performance of the three attacks. The first-order success rate expresses the probability that, given $n$ measurements, the attack's best guess is the correct key. For each scenario, we acquired a set of 25,000 power consumption traces using random masks and plaintexts. To evaluate the scenario, we carry out the following algorithm:

for $n$ := 1000 to 25000 step 1000
1) *counter* := 0
2) for $i$ := 1 to 20
   a) select random $n$ power consumption traces from set $v_i$
   b) run the attack for the key $k \in [0,63]$
   c) increment *counter* if attack successful
3) compute success rate for $n$ traces as *counter*/20

Figure 8 shows our experimental results for those three attacks on the ROM masked DES implementation. We can see that VPA performs well in this scenario. About 5,000 traces suffice to achieve a success rate of 50% and starting from about 14,000 traces the VPA attack reveals the correct key with success rate of 95%. The EPA attack performs well also. The success rates stay well above 50% even when using 11,000 measurements, but eventually reaches success rate of

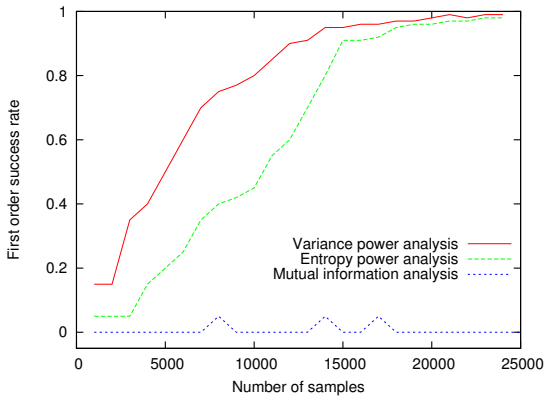


Fig. 8. First order success rate of 3 distinguishers on a ROM masked DES.

95% using 18,000 traces. MIA attack performs much worse. The success rates stay under 10% even when using 25,000 measurements. We conclude that the distinguisher based on the computation of the difference between entropy is more efficient than the MIA attack in the context of attacking a masked implementation. The VPA remains the best attack, since the leakage model is well known (see Eqn. (3)); however, in the context of an unknown model, an information-theoretic attack would be necessary; the EPA would be in this scenario the preferred distinguisher, since it would outperform the MIA.

## V. Conclusion and Perspectives

This paper shows the limitations of the mutual information attack, when masking is used to protect the implementation. We presented a 2O-DPA attack based on entropy analysis which succeeds in breaking a hardware masked DES implemented in an FPGA. This is the first high-order information-theoretic attack reported so far on a hardware accelerator. This attack is quite efficient on ROM implementation (all the S-Boxes are cracked) and requires a reasonable number of traces (15K). We compared it with an efficient partition distinguisher which is the variance. We observe that the variance performs better than our EPA attack when the leakage model is known.

A perspective is to compare EPA and VPA attacks with an MMIA [7], using multiple sensors (*e.g.* two magnetic probes) placed at different $(X, Y, Z, \vartheta)$ locations over a zero-offset masked cryptoprocessor.

## Acknowledgment

This work has been supported by the french Agency "Agence Nationale de la Recherche" in the framework of the ARPEGE SecReSoC project ANR-09-SEGI-013.

## References

[1] M.-L. Akkar and C. Giraud. An Implementation of DES and AES Secure against Some Attacks. In LNCS, editor, *Proceedings of CHES'01*, volume 2162 of *LNCS*, pages 309–318. Springer, May 2001. Paris, France.

[2] É. Brier, C. Clavier, and F. Olivier. Correlation Power Analysis with a Leakage Model. In *CHES*, volume 3156 of *LNCS*, pages 16–29. Springer, August 11–13 2004. Cambridge, MA, USA.

[3] S. Chari, C. Jutla, J. Rao, and P. Rohatgi. Towards Sound Approaches to Counteract Power-Analysis Attacks. In *CRYPTO*, volume 1666 of *LNCS*, August 1999. ISBN: 3-540-66347-9.

[4] S. Chari, J. R. Rao, and P. Rohatgi. Template Attacks. In *CHES*, volume 2523 of *LNCS*, pages 13–28. Springer, August 2002.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society B*, 39:1–38, 1977.

[6] A. M. Fraser and H. L. Swinney. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, 33(2):1134–1140, Feb 1986.

[7] B. Gierlichs, L. Batina, B. Preneel, and I. Verbauwhede. Revisiting Higher-Order DPA Attacks: Multivariate Mutual Information Analysis. In *CT-RSA*, volume 5985 of *LNCS*, pages 221–234. Springer, March 1-5 2010. San Francisco, CA, USA.

[8] B. Gierlichs, L. Batina, P. Tuyls, and B. Preneel. Mutual Information Analysis – A Generic Side-Channel Distinguisher. In E. Oswald and P. Rohatgi, editors, *Cryptographic Hardware and Embedded Systems - CHES 2008*, volume 5154 of *Lecture Notes in Computer Science*, pages 426–442, Washington DC, USA, 2008. Springer-Verlag.

[9] L. Goubin and J. Patarin. DES and differential power analysis (The "Duplication" Method). In *CHES*, volume 1717 of *LNCS*, pages 158–172. Springer, August 12-13 1999. Worcester, MA, USA.

[10] S. Guilley, P. Hoogvorst, and R. Pacalet. Differential Power Analysis Model and some Results. In Kluwer, editor, *Proceedings of WCC/CARDIS*, pages 127–142, Aug 2004. Toulouse, France.

[11] P. Kocher, J. Jaffe, and B. Jun. Differential Power Analysis. In *CRYPTO*, volume 1666 of *LNCS*, pages pp 388–397. Springer, 1999.

[12] A. Kolmogorov. Logical basis for information theory and probability theory. *Information Theory, IEEE Transactions on*, 14(5):662–664, Sep 1968.

[13] H. Maghrebi, J.-L. Danger, F. Flament, and S. Guilley. Evaluation of Countermeasures Implementation Based on Boolean Masking to Thwart First and Second Order Side-Channel Attacks. In *SCS*, IEEE, November 6–8 2009. Jerba, Tunisia. Complete version online: http://hal.archives-ouvertes.fr/hal-00425523/en/.

[14] S. Mangard and K. Schramm. Pinpointing the Side-Channel Leakage of Masked AES Hardware Implementations. In *CHES*, volume 4249 of *LNCS*, pages 76–90. Springer, 2006. PDF.

[15] J. Proakis. *Digital Communications*. McGraw-Hill Science/Engineering/Math, 4 edition, August 2000.

[16] E. Prouff and M. Rivain. Theoretical and practical aspects of mutual information based side channel analysis. In *ACNS '09: Proceedings of the 7th International Conference on Applied Cryptography and Network Security*, pages 499–518, Berlin, Heidelberg, 2009. Springer-Verlag.

[17] SASEBO board from the Japanese RCIS-AIST: http://www.rcis.aist.go.jp/special/SASEBO/index-en.html.

[18] C. E. Shannon. *A Mathematical Theory of Communication*. CSLI Publications, 1948.

[19] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, April 1986.

[20] F.-X. Standaert, B. Gierlichs, and I. Verbauwhede. Partition vs. comparison side-channel distinguishers: An empirical evaluation of statistical tests for univariate side-channel attacks against two unprotected cmos devices. In *ICISC*, pages 253–267, 2008.

[21] F.-X. Standaert, G. Rouvroy, and J.-J. Quisquater. FPGA Implementations of the DES and Triple-DES Masked Against Power Analysis Attacks. In *proceedings of FPL 2006*, August 2006. Madrid, Spain.

[22] N. Veyrat-Charvillon and F.-X. Standaert. Mutual information analysis: How, when and why? In *CHES 2009*, volume 5747/2009, pages 429–443. Springer, 2009. http://dx.doi.org/10.1007/978-3-642-04138-9-30.

[23] M. Wand. Data-based choice of histogram bin width. Statistics working paper, Australian Graduate School of Management, 13th May 1996.