Analysis Of Variance and CPA in SCA

Sébastien Tiran[1], Guillaume Reymond[3], Jean-Baptiste Rigaud[5], Driss Aboulkassimi[3], Benedikt Gierlichs[6], Mathieu Carbone[1,2], Gilles Ducharme[4], Philippe Maurine[3,1]

[1] LIRMM, Université Montpellier II
161 rue Ada, 34392 Montpellier CEDEX 5, France
[2] STMicroelectronics, Advanced System Technology (AST)
190 Avenue Célestin Coq, Z.I. Peynier-Rousset, 13106 Rousset CEDEX, France
[3] CEA - Commissariat à l'Énergie Atomique et aux Énergies Alternatives
880 route de Mimet, 13120 Gardanne, France
[4] EPS - Institut de Mathématiques et de Modélisation de Montpellier
2, Place Eugène Bataillon, Université Montpellier 2, 34095 Montpellier Cedex 5, France
[5] Ecole Nationale Supérieure des Mines de Saint Etienne
CMPGC, 880 Route de Mimet, 13541 Gardanne, France
[6] Katholieke Universiteit Leuven, COSIC and IBBT
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium

Abstract. This paper introduces Side-Channel Analysis results obtained on an unprotected circuit characterized by a surprisingly non-linear leakage. While in such a case, Correlation Power Analysis is not adapted, we show that a more generic attack, based on the Analysis Of Variance (AOV) outperfoms CPA. It has the advantage of detecting non-linear leakage, unlike Correlation Power Analysis, and of providing similar or much better results in all cases, with a similar computation time.

Keywords: SCA, AOV, CPA, leakage

1 Introduction

Since the works of Kocher *et al.* [9], new Side Channel Attacks (SCA) or techniques to increase their efficiency have been proposed in the literature. E. Brier, C. Clavier and F. Olivier proposed the use of Pearson correlation instead of the Difference of Means (DoM) to exploit the dependency between power consumption and processed data [5]. This leads to the so called CPA which looks for a linear relation between these two variables. Mutual Information Analysis (MIA) was introduced by S. Aumonier [3] and B. Gierlichs [8] who have proposed the use of the Mutual Information index, a more generic distinguisher able to detect any kind of relation between these variables. This proposal was then further enhanced in [15, 10, 14]. However, even if the MI index is theoretically the most generic distinguisher, its use raises several difficulties. Indeed, the choice of hyper-parameters to obtain an efficient analysis is crucial [6].

Within this context, little attention has been paid to the Analysis Of Variance (AOV). Its first use in the context of SCA was introduced by F. Standaert and B. Gierlichs [12], and further analyzed in [4]. In [12], AOV is applied to different sets of traces in order to experimentally compare its efficiency to that of different distinguishers, namely: the Difference of Means (DoM), the Pearson correlation (ρ) and the Mutual Information (MI) index. However, no general conclusion could be drawn from these results except that an Hamming Weight (HW) partitioning seemed to be the best choice. Indeed, this paper does not provide any decisive information about the superiority of AOV over CPA and MIA. In [4], it is however argued that AOV and CPA give similar results in practice. The resulting question is then: what kind of practice ? This question is especially important as AOV can also detect non-linear relations between two variables and can easily be extended to multivariate analyses.

The main goal of this paper is to highlight that in the case of a linear leakage with the Hamming Weight (HW) or Distance (HD), the AOV gives similar results as the Pearson correlation. To proceed toward this goal, we first present a theoretical analysis of the difference between a distinguisher based on AOV and CPA. These theoretical results indicate that AOV provides at least the same efficiency as CPA in most cases and better results in uncommon cases characterized by a non-linear leakage model. The AOV appears thus superior to CPA because it offers a kind of theoretical warranty of capturing more complex leakages than CPA, while providing similar results in case (the most common one) of a linear leakage. This aspect is of prime importance while evaluating a design. Indeed, CPA could fail to recover the key in some cases and AOV could succeed, while the reverse is theoretically impossible.

To assess the impact of these theoretical results in realistic situations, we secondly present experimental results confirming that, while the AOV performs similar results to CPA in the case of a linear leakage, it can provide much better results with an unusual leakage, and, results similar to MIA with the same computational burden as CPA.

The rest of the paper is organized as follows. In section 2, a reminder of SCA principles is presented. Section 3 reminds the basics of AOV which is then compared to CPA. Experimental results are presented in section 4, starting by the description of the devices under study. The leakage of these unprotected devices is profiled using Akaike Information Criterion (AIC) and the efficiency of CPA and AOV are then compared. Finally a conclusion is drawn from the obtained results.

2 SCA principle

A SCA aims at recovering secret information by exploiting a physical leakage (e.g. power consumption) which depends on the data and the secret (e.g. a cryptographic key) processed by the circuit under analysis. Among all possible SCA, vertical SCA requires multiple executions of the algorithm implemented on the targeted device to recover the key.

A univariate and non-profiled differential SCA is usually performed following three steps. First, the adversary acquires a set of power traces, corresponding to the encryption or decryption of messages. Second he makes guesses on the key to predict a targeted intermediate binary word processed by the algorithm, and sorts traces according to its values. The sorting is done according to, one, several or all bits of this word, or also to its Hamming Weight value, and to a given model of the power consumption. In practice, two models have been proved efficient: the Hamming Weight Model (HWM) and the Hamming Distance Model (HDM). According to the first one, the consumption is greater when a target bit is equal to 1, while according to the second, the consumption is higher when the bit toggles during the calculation. Third, the adversary applies a distinguisher between the traces and the predicted values to get a score for each key guess. Finally, he identifies the secret key as the key guess corresponding to the maximum score. Indeed this is the one corresponding to the best prediction of the power consumption.

In this context, the choice of the distinguisher is important as it is used to detect the dependency between the two variables: the traces and the predicted values. A very popular distinguisher in SCA is the Pearson correlation. However, it has some limitations: it only detects an eventual linear dependency between two variables and does not detect other kinds of relations. The use of another distinguisher has been proposed in [12, 4]; it is based on AOV. It presents many advantages over the Pearson correlation. In the next section we propose to revisit the theoretical advantages of this distinguisher leading to a more favorable conclusion for AOV.

3 Analysis Of Variance : AOV

The one-way AOV allows to study the behavior of a random variable of interest, noted $\mathcal{L}$, according to the values of one discrete explanatory variable or factor, noted $\mathcal{H}$, taking H distinct values that we generically denote by $h \in \{1, \ldots, H\}$. AOV achieves this by partitioning the variance of $\mathcal{L}$ into components that are expected to be responsible of different sources of variation.

3.1 Total sum of squares decomposition for AOV

A sample of n_h values of $\mathcal{L}$, noted $\{L_{h,1}, \ldots L_{h,n_h}\}$, is observed at each value $h \in \{1, \ldots, H\}$ of factor $\mathcal{H}$. These observations are assumed independent and identically distributed with expected value $\mu_h = \mathbb{E}(L_{h,1})$. We regroup these observations into $\boldsymbol{L} = (L_{1,1}, \ldots, L_{1,n_1}, L_{2,1}, \ldots, L_{H,n_H})^T$ (here "^{T}" denotes transposition), the vector of length $n = n_1 + \cdots + n_H$ of all observations, where $L_{h,i} = \mu + \epsilon_{h,i}$. The noise $\epsilon_{h,i} \approx \mathbb{N}(0, \sigma^2)$ is usually assumed Gaussian with mean 0 and constant variance σ^2.

AOV seeks to determine if the factor $\mathcal{H}$ has an effect on $\mathcal{L}$ by assessing the assumption $\mu_1 = \mu_2 = \cdots = \mu_H$. The total sum of squares is defined as :

$$SS_{tot} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} (L_{h,i} - \bar{L})^2 \tag{1}$$

where $\bar{L} = n^{-1} \sum_{h=1}^{H} \sum_{i=1}^{n_h} L_{h,i}$ is the *global* mean of all components of $\boldsymbol{L}$. Using the centering matrix $K_n = n^{-1} 1_n 1_n^T$, where $1_n \in \mathbb{R}^n$ is the vector of "ones", this can be written as the quadratic form $SS_{tot} = \boldsymbol{L}^T (I_n - K_n) \boldsymbol{L}$, where I_n denotes the identity matrix of order n. For any $n \times n$ matrix A, we obviously have the "decomposition" : $\boldsymbol{L}^T (I_n - K_n) \boldsymbol{L} = \boldsymbol{L}^T (I_n - A) \boldsymbol{L} + \boldsymbol{L}^T (A - K_n) \boldsymbol{L}$. The AOV uses for A the matrix composed of the H^2 blocks of size $n_h \times n_{h'}$

$$A_{h,h'} = \begin{cases} \frac{1}{n_h} 1_h 1_h^T & \text{if} \quad h = h' \\ 0 & \text{otherwise} \end{cases}$$

where now $1_h \in \mathbb{R}^{n_h}$. We will write A_{aov} for this matrix. The ensuing decomposition can then be written as

$$\begin{aligned} SS_{tot} &= \boldsymbol{L}^T (I_n - A_{aov}) \boldsymbol{L} + \boldsymbol{L}^T (A_{aov} - K_n) \boldsymbol{L} \\ &= \sum_{h=1}^{H} \sum_{i=1}^{n_h} (L_{h,i} - \bar{L}_{h\cdot})^2 + \sum_{h=1}^{H} n_h (\bar{L}_{h\cdot} - \bar{L})^2 \\ &= SS_{err} + SS_{treat} \end{aligned} \tag{2}$$

with $\bar{L}_{h\cdot}$, the mean of $\{L_{h,1}, \ldots L_{h,n_h}\}$, being an estimator of μ_h. The term SS_{err} is the "error sum of squares" and reflects the variation of the data about their mean $\bar{L}_{h\cdot}$ within each values h of $\mathcal{H}$; when the values $\{L_{h,1}, \ldots L_{h,n_h}\}$ are close to their respective $\bar{L}_{h\cdot}$, SS_{err} will be close to zero. The second term, the "treatment sum of squares", captures the weighted variations of $\bar{L}_{h\cdot}$ as $\mathcal{H}$ varies : if the $\bar{L}_{h\cdot}$ are close to the global mean $\bar{L}$, SS_{treat} will be close to zero and support the assumption $\mu_1 = \mu_2 = \cdots = \mu_H$.

A unitless measure of the support offered by the data toward this assumption (which has been used as a distinguisher in [12] in a SCA context) is:

$$R^2_{aov} = 1 - \frac{SS_{err}}{SS_{tot}} = \frac{SS_{treat}}{SS_{tot}}. \tag{3}$$

R^2_{aov}, which lies in $[0, 1]$, will tend to be in the neighborhood of zero when the assumption $\mu_1 = \mu_2 = \cdots = \mu_H$ is true and increases toward 1 as the data departs more strongly. In classical AOV, the usual F-test for the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_H$ is based on the test statistic $((n-H) R^2_{aov} / ((H-1)(1 - R^2_{aov}))$ which, under H_0 (and Gaussian noise) follows a Fisher distribution with degrees of freedom $(H-1, n-H)$. It is important to stress that any departures from H_0 can be detected via R^2_{aov}. Thus AOV tries to answer the question *"Is the effect of $\mathcal{H}$ on $\mathcal{L}$ significantly different across its values* ?", e.g. are there at least two values $\mu_h, \mu_{h'}$ such that $\mu_h \neq \mu_{h'}$?

3.2 Total sum of squares decomposition for Regression

When trying to relate a factor to a variable of interest, another possible question is "*Do the values of* $\mathcal{H}$ *affect* $\mathcal{L}$?" An answer to this question can be obtained via regression analysis where it is implicitly assumed that, if such an effect exists, increasing the value of $\mathcal{H}$ modifies linearly (approximately) the values of $\mathcal{L}$. To answer this question, we assume $L_{h,i} = a + b \times h + \epsilon_{h,i}$, where the "noise" $\epsilon_{h,i}$ is again usually assumed Gaussian with mean 0 and constant variance σ^2, noted $\epsilon_{h,i} \sim N(0, \sigma^2)$. The question can then be recasted as "is $b = 0$?".

To answer this, regression analysis uses the matrix

$$A_{reg} = X(X^T X)^{-1} X^T$$

where X^T is a $2 \times n$ matrix with blocks $X_h^T = \begin{pmatrix} 1 \cdots 1 \\ h \cdots h \end{pmatrix}$ of dimension $2 \times n_h$. The matrix A_{reg} is composed of the $n_h \times n_{h'}$ blocks $c_{h,h'} 1_h 1_{h'}^T$ with $c_{h,h'} = \frac{1}{n} + \frac{(h-\bar{h})(h'-\bar{h})}{nS_{\mathcal{H}}^2}$ where $S_{\mathcal{H}}^2 = n^{-1} \sum_{h=1}^{H} n_h h^2 - \left(n^{-1} \sum_{h=1}^{H} n_h h \right)^2 = \bar{h^2} - \bar{h}^2$ and $\bar{h} = n^{-1} \sum_{h=1}^{H} n_h h$. The decomposition $\boldsymbol{L}^T (I_n - K_n) \boldsymbol{L} = \boldsymbol{L}^T (I_n - A_{reg}) \boldsymbol{L} + \boldsymbol{L}^T (A_{reg} - K_n) \boldsymbol{L}$ can also be written as $SS_{tot} = SS_{err} + SS_{reg}$ and a unitless measure of the validity of the assumption $b = 0$ is obtained through

$$R_{reg}^2 = 1 - \frac{SS_{err}}{SS_{tot}} = \frac{SS_{reg}}{SS_{tot}}.$$

Again R_{reg}^2, which lies in $[0, 1]$, will tend to be in the neighborhood of zero when $b = 0$ and increases toward 1 as the data cluster more closely about the line $a + b \times h$. In classical linear regression analysis, the usual F-test for the null hypothesis $H_0 : b = 0$ is based on $(n-2) R_{reg}^2 / (1 - R_{reg}^2))$ which, under H_0 (and Gaussian noise) follows a Fisher F-distribution with degrees of freedom $(1, n-2)$. In the present context where $\mathcal{H}$ takes only H distinct values, the null hypothesis $H_0 : b = 0$ of regression is equivalent to the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_H$ of AOV. It is important to stress however that in contrast to AOV, all departures from $L_{h,i} = a + b \times h + \epsilon_{h,i}$ cannot be detected with R_{reg}^2. Indeed if the true model is a perfect quadratic polynomial centered on $\bar{h} = n^{-1} \sum_{h=1}^{K} n_h h$, R_{reg}^2 will be close to zero, whereas R_{aov}^2 should be much greater. Also important to stress is the fact that R_{reg}^2 is the square of Pearson's correlation coefficient $\rho_{Pearson}$, which is the distinguisher used in CPA.

3.3 Linking Pearson's Correlation Coefficient with AOV

From the above decompositions, we have

$$R_{aov}^2 = R_{reg}^2 + \frac{\boldsymbol{L}^T (A_{aov} - A_{reg}) \boldsymbol{L}}{\boldsymbol{L}^T (I_n - K_n) \boldsymbol{L}},$$

so the above stated differences in the behavior of R_{aov}^2 with respect to R_{reg}^2 are caused by the difference term

$$\begin{aligned} R^2_{dif} &= \frac{\boldsymbol{L}^T(A_{aov} - A_{reg})\boldsymbol{L}}{\boldsymbol{L}^T(I_n - K_n)\boldsymbol{L}} \\ &= \frac{\boldsymbol{L}^T(A_{aov} - A_{reg})\boldsymbol{L}/n}{\boldsymbol{L}^T(I_n - K_n)\boldsymbol{L}/n} \\ &= \frac{Z_n}{S^2_{\mathcal{L}}} \end{aligned}$$

where $S^2_{\mathcal{L}}$ is the empirical variance of the data in $\boldsymbol{L}$ and converges toward the variance $\sigma^2_{\mathcal{L}}$ of the marginal distribution of $\mathcal{L}$. Thus we need only to study the term $Z_n = \boldsymbol{L}^T(A_{aov} - A_{reg})\boldsymbol{L}/n$ which is a quadratic form in the matrix $A_{aov} - A_{reg}$ composed of the H^2 blocks of size $n_h \times n_{h'}$

$$B_{h,h'} = \begin{cases} \left(\frac{1}{n_h} - c_{h,h'}\right) 1_h 1_h^T & \text{if} \quad h = h' \\ -c_{h,h'} 1_h 1_{h'}^T & \text{otherwise} \end{cases}.$$

It is easy to see that $A_{aov} - A_{reg}$ is idempotent and symmetric. Hence it is semi-definite positive and, by standard results on extrema of quadratic forms, $0 \leq \boldsymbol{L}^T(A_{aov} - A_{reg})\boldsymbol{L} \leq \boldsymbol{L}^T\boldsymbol{L}$. Hence, we always have

$$R^2_{aov} \geq R^2_{reg}.$$

To better understand the behavior of $Z_n = \boldsymbol{L}^T(A_{aov} - A_{reg})\boldsymbol{L}/n$, and its effect on the genericity of R^2_{aov} in the context of SCA, we now look into its expectation and variance. Write $\mathbb{E}(\boldsymbol{L}) = \boldsymbol{\mu}_{\mathcal{L}}$ with the first n_1 components being μ_1, the n_2 following being μ_2 etc. We have, conditional on the values of $\mathcal{H}$,

$$\begin{aligned} \mathbb{E}(\boldsymbol{L}^T(A_{aov} - A_{reg})\boldsymbol{L}) &= tr(A_{aov} - A_{reg})\mathbb{E}(\boldsymbol{L}\boldsymbol{L}^T) \\ &= tr(A_{aov} - A_{reg})\left(\mathbb{V}(\boldsymbol{L}) + \boldsymbol{\mu}_{\mathcal{L}}\boldsymbol{\mu}_{\mathcal{L}}^T\right) \\ &= \sigma^2_{\mathcal{L}} tr(A_{aov} - A_{reg}) + \boldsymbol{\mu}_{\mathcal{L}}^T(A_{aov} - A_{reg})\boldsymbol{\mu}_{\mathcal{L}}, \end{aligned}$$

because by assumption the variance-covariance matrix $\mathbb{V}(\boldsymbol{L})$ of $\boldsymbol{L}$ is $\sigma^2_{\mathcal{L}} I_n$. Now

$$tr(A_{aov}) = \sum_{h=1}^{H}\sum_{i=1}^{n_h}\frac{1}{n_h} = H,$$

$$tr(A_{reg}) = \sum_{i=1}^{n}\frac{1}{n} + \sum_{h=1}^{H}\frac{n_h(h-\bar{h})^2}{n(\bar{h^2} - \bar{h}^2)} = 2,$$

so that

$$\mathbb{E}(Z_n) = \frac{\sigma^2_{\mathcal{L}}(H-2)}{n} + \frac{\boldsymbol{\mu}_{\mathcal{L}}^T A_{aov}\boldsymbol{\mu}_{\mathcal{L}}}{n} - \frac{\boldsymbol{\mu}_{\mathcal{L}}^T A_{reg}\boldsymbol{\mu}_{\mathcal{L}}}{n}.$$

Now

$$\boldsymbol{\mu}_{\mathcal{L}}^T A_{aov} \boldsymbol{\mu}_{\mathcal{L}} = \sum_{h=1}^{H} \sum_{h'=1}^{H} \mu_h 1_h^T A_{h,h'} \mu_h 1_{h'} = \sum_{h=1}^{H} \mu_h^2 1_h^T A_{h,h'} 1_{h'} = \sum_{h=1}^{H} n_h \mu_h^2$$

Similarly

$$\begin{aligned}
\boldsymbol{\mu}_{\mathcal{L}}^T A_{reg} \boldsymbol{\mu}_{\mathcal{L}} &= \sum_{h=1}^{H} \sum_{h'=1}^{H} \mu_h 1_h^T c_{h,h'} 1_h 1_{h'}^T 1_{h'} \mu_h \\
&= \frac{1}{n} \sum_{h=1}^{H} \sum_{h'=1}^{H} \mu_h \mu_{h'} n_h n_{h'} + \sum_{h=1}^{H} \sum_{h'=1}^{H} \mu_h \mu_{h'} \frac{(h-\bar{h})(h'-\bar{h})}{n S_{\mathcal{H}}^2} n_h n_{h'} \\
&= \frac{1}{n} \left(\sum_{h=1}^{H} \mu_h n_h \right)^2 + \frac{1}{n S_{\mathcal{H}}^2} \left(\sum_{h=1}^{H} \mu_h (h-\bar{h}) n_h \right)^2
\end{aligned}$$

Thus, upon noticing that $n^{-1} \sum_{h=1}^{H} \mu_h (h-\bar{h}) n_h$ is the weighted (by n_h) empirical covariance $S_{\mathcal{H},\mu}$ of the points $(h, \mu_h)_{h=1,\ldots,H}$, we get

$$\begin{aligned}
\mathbb{E}(Z_n) &= \frac{\sigma_{\mathcal{L}}^2 (H-2)}{n} + \left(\sum_{h=1}^{H} \frac{n_h}{n} \mu_h^2 - \left(\frac{1}{n} \sum_{h=1}^{H} \mu_h n_h \right)^2 \right) - \frac{1}{S_{\mathcal{H}}^2} \left(S_{\mathcal{H},\mu} \right)^2 \\
&= \frac{\sigma_{\mathcal{L}}^2 (H-2)}{n} + S_\mu^2 - S_\mu^2 \frac{\left(S_{\mathcal{H},\mu} \right)^2}{S_{\mathcal{H}}^2 S_\mu^2} \\
&= \frac{\sigma_{\mathcal{L}}^2 (H-2)}{n} + S_\mu^2 \left(1 - \rho_{\mathcal{H},\mu}^2 \right)
\end{aligned}$$

where S_μ^2 is the empirical variance of the terms in $\boldsymbol{\mu}_{\mathcal{L}}$ and $\rho_{\mathcal{H},\mu}^2$ is the empirical weighted (by n_k) Pearson correlation coefficient for the points $(h, \mu_h)_{h=1,\ldots,H}$. Hence the expectation of the term R_{dif}^2 is approximately

$$\mathbb{E}(R_{dif}^2) \simeq \frac{\sigma_{\mathcal{L}}^2}{\mathbb{E}(S_{\mathcal{L}}^2)} \frac{(H-2)}{n} + \frac{S_\mu^2}{\mathbb{E}(S_{\mathcal{L}}^2)} \left(1 - \rho_{\mathcal{H},\mu}^2 \right). \tag{4}$$

As for the variance of this term, we need only its order so that we consider the particular case where the errors $\epsilon_{h,i}$ are Gaussian, for which the calculations are easy. In this case, standard results on the variance of a quadratic form show that (recall that $A_{aov} - A_{reg}$ is idempotent)

$$\mathbb{V}(Z_n) = \frac{2}{n^2} \sigma_{\mathcal{L}}^4 tr\left(A_{aov} - A_{reg} \right) - \frac{4}{n^2} \left(\boldsymbol{\mu}_{\mathcal{L}}' (A_{aov} - A_{reg}) \boldsymbol{\mu}_{\mathcal{L}} \right)^2 = O(n^{-1}).$$

Hence, by Tchebychev's inequality, we get that the dominant terms in R_{dif}^2 (as n increases) is

$$R_{dif}^2 = \frac{S_\mu^2}{\sigma_{\mathcal{L}}^2} \left(1 - \rho_{\mathcal{H},\mu}^2 \right) + \frac{(H-2)}{n} + o_p(n^{-1}).$$

This expression allows the analysis of the term R^2_{dif}. First, when $H = 2$, and because the correlation between a pair of points is always ± 1, we have that $R^2_{dif} \simeq 0 + o_p(n^{-1})$, so that an SCA based on AOV will give results similar to the corresponding CPA. Actually here, in view of the fact that AOV with $H = 2$ is the same as a squared student t test, the equality holds exactly and the student version of Kocher's DPA is equivalent to a CPA, a fact already noticed by [11].

When $H > 2$, then R^2_{aov} is equivalent up to the constant term $\frac{(H-2)}{n} + o_p(n^{-1})$ to R^2_{reg} if the points $\{(h, \mu_h), h = 1, \ldots, H\}$ fall on a strait line where then $\rho_{\mathcal{H},\mu} = 1$.

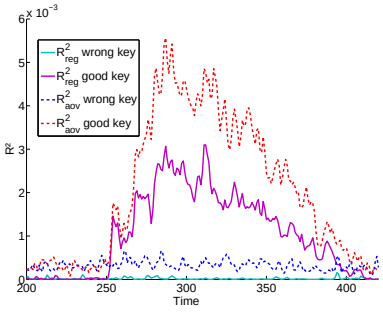

Fig. 1. DPAv2-AES (20000 traces) - Evolution of R^2_{reg} and R^2_{aov}, for S-box 2, according to time

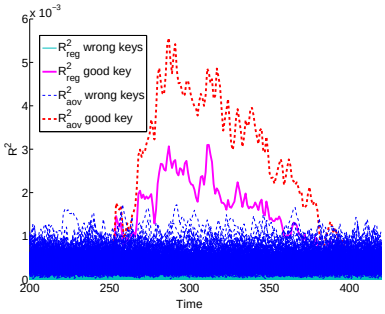

Fig. 2. DPAv2-AES (20000 traces) - Evolution of R^2_{reg} and R^2_{aov}, for S-box 2, according to time

As an experimental demonstration, Fig. 1 and 2 show the evolution of R^2_{reg} and R^2_{aov} obtained on the DPAv2 traces, with a leakage assumed close to linearity. This point will be discussed later in the paper. Fig. 1 only shows R^2_{reg} and R^2_{aov} for the correct key guess and one wrong guess, while 2 show them for all key values. As expected R^2_{aov} is always above R^2_{reg}.

Considering this fact, it appears that an SCA using the AOV should be preferred over Pearson correlation because it has the advantage of being more generic *and* provides theoretically the same results (up to the order $o_p(n^{-1})$ in the case of a linear leakage. Indeed, when the leakage is unknown, which is often the case, it is rather a risky choice to perform a CPA. The use of a more generic distinguisher like the distinguisher based on AOV, appears thus a reasonable choice, because it enables to cover cases where the leakage is not linear.

However, it should be noted that, theoretically, AOV is not as general as MIA because it works on means and is therefore not able to detect links hidden in higher statistical moments. Nonetheless, this loss of genericity with respect to MIA is compensated by a drastic reduction in the computational burden, burden which is comparable to that of CPA. At this point, the question is to decide if leakages can be brought by higher moments than the mean in practice. Whatever the answer, to confirm our theoretical results about CPA and AOV,

experiments were conducted on two different testcases characterized by radically different leakage models.

4 Experimental Results

To compare CPA and AOV, attacks on the last two rounds of the AES-128 were coded in C language. To get deeper insights, we also compared the results obtained with these two analyses with those provided by a MIA [8] based on kernel density estimation with adaptive bandwidth selection [6]. These attacks were applied to two different testcases.

4.1 First testcase

As the first testcase, denoted DPAv2-AES in the rest of the paper, we select the traces from the DPA contest v2, which are publicly available traces. They correspond to power traces of a AES-128, implemented on a SASEBO GII board [2], its design being the one from AIST and Tohoku University. More information about these traces can be found in [1].

4.2 Second testcase

The second testcase, denoted by 65nm-AES afterward, is an AES-128 designed with a 65nm Low Power High Threshold Voltage CMOS technology. This circuit presents some specific characteristics with respect to smartcards. First, it integrates an in-house communication protocol and second it is supplied by 16 pads so that the power consumed by the AES is not drawn from a single power pad. A picture of the IC showing the location of the AES on the die is given Fig. 3.

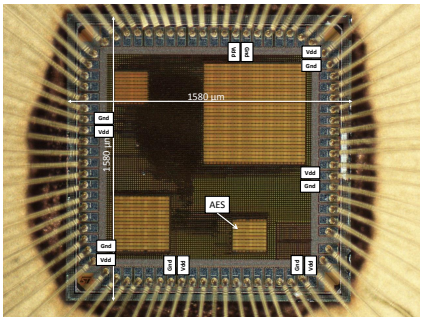

Fig. 3. Picture of the first testchip showing the position of the 16 Vdd and Gnd pads

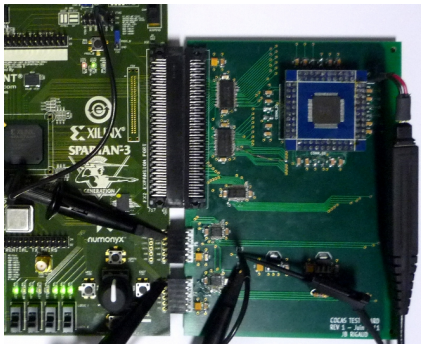

Fig. 4. Measurement Setup

The experiment set-up used to collect power traces is given Fig. 4. A Xilinx Spartan 3 FPGA board is used to drive the IC via a state machine while a serial port transfers from a PC to the FPGA the plain texts, the secret key and the configuration commands such as the encryption start and reset signals. The state machine manages the bidirectional communication with the circuit and sets the controls signals with the right timing according to the circuit specifications.

Power traces are acquired with a differential probe measuring the variations of Vdd and Gnd, and a oscilloscope with a 20GS/s sampling rate. The bandwidth of the whole acquisition setup was 1Mhz-4GHz.

4.3 Leakage Profiling

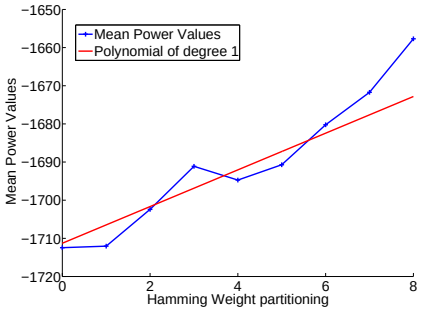

Fig. 5. DPAv2-AES (20000 traces) - Polynomial fitting the best the leakage according to the AIC

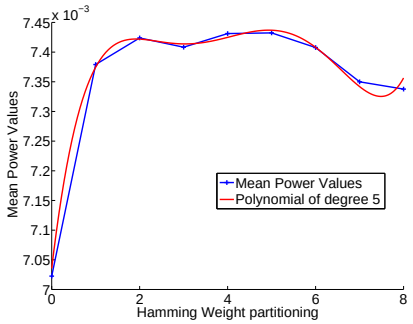

Fig. 6. 65nm-AES - Polynomial fitting the best the leakage according to the AIC

Before performing any attack on these testcases, one key point was to profile the leakage of both testcases. The goal was to verify our engineering intuition according to which the first testcase (65nm-AES) is leaking according to a non-linear leakage model. Even if it was not crucial, the same profiling step was applied to the second testcase, the DPAv2-AES.

This profiling step was conducted as follows. Traces were sorted for each S-box according to a Hamming Distance model (HD), with a Hamming Weight (HW) partitioning *i.e.* were sorted according to the value of $HW(T_9(i) \oplus T_{10}(i))$, where $T_9(i)$ is the i^{th} byte of the message before the last Subbyte and Addroundkey operations, $T_{10}(i)$ is the i^{th} byte of the ciphertext at the end of round 10, and $HW()$ is the Hamming Weight of the byte.

The solid lines of Fig. 5 and 6 show, for the DPAv2-AES and the 65nm-AES respectively, the mean values of a leaking sample extracted from traces according to the aforementioned partitioning for S-box 2. It is therefore possible to observe the leakage evolutions with $HW(T_9(i) \oplus T_{10}(i))$. For the DPAv2-AES, this evolution seems linear while it is far from being linear for the 65nm-AES.

At that stage, we could have applied a weighted least squares method directly on the mean of each partition to find the polynomials fitting the data, namely the sample amplitude with respect to $HW(T_9(i) \oplus T_{10}(i))$. However, because the cardinality of each partition is not balanced, we used the Akaike Criterion (AIC), with the weighted least squares method to find these polynomials. The AIC is a trade-off between goodness of fit and the complexity of the model. Its expression is:

$$AIC = 2(N+1) + n \times ln\left(\frac{SS_{reg}}{n}\right) \tag{5}$$

with N the degree of the polynomial, n the number of points, and SS_{reg} the residual sum of squares. With such a tool, we found the polynomials with the minimum degree, that best fit the two leakages, by searching the degrees of the polynomials that minimize the AIC criterion.

The dotted lines of Fig. 5 show that for S-box 2 the best polynomial is of degree 1 while it is of degree 5 in Fig. 6. We are therefore facing a linear leakage when analysing the S-box 2 of the DPAv2-AES and a non linear model in case of the 65nm-AES.

This procedure was applied to all remainder S-boxes. The degrees of the best fitting-polynomial are reported for all S-boxes in Table 1. For the 65nm-AES testcase, the degree of the polynomials ranges between 4 and 8 while for the DPAv2-AES it ranges between 1 and 3. It may be noted that, in the 65nm-AES case, the degree of this polynomial for S-box 5 is 0, as the subkey cannot be retrieved, suggesting that there is no leakage.

Table 1. Degree of polynomials found for each sbox according to the AIC

Sbox	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
DPAv2-AES	2	1	1	2	1	1	2	1	3	1	1	2	3	1	1	1
65nm-AES	4	5	6	5	0	5	6	6	4	8	5	5	8	5	5	5

In order to confirm these results, and to make the link with section 3, the R^2_{aov} and R^2_{reg} coefficients were computed for all S-boxes in both testcases. Results are shown in Table 2. As proved in section 3, R^2_{aov} is always greater than R^2_{reg}. Moreover it may be noticed that in the DPAv2-AES case, R^2_{aov} and R^2_{reg} are, as expected from the theoretical results of section 3.3, very close to each other, confirming that the leakage is linear. On the contrary, much lower values are obtained for R^2_{reg} on the 65nm-AES case, and there is a great difference between R^2_{aov} and R^2_{reg}. This also confirms that the leakage is not linear in this case.

According to the results listed in the Tables 1 and 2, and to the conclusion we drew in section 3, we were expecting, before launching further analyses, that:

- CPA and AOV give similar results when applied to the DPAv2-AES testcase (except maybe for the S-box 9 and 13)
- AOV outperforms CPA when applied to the 65nm-AES testcase.

Table 2. Comparison between R^2_{aov} and R^2_{reg} for both testcases.

	DPAv2-AES			65nm-AES		
Sbox	R^2_{reg}	R^2_{aov}	R^2_{reg}/R^2_{aov}	R^2_{reg}	R^2_{aov}	R^2_{reg}/R^2_{aov}
1	0.00248	0.00335	0.74061	0.00001	0.00143	0.00840
2	0.00254	0.00426	0.59540	0.00005	0.00241	0.01952
3	0.00188	0.00306	0.61584	0.00019	0.00190	0.10026
4	0.00291	0.00475	0.61391	0.00014	0.00268	0.05108
5	0.00117	0.00353	0.33012	0.00002	0.00023	0.09829
6	0.00377	0.00508	0.74183	0.00006	0.00287	0.02022
7	0.00283	0.00544	0.51940	0.00011	0.00256	0.04225
8	0.00230	0.00458	0.50306	0.00031	0.00282	0.11068
9	0.00524	0.00710	0.73869	0.00006	0.00398	0.01408
10	0.00504	0.00738	0.68247	0.00008	0.00333	0.02432
11	0.00148	0.00286	0.51766	0.00017	0.00312	0.05479
12	0.00303	0.00575	0.52750	0.00011	0.00278	0.03846
13	0.00380	0.00610	0.62219	0.00012	0.00243	0.04768
14	0.00139	0.00327	0.42603	0.00015	0.00216	0.06812
15	0.00303	0.00438	0.69196	0.00043	0.00243	0.17763
16	0.00134	0.00311	0.43115	0.00003	0.00261	0.01304

4.4 Results obtained with CPA and AOV and MIA

CPA, MIA and AOV were applied to the two cases. Results from these attacks are presented in this section. They were carried out on the last round of the AES with a HD model, this is to say by partitioning according to the value: $HW(T_9(i) \oplus T_{10}(i))$. The MIA used here is based on kernels [15], and the choice of the bandwidth is made according to the method proposed in [6].

In order to compare the results of these attacks, a metric presented in [13] was used, namely the Guessing Entropy (GE). It represents the mean position of the correct key among the guesses, after a given number of traces. It should be noticed that the Mean Guessing Entropy (MGE) is also used; it represents the mean position of all the 16 correct sub-keys. To compute these metrics on a given set of traces, a same attack is applied several times with traces processed in different and random orders.

Fig. 7 shows the evolution of the Mean Guessing Entropy obtained with these attacks applied to the DPAv2-AES. All attacks provide similar results and no distinguisher enables to recover the key with significantly less traces than the others. This result was expected because the leakage is linear for most S-boxes. Note however, that in that case, CPA and AOV are more interesting than MIA because they are easier to apply (no hyper-parameter needs to be fixed) and faster to compute. Indeed, with our PC, the time spent by CPA, AOV and MIA to process 1000 traces was respectively equal to 3 s, 5 s and 6 min.

Fig. 8 also shows the evolution of the Mean Guessing Entropy but for the 65nm-AES. CPA clearly provides the worst results while MIA and AOV give the same results. Indeed, after the processing of 50000 traces with CPA, only 3 good subkeys are ranked first and 8 are ranked among the ten best hypotheses while

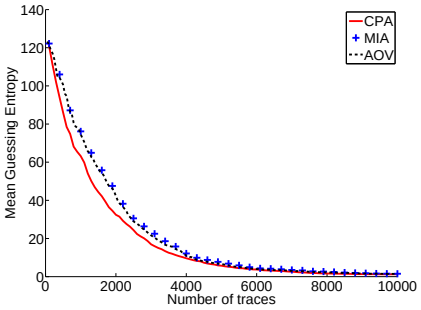

Fig. 7. Mean Guessing Entropy obtained with CPA, AOV and MIA after the processing of 20000 traces of the DPAv2-AES

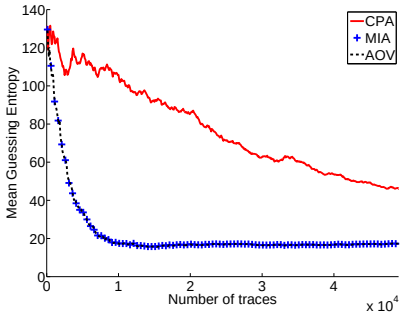

Fig. 8. Mean Guessing Entropy obtained with CPA, AOV and MIA after the processing of 60000 traces collected above the 65nm-AES

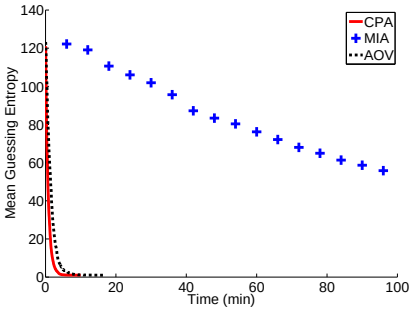

Fig. 9. Mean Guessing Entropy obtained with CPA, AOV and MIA after the processing of 20000 traces of the DPAv2-AES, depending on the computation time

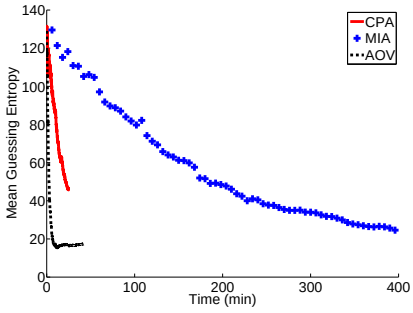

Fig. 10. Mean Guessing Entropy obtained with CPA, AOV and MIA after the processing of 60000 traces collected above the 65nm-AES, depending on the computation time

AOV and MIA are able to disclose 14 subkeys among 16 after the processing of only 10000 traces. It is to notice that the 2 last subkeys cannot be retrieved with a HD or a HW model, neither with AOV nor with MIA. However, because AOV requires a much reduced computational effort compared to MIA, it follows that AOV leads to the best trade off between efficiency, genericity and computational burden. Note also that because the leakage related to most S-boxes is far from being linear, all these expected results confirm our theoretical results according to which AOV gives similar results as CPA when the leakage is linear but may give significantly better results when the leakage is far from being linear. AOV must therefore be preferred to CPA.

Fig. 9 and 10 show the evolution of the Mean Guessing Entropy for the DPAv2-AES and the 65nm-AES, but this time depending on the time taken by the attacks on our PC. As can be seen in Fig. 9, in a linear case, CPA and AOV require approximately the same number of traces to recover the key and thus took approximately the same amount of time. However, the MIA that also requires approximately the same number of traces in this case but is much more time consuming. Thus it takes more time to recover the key. On a non-linear case represented in Fig. 10, it is clear that the AOV is the best compromise between genericity and speed. However it is to notice that it only detects relations on the means, and it is not able to capture higher moments. Thus, in case of a leakage present in higher moments, the MIA should perform the best results as it would be the only one to recover the key despite its low computation time.

5 Conclusion

Despite the proposal of many distinguishers in the literature, the CPA remains the most used SCA. This choice is due to its simplicity of use and its low computation time. However, in most cases, the shape of the leakage of a device is unknown for an attacker. The distinguisher used by the CPA is the Pearson correlation. Thus, performing such an attack, that can only detect relations not too far from linearity, is an irrelevant choice without knowledge about the leakage.

In this context, the Analysis of Variance for SCA is a safer alternative and should be preferred. Indeed, it is more generic than CPA, while keeping approximately the same simplicity of use and the same computation time. It can recover the key in cases where the leakage deviates from linearity (leakage carried by the means) and where the CPA can't find it. And it should perform similar results to CPA in cases of a linear leakage, as theoretically and empirically showed in this paper.

However, it should be noticed that AOV is less generic than MIA that can detect any kind of relation between two variables, and not only dependence on their means. A further study of the difference between these two distinguishers would be interesting.

References

1. DPA contest v2. http://www.dpacontest.org/v2/index.php.

2. SASEBO project. http://www.risec.aist.go.jp/project/sasebo/.
3. Sébastien Aumonier. Generalized correlation power analysis. In *ECRYPT Workshop on Tools For Cryptanalysis*, Kraków, Poland, September 2007.
4. Lejla Batina, Benedikt Gierlichs, and Kerstin Lemke-Rust. Differential cluster analysis. In Clavier and Gaj [7], pages 112–127.
5. Eric Brier, Christophe Clavier, and Francis Olivier. Correlation power analysis with a leakage model. In Marc Joye and Jean-Jacques Quisquater, editors, *CHES*, volume 3156 of *Lecture Notes in Computer Science*, pages 16–29. Springer, 2004.
6. Mathieu Carbone, Sébastien Tiran, Sébastien Ordas, Michel Agoyan, Yannick Teglia, G. Ducharme, and Philippe Maurine. On adaptive bandwidth selection for efficient MIA. In *COSADE 2014*, April 2014.
7. Christophe Clavier and Kris Gaj, editors. *Cryptographic Hardware and Embedded Systems - CHES 2009, 11th International Workshop, Lausanne, Switzerland, September 6-9, 2009, Proceedings*, volume 5747 of *Lecture Notes in Computer Science*. Springer, 2009.
8. Benedikt Gierlichs, Lejla Batina, Pim Tuyls, and Bart Preneel. Mutual information analysis. In Elisabeth Oswald and Pankaj Rohatgi, editors, *CHES*, volume 5154 of *Lecture Notes in Computer Science*, pages 426–442. Springer, 2008.
9. Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential Power Analysis. In Michael J. Wiener, editor, *CRYPTO*, volume 1666 of *Lecture Notes in Computer Science*, pages 388–397. Springer, 1999.
10. Thanh-Ha Le and Maël Berthier. Mutual Information Analysis under the View of Higher-Order Statistics. In Isao Echizen, Noboru Kunihiro, and Ryôichi Sasaki, editors, *IWSEC*, volume 6434 of *Lecture Notes in Computer Science*, pages 285–300. Springer, 2010.
11. Stefan Mangard, Elisabeth Oswald, and François-Xavier Standaert. One for all - all for one: unifying standard differential power analysis attacks. *IET Information Security*, 5(2):100–110, 2011.
12. François-Xavier Standaert, Benedikt Gierlichs, and Ingrid Verbauwhede. Partition vs. comparison side-channel distinguishers: An empirical evaluation of statistical tests for univariate side-channel attacks against two unprotected cmos devices. In Pil Joong Lee and Jung Hee Cheon, editors, *ICISC*, volume 5461 of *Lecture Notes in Computer Science*, pages 253–267. Springer, 2008.
13. François-Xavier Standaert, Tal Malkin, and Moti Yung. A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In Antoine Joux, editor, *EUROCRYPT*, volume 5479 of *Lecture Notes in Computer Science*, pages 443–461. Springer, 2009.
14. Alexandre Venelli. Efficient Entropy Estimation for Mutual Information Analysis Using B-Splines. In Pierangela Samarati, Michael Tunstall, Joachim Posegga, Konstantinos Markantonakis, and Damien Sauveron, editors, *WISTP*, volume 6033 of *Lecture Notes in Computer Science*, pages 17–30. Springer, 2010.
15. Nicolas Veyrat-Charvillon and François-Xavier Standaert. Mutual information analysis: How, when and why? In Clavier and Gaj [7], pages 429–443.