On the Optimality of Correlation Power Attack on Embedded Cryptographic Systems

Youssef Souissi, Nicolas Debande, Sami Mekki, Sylvain Guilley, Ali Maalaoui, Jean-Luc Danger

On the Optimality of Correlation Power Attack on Embedded Cryptographic Systems

Youssef Souissi[1], Nicolas Debande[1,2], Sami Mekki[1],
Sylvain Guilley[1], Ali Maalaoui[3], Jean-Luc Danger[1].

[1] TELECOM ParisTech, 46 rue Barrault, 75634 Paris, France.
[2] Morpho, 95 523 OSNY, FRANCE.
[3] Rutgers University, NJ, USA.
Email: firstname.lastname@TELECOM-ParisTech.fr
This project is partially funded by the JST/ANR SPACES project.

Abstract. In this paper, we answer the question of what are the necessary conditions under which Correlation Power Attack (CPA), that essentially targets embedded cryptographic implementations, is optimal with regards to attacks that exploit the same leakage model. For this purpose, we offer an in-depth theoretical study which aims at determining the conditions under which the Pearson correlation coefficient is maximized. Moreover, we propose theoretical metrics to practically verify the validity of those conditions. Besides, we illustrate our theoretical study by an experiment on real electromagnetic traces acquired from a DES cryptographic implementation.

Keywords: Correlation Power Attack (CPA), Estimation theory, Security metrics, Spearman attack.

1 Introduction

Recently, E.Prouff *et al.* have shown in [1] that Side-channel distinguishers are not only asymptotically equivalent but also can be rewritten one in function of the other, only by modifying the power consumption model. In particular, they have established an equivalence between most univariate Side-channel distinguishers and Correlation Power Analysis (CPA) performed with different leakage models. Besides, based on the same statistical tool (*i.e.* Pearson coefficient), it is shown that it is possible to break protected implementations (masking countermeasure) by considering the leakage at different time samples. Such attacks, called *Higher-Order Power Correlations*, were suggested and investigated by T.Messerges in [2]. In this paper, we answer the question of what are the conditions under which CPA is optimal with regards to attacks that exploit the same leakage model. The answer we provide is principally based on *Estimation theory*. For more in-depth study about Estimation theory, we refer the reader to [3–5]. The overall goal of this study is to put the Correlation Power Analysis on a sound theoretical basis, and therefore brighten the task of an evaluator when assessing the robustness of secure embedded systems against CPA.

The rest of the paper is organized as follows: first, in Section 2, we discuss the optimality of CPA. Actually, we define the sufficient conditions to maximize Pearson correlation coefficient, thereby reaching the optimality of CPA. Second, in Section 3, we propose theoretical and practical metrics to validate those conditions. Third, in Section 4, we illustrate the theoretical study by an experiment on real electromagnetic traces acquired from a DES cryptographic implementation. Eventually, we conclude the paper in Section 5.

2 The Optimality From the Estimation Theory View Point

The Approximation Problem Suppose we want to best approximate Y with another variable X based on their joint distribution. The approximation problem is to seek for a function $\phi(\cdot)$ of X that best fits Y among all possible forms of $\phi(\cdot)$. We write $\widehat{Y} = \phi(X)$ and we call $\widehat{Y}$ an estimator of Y. In our study, the variable X is deterministic since it is theoretically predicted from a known cryptographic process. Whereas, the variable Y is a real measure acquired by an oscilloscope. For sake of clarity, in what follows the variable X is called *the prediction* and Y *the measurement* (or the observation). Let $\epsilon = Y - \widehat{Y}$ denotes the error in estimating Y, and let $pos(\epsilon) = pos(Y, \widehat{Y})$ denotes a non negative function of ϵ. $pos(\epsilon)$ can be for instance the absolute difference or the square difference between Y and $\widehat{Y}$ (*i.e.* $|Y - \widehat{Y}|$ or $(Y - \widehat{Y})^2$ respectively). The average cost, *i.e.*, $\mathbb{E}[pos(Y, \widehat{Y})]$, is referred to as the *Bayes risk* $\Re_B$. Obviously, the approximation problem comes down to a minimization problem. In fact, minimizing the *Bayes risk* with respect to $\widehat{Y}$ for a given cost function is a proper solution of the problem. The most popular $\Re_B$ is the *Mean Square Error (MSE)*, since it is parameter free, straightforward to implement and memory-less. The MSE measures the average of the squares of the errors. In this case, it is clear that $pos(Y, \widehat{Y}) = (Y - \widehat{Y})^2$. In what follows, we will focus on the important role played by the MSE in the approximation problem. There are several ways in which the role of the MSE can be introduced. A particular way for especial convenience is to work with the L^2 space that is defined as the space of square summable variables[1]. If Z is a random variable belonging to this space, then the corresponding norm, called L_2 norm, is expressed as $\|Z\|_2 = \sqrt{\mathbb{E}[Z^2]}$; so that the distance between two elements Z_1 and Z_2 of L^2 space can be written as $\|Z_1 - Z_2\|_2 = \sqrt{\mathbb{E}[(Z_1 - Z_2)^2]}$. Z_1 and Z_2 are said to be *orthogonal* ($Z_1 \perp Z_2$) if and only if $\mathbb{E}[(Z_1 Z_2)] = 0$. Orthogonality property and mean square convergence will allow us in the following to introduce the notion of optimal estimation in the sense of L_2 norm. With these notations, the optimal estimator of Y given X, in the sense of the L_2 norm, is the function $\widehat{Y} = \phi(X)$ for which $\|Y - \widehat{Y}\|_2^2$ is a minimum [4]. But more importantly, it is proved that the conditional expectation $\widehat{Y} = \mathbb{E}[Y|X]$ is the estimator that gives such a minimum. Incidentally, using the error notation, ϵ, the MSE is written in the following form:

$$MSE(\widehat{Y}) = \mathbb{E}[\epsilon^2] = \|Y - \widehat{Y}\|_2^2 \ .$$

[1] The L^2 space is also often referred to as a weighted Euclidean norm.

Besides, in [6], it is shown that MSE can be expressed as follows:

$$MSE(\widehat{Y}) = Var(\widehat{Y}) + bias(\widehat{Y})^2 \ ,$$

where $Var(\widehat{Y})$ is the variance of $\widehat{Y}$ and $bias(\widehat{Y}) = \mathbb{E}(\widehat{Y}) - Y$. Note that for an unbiased estimator (*i.e.* $bias = 0$), the MSE is just the variance of the estimator. In the literature of estimation theory [7], two naturally desirable properties of estimators are for them to have minimal MSE and to be unbiased. Common criteria for estimation are Maximum Likelihood Estimator (MLE), Minimum Mean Squared Error (MMSE) and Maximum A Posteriori Probability (MAP [4]). From the theoretical point of view, MLE approach is more efficient than the rest of criteria. But more importantly, estimation theory says that no asymptotically unbiased estimator has lower MSE than the MLE (see *Cramer-Rao Lower Bound theory*) [8–10]. However, in practice, statisticians prefer using MMSE estimator, specifically in the linear case, which is in fact the approach that minimizes the MSE in the sense of the L_2 norm, because of its simplicity relatively to the other criteria. Additionally, later on, we will show that, under few assumptions, MMSE estimator produces the lowest MSE among all estimators, in particular unbiased ones, and can be derived as a maximum likelihood estimator.

Optimal Linear MMSE Estimation & Connection with Pearson coefficient As stated before, the conditional expectation is the optimal estimator in the sense of the L_2 norm, which is indeed the MMSE estimator. Hence, the MSE can be rewritten as $MSE(\widehat{Y}) = \mathbb{E}[\epsilon^2] = \|Y - \mathbb{E}[Y|X]\|_2^2$. A useful property of the MMSE estimator is that the estimation error $Y - \mathbb{E}[Y|X]$ is orthogonal to every function of the variable X. This property is known as the *Orthogonality Principle* that provides a necessary and sufficient condition for the optimal estimation in the L^2 space. More formally, $\phi(X)$ is the MMSE estimator $\hat{Y}_{MMSE}$ if and only if the error $Y - \phi(X)$ is orthogonal to every function $\gamma(X)$ that is:

$$\mathbb{E}\left[(Y - \phi(X)) \cdot \gamma(X)\right] = 0 \ . \tag{1}$$

Now, the problem is that MMSE is very general; and therefore, the conditional expectation can be complicated to compute. Nonetheless, the analysis is very simple when the *linear assumption* is made (*i.e.* Linear MMSE, often termed by LMMSE). For this purpose, statisticians usually make such assumption as a first approximation. However, when the true data does not fit the linear case, we say that LMMSE is sub-optimal to the optimal estimate of MMSE. In the context of side-channel analysis, the linear case has a pure theoretical flavour for us especially when considering unprotected implementations. But it is noteworthy that even for unprotected implementations it is possible to have recourse to what we call *linear transformations* [11]; and therefore to fall into the linear case. In "Introduction to optimal estimation" book ([4] Chapter 3), using the orthogonality principle (Eqn. (1)), authors show that when the true data fits exactly the linear case (*i.e.* LMMSE is optimal) the associated MSE of $\hat{Y}_{LMMSE}$ is expressed with Pearson coefficient ρ, as follows:

$$MSE_{LMMSE} = {\sigma_Y}^2(1 - \rho_{X,Y}^2) \ .$$

In the linear case, $\hat{Y}_{LMMSE}$ is the optimal estimate in the sense of MMSE estimation. But more importantly and always from the MMSE estimation point of view, it is clear that ρ is the optimal metric to measuring the linear association between involved variables. Actually, the **maximization** of ρ^2 implies the **minimization** of MSE_{LMMSE}.

Limitations of Optimal MMSE Estimation Up to this point, we have shown that in the linear case Pearson correlation coefficient is an optimal indicator of linearity in the sense of MMSE estimation. However, the MMSE does not make any assumption about the joint distribution. One may ask: is the Pearson correlation still the best linear indicator even if the joint distribution is not bivariate normal? Indeed, the fact that the MMSE is distribution free[2] is often seen as a weak point in the estimation literature, specifically when performing a linear estimation (LMMSE). Generally, when no assumption is made about the joint distribution, it exists two important cases in which the optimality of LMMSE, relatively to all estimators, is not guaranteed. In other words, in these cases LMMSE does not give the lowest MSE among the other estimators such as the Maximum Likelihood Estimator (MLE).

Case 1: Heteroscedasticity This basically occurs when the error of estimation ϵ depends on the prediction X. The LMMSE only states that the error of estimation ϵ is uncorrelated with the prediction X. In the linear case, this statement follows since $Cov(X, \epsilon)$ is null. However, $Cov(X, \epsilon) = 0$ does not imply the independence of X and ϵ. In other words, even if the linear estimation is optimal in the sense L_2 norm (*i.e.* $\mathbb{E}[Y|X] = \alpha + \beta X$), it could exist a relation between X and ϵ which compromises the efficiency of the LMMSE in estimating the parameters α and β of the linear model. In this case, the linear model is said to display a *heteroscedasticity*. A frequent situation of heteroscedasticity is that the error is linearly increasing with the values taken by the prediction X. For such situation, it is easy to verify that the MLE estimator is more efficient than the LMMSE as it produces the lowest MSE ([12] page 398).

Case 2: Imperfect data (*aka* outliers problem) The data, which is composed by the prediction X and the measurement Y, is often disturbed by the presence of what we call outliers. An *outlier* can vaguely be defined as an observation which shows a different behaviour with regards to observations composing the data. The reason might be due to the type of variables (continuous, discrete) and the shape of the marginal distributions of X and Y respectively.

Overall Optimality of MMSE According to **case 1** and **case 2**, the MMSE is not sufficient to **totally** characterise the dependence between X and Y, even

[2] In statistics, a statistical criterion that does not make any assumption about the joint distribution is said to be "distribution free".

if the true relationship between them is **linear**. More importantly, the Pearson correlation coefficient could not be considered as the best linear metric to measuring the true relationship. In statistic, several candidates exist, such as Spearman, Kendall or intra-class coefficient correlations, that are designed to be less sensitive (more robust) to outliers or heteroscedasticity and therefore they would be better than Pearson coefficient. However, the estimation theory proved that there exists one and only one condition if satisfied then the MMSE is equivalent to the MLE; and therefore considered to be the optimal estimator among all estimators, in particular unbiased ones, as it gives the lowest MSE. Thus, the Pearson coefficient ρ is the best metric for measuring a linear association. This condition requires that the joint distribution should be bivariate normal [4, 13]. In fact, under the Gaussian assumption, the true relationship is **linear**, not heteroscedastic (*i.e.* homoscedastic) and not disturbed by some undesirable effects like the outliers. Note that in this case the error of estimation follows a normal distribution. Hence, we can state that ρ is the best linear metric only when the true relationship in the MMSE sense satisfies the Gaussian assumption. If such assumption is not validated, the MMSE is less efficient than MLE; and therefore the optimality of CPA is compromised.

Sufficient Conditions for the Optimality A common pitfall about the validity of the Gaussian assumption is to check only that X and Y are drawn from normal distributions. This is not sufficient. Indeed, if X and Y are each individually Gaussian then this does not imply that they are jointly Gaussian. Generally, a joint distribution is said to be bivariate normal if all following conditions are satisfied ([14] page 54):

1. **Linearity** The true relationship between X and Y is linear.
2. **Normal conditional distribution** The conditional distribution of Y given $X = x$ is normal.
3. **Homoscedasticity** The conditional distribution of Y given $X = x$ has a constant variance (*i.e.* the variance of the error) for each x.
4. **Normal marginal distribution** The marginal distribution of X is normal (Gaussian).

We note that, the last condition is independent from the measurements; it is only dependent on the predictions that are provided by the leakage model. Moreover, under these conditions, the error ϵ must be drawn from zero mean normal distribution. In other words, ϵ is a random variable strictly independent from X and that a linear function ϕ characterizes the dependence between X and Y, entirely. In practice, these conditions are not supposed to be strictly verified but to hold to a certain degree. Actually, in real situations, it is mostly hard to get a perfect binormal joint distribution. In such situations, the higher the departure from the Gaussian assumption is, the lower the efficiency of Pearson correlation coefficient ρ will be.

3 Practical Metrics Computation

3.1 Deviation from Linearity Metric (*DLM*)

In statistics, the *Correlation ratio* coefficient [12] between X and Y is defined as follows:

$$\eta_{Y|X}{}^2 = \frac{Var\big[\mathbb{E}[Y|X]\big]}{Var(Y)} = 1 - \frac{\mathbb{E}\big[Var[Y|X]\big]}{Var(Y)}\,.$$

Unlike the Pearson correlation coefficient ρ which only detects the linear dependency between two variables, the *Correlation ratio* measures the functional dependency. In other words, η quantifies the dependency strength whatever the relation between the two variables, linear or non linear. Similarly to $\rho^2_{X,Y}$, the Correlation ratio takes on values between 0 and 1. The higher the value of η is, the higher the functional dependency is. Furthermore, η is asymmetric (*i.e.* $\eta_{Y|X} \neq \eta_{X|Y}$) since the two variables fundamentally do not play the same role in the functional relationship. In the general context of this paper, the most important additional properties of η are those which characterize the relation between η and ρ. These properties are summarized as follows:

$$\begin{aligned} &\eta_{Y|X}{}^2 \geq \rho_{X,Y}{}^2\,. \\ \eta_{Y|X}{}^2 = \rho^2_{X,Y} \quad &\Longleftrightarrow \quad \exists(\alpha,\beta),\ \mathbb{E}[Y|X] = \alpha + \beta X\,. \end{aligned} \tag{2}$$

From (2), we can design a new metric which aims at measuring the deviation from a perfect linear relationship. This metric that we name *Deviation from Linearity Metric (DLM)* is expressed by the ratio between the squared Pearson coefficient and the Correlation ratio as follows:

$$DLM = \frac{\rho^2}{\eta^2} \in [0,1]\,.$$

The DLM ratio takes on values between 0 when the relation is totally curved and 1 when it is perfectly linear.

3.2 Deviation from Normality Metric (*DNM*)

In the literature, there exist many variants of statistical tools (*e.g.* Shapiro-Wilk test, Lilliefors test, D'Agostino test and Jarque-Bera test [12]) that aim at measuring the deviation from a normal distribution. For sake of simplicity and convenience, we used the *Jarque-Bera* (JB) test as a "Deviation from Normality Metric" (DNM_{JB}) as it is mainly based on the computation of the skewness (S) and the kurtosis (K) ([12]). Hence, the Jarque-Bera metric can be defined as follows:

$$DNM_{JB} = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right)\,,$$

where n is the number of observations. The smaller the value of DNM_{JB}, the better the approximation by a normal distribution. Indeed, in the case of a perfect normal distribution, we have $DNM_{JB} = 0$.

3.3 Deviation from Homoscedasticity Metric (DHM)

As stated before, homoscedasticity simply requires that the conditional variances $Var(Y|X = x)$ are equals. For this purpose, we propose to compute the *Coefficient of variation* [12] based dispersion of such conditional variances to measure the deviation of data from homoscedasticity. The coefficient of variation, which is a normalized metric for dispersion, is defined as the ratio of the standard deviation to the mean. Hence, the metric proposed (DHM) can be expressed as follows:

$$DHM = \frac{\sqrt{Var(Var(Y|X = x))}}{\mathbb{E}(Var(Y|X = x))}.$$

The smaller the dispersion DHM, the better the homoscedasticity. Actually, in the case of a perfect homoscedasticity, we have $DHM = 0$.

4 Experiments on Real DES Cryptographic Implementation

In this study, we are interested in the basic attack of the Data Encryption Standard (DES), that targets the two first rounds of the encryption algorithm. For this purpose, we acquired real electromagnetic leakage traces from an unprotected DES implementation based ASIC (Secmat V1) [15] circuit (easy to attack by CPA). Recall that the DES implementation is composed of eight different Sboxes. There are thus eight secret keys to retrieve[3]. Therefore, in this case, a Hamming distance HD model can take five possible values, $HD = \{0, 1, 2, 3, 4\}$, and 2^6 key hypotheses are required to break one Sbox. Now, let us analyse the marginal distributions of the prediction X. The variable X that is represented by the values taken by HD, is a discrete type variable following a symmetric distribution [16] $\beta_{(n_b=4,p=\frac{1}{2})}$ where n_b represents the predicted bits in the targeted register R, and p is the success probability. In statistics, if n_b is large, say $n_b > 20$, and $p = \frac{1}{2}$, then the binomial distribution is approximately equal to the normal distribution [12, 17, 18]. In our case $n_b \ll 20$, therefore X can not be strictly approximated to a normal distribution. We note that for a binomial distribution, the Skewness (S) and the Kurtosis (K) are expressed by the parameters n_b, p and q [12]. In our experiment, we assume that enough traces are acquired. Thus, $p = q = 0.5$. Hence, the DNM_{JB} metric defined previously, to measure the deviation from a normal distribution, is computed as follows:

$$DNM_{JB} = \frac{n_b}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right) = \frac{4}{6}\left(0^2 + \frac{1}{4}(2.5-3)^2\right) = 0.041.$$

Up to this point, we have only studied the marginal distribution of the prediction X and calculated theoretically its deviation from a perfect normal distribution. Therefore, three other conditions are still to be verified to assume the Gaussian case. Fig. 1 represents a CPA differential trace superposed with DNM_{JB} (for

[3] For misuse of language, we often use the notion of "broken Sbox" to say that the secret key corresponding to the attacked Sbox is found.

example HD = 4), DLM and DHM metrics. The idea behind calculating such metrics over the whole time samples of the real DES traces is to reveal more details, especially when the knowledge about the implementation is not total to know exactly where the secret information is happened. In this figure, we distinguish four zones ($Z1$, $Z2$, $Z3$ and $Z4$), depending on the distribution of the real leakage over the time samples. In fact, the secret information corresponds to the zones $Z2$, $Z3$ and $Z4$ in which CPA differential trace shows high values (*i.e.* peaks). But more importantly, we show that the proposed metrics are in agreement with our theoretical study, as $DLM \simeq 1$, $DHM \simeq 0$ and $DNM \simeq 0$. Therefore, CPA performance is close to the optimality (Gaussian assumption). However, $Z1$ is not suitable for CPA as the linearity metric (DLM) is virtually equal to zero. Besides, $Z4$ can not be the best zone for CPA (CPA peak is not high), because the deviations from normality DNM and homoscedasticity DHM are relatively high. In what follows, we will be interested only in the

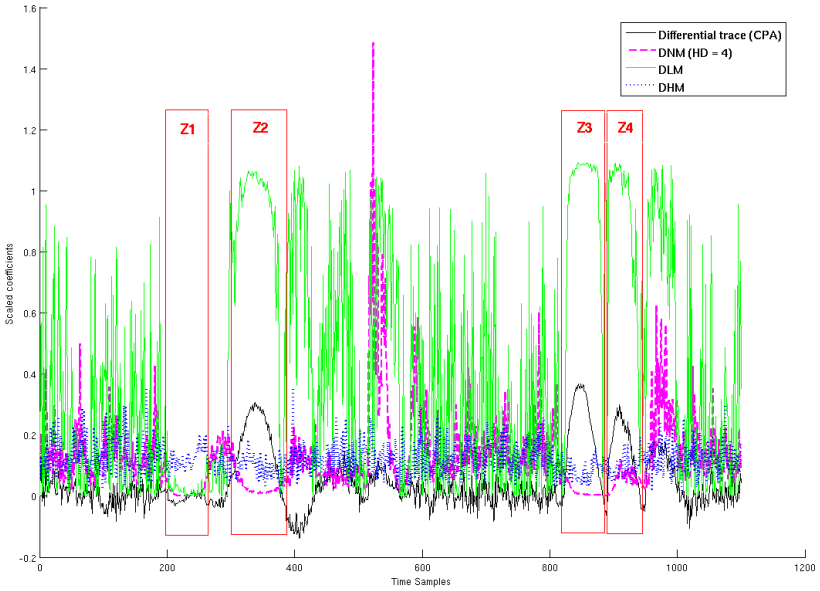

Fig. 1. Illustration of Gaussian assumption metrics on unprotected DES.

zone $Z3$ as it shows better performance of CPA than it does for $Z2$ and $Z4$. Actually, we conducted two operations: the first operation consists in performing and comparing the efficiency of two correlation attacks based on Pearson (CPA) and Spearman coefficients, respectively. We note that Spearman correlation has been developed to be more robust[4] than the Pearson correlation. Spearman correlation measures both the linear and the non-linear relationship between the two variables, as it does not require that the observations are drawn from a

[4] A statistical criterion that does not make any assumption about the joint distribution is said to be robust or distribution free.

bivariate normal distribution. It is a *non-parametric* correlation that was first applied in side-channel context in [19]. In this operation, we will show that the conditions required to maximize the Pearson coefficient (Gaussian assumption) hold to a sufficient degree, which makes the CPA more powerful than Spearman attack. The second operation consists in degrading the quality of acquired traces, by creating outliers, in order to simulate a situation in which the deviation from the Gaussian assumption is relatively high. Therefore, we show that in such case, despite the fact that the true relationship is linear, CPA becomes less powerful than Spearman attack. Recently, an evaluation metric has been proposed in [20] to assess the performance of Side-channel analysis: the *Guessing Entropy*, termed by GE. In fact, GE metric measures the average position of the secret key in a list of key hypotheses ranked by a statistical test referred to as distinguisher (*e.g.* Pearson coefficient, Spearman coefficient). According to Fig. 2, as expected, CPA outperforms Spearman attack when applied on original traces. Actually, to reach the five first ranks, we need 12 traces for CPA. Whereas, 18 traces are needed for Spearman attack. However, when the quality of original set of traces is degraded by creating outliers, Spearman attack becomes more powerful than CPA; which is in agreement with our study.

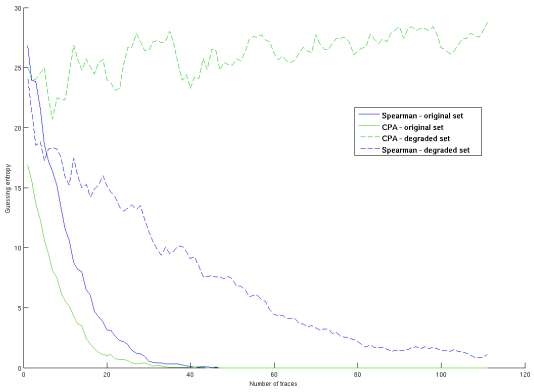

Fig. 2. GE for CPA and Spearman attacks on original and degraded traces.

5 Conclusion

In this paper, we have studied the efficiency of Correlation Power Attack (CPA) from the estimation theory point of view. This study is useful in that it allows to assess the performance of CPA; and therefore to decide on the choice of an appropriate Side-channel distinguisher for the analysis. Actually, if the Gaussian assumption is satisfied, then CPA must be the best analysis to quantify the secret leakage. Besides, if these conditions do not hold to a certain degree, then CPA might not be efficient and therefore is not the best analysis anymore. In this case, more powerful attacks and distinguishers like Spearman coefficient should be investigated.

References

1. Doget, J., Prouff, E., Rivain, M., Standaert, F.X.: Univariate side channel attacks and leakage modeling. Journal of Cryptographic Engineering **1** (2011) 123–144
2. Messerges, T.S.: Using second-order power analysis to attack dpa resistant software. In: Proceedings of the Second International Workshop on Cryptographic Hardware and Embedded Systems. CHES '00, London, UK, Springer-Verlag (2000) 238–251
3. Mendel, J.: Lessons in Estimation Theory for Signal Processing, Communications, and Control. Pearson Education (1995)
4. Kamen, E., Su, J.: Introduction to optimal estimation. Advanced textbooks in control and signal processing. Springer (1999)
5. Candy, J.: Bayesian Signal Processing: Classical, Modern and Particle Filtering Methods. Adaptive and Learning Systems for Signal Processing, Communications and Control Series. John Wiley & Sons (2011)
6. Anderson, S.: Statistical methods for comparative studies: techniques for bias reduction. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley (1980)
7. Edward W. Kamen, J.S.: Introduction to optimal estimation Advanced textbooks in control and signal processing Control and Signal Processing Series. Springer (1999)
8. Bar-Shalom, Y., Li, X., Kirubarajan, T.: Estimation with applications to tracking and navigation. A Wiley-Interscience publication. Wiley (2001)
9. Bos, A.: Parameter estimation for scientists and engineers. Wiley-Interscience (2007)
10. Sorensen, D., Gianola, D.: Likelihood, Bayesian and MCMC methods in quantitative genetics. Statistics for biology and health. Springer-Verlag (2002)
11. Sharma, A., Prakash, M.: Linear Transformation. Discovery Publishing House (2007)
12. Saporta, G.: Data mining et statistique décisionnelle. L'intelligence des données. Technip (2010)
13. Proakis, J., Salehi, M.: Digital communications. McGraw-Hill higher education. McGraw-Hill (2008)
14. Arnold, B., Castillo, E., Sarabia, J.: Conditional specification of statistical models. Springer series in statistics. Springer (1999)
15. Guilley, S.: Documentation technique de la conception physique (ou *back-end*) du circuit SECMAT (2006) `http://perso.telecom-paristech.fr/~guilley/backend.pdf`.
16. Guilley, S.: Geometrical Counter-Measures against Side-Channel Attacks. PhD thesis, ENST / CNRS LTCI (2007) 219 pages; Id: `2007 E 003`, `http://pastel.paristech.org/2562/`.
17. Russo, R.: Statistics for the behavioural sciences: an introduction. Psychology Press (2003)
18. Rosner, B.: Fundamentals of biostatistics. Brooks/Cole Cengage Learning (2010)
19. Batina, L., Gierlichs, B., Lemke-Rust, K.: Comparative Evaluation of Rank Correlation Based DPA on an AES Prototype Chip. In: ISC. Volume 5222 of Lecture Notes in Computer Science., Springer (2008) 341–354 Taipei, Taiwan.
20. Standaert, F.X., Malkin, T., Yung, M.: A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In: EUROCRYPT. Volume 5479 of LNCS., Springer (2009) 443–461 Cologne, Germany.