On the Exact Success Rate of Side Channel Analysis in the Gaussian Model

Matthieu Rivain

Oberthur Technologies & University of Luxembourg
m.rivain@oberthurcs.com

Abstract. Nowadays, Side Channel Analysis is one of the most powerful cryptanalytic technique against cryptosystems embedded in portable devices such as smart cards. Faced with this threat, it is of crucial importance to precisely determine what is achievable by a given side channel adversary against a cryptosystem producing a given side channel leakage. This can be answered by evaluating the success rate of an attack according to the adversary capacities and to the leakage properties.
In this paper, we investigate the issue of evaluating the success rate of side channel analysis in the widely admitted Gaussian leakage model. We introduce a new approach that allows us to efficiently compute the success rate of an attack in this model and we apply it to the two main families of side channel analysis: differential side channel analysis and profiling side channel analysis.

1 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 (*e.g.* execution time [13], power consumption [12], electromagnetic emanations [8]). Some kinds of SCA exploit this *side channel leakage* to recover information on the operation flow that may depend on the secret key (*e.g.* Simple SCA [12], Timing Attacks [13]). These can be circumvent by ensuring that the operation flow is independent of the secret key. Other kinds of SCA exploit the fact that the side channel leakage is statistically dependent on the intermediate variables of the computation. Some of these variables are themselves related to small parts of the secret key which enables key recovery attacks. These second kinds of SCA are particularly powerful and securing cryptographic implementation against them constitutes a real challenge.

SCA targeting intermediate variables divides into two main categories: *differential SCA* and *profiling SCA*. Differential SCA relies on correlation techniques [12, 4]. Based on several leakage observations, the attacker estimates a correlation between the leakage and different predictions on the value of a key-dependent intermediate variable. According to the obtained correlation values, the attacker is able to (in)validate some hypotheses on the secret key. Profiling SCA [6, 19] is based on the maximum likelihood approach. It assumes that the attacker owns a profile of the leakage according to the values of some key-dependent intermediate variables. This profile is involved to derive the likelihood of some key hypotheses given the observed leakage.

Faced with the threat of side channel analysis, a crucial issue is to quantify the efficiency of the different attacks according to the adversary capacities and the leakage statistical properties. For this purpose, a natural metric is the success rate, namely the probability that an attack succeeds in recovering the correct key (or in isolating it in a restricted set). A straightforward way to evaluate the success rate is to estimate it empirically by performing the attack several times. However such an approach is costly in time and may even become impossible for attacks with medium or high complexity. It is therefore not suitable to efficiently and precisely determine the resistance of an embedded device if this one is not quite weak. To tackle this issue, it is of particular interest to investigate efficient ways to compute (or at least to precisely estimate) the success rate of an attack without requiring to perform it many times.

Previous investigations have been done regarding this issue [7, 16, 22]. These works investigate differential SCA in a noisy context. They provide an approximation of the required number of leakage measurements for a successful attack [7, 16] and an approximation of the success rate [22]. For the sake of generality, these works do not take into account the relationship between the different key candidates (which depends on the target algorithm logical properties and on the leakage statistical properties) and only focus on the good key guess. However, the success rate depends on the joint behavior of the different candidates and this relationship cannot be neglected while looking for a precise estimation of the success rate. Concerning profiling SCA, to the best of our knowledge no solution for the success rate evaluation has been proposed in the literature so far. This is a lack since these attacks are considered as the strongest form of side channel analysis.

In this paper, we address the issue of evaluating the success rate of a side channel key recovery attack. We analyze both differential SCA and profiling SCA under the widely admitted assumption that the noise in the leakage has a Gaussian distribution. We show that the result of these attacks can be expressed as a multivariate Gaussian random variable which leads to an efficient way for determining their success rates.

The rest of the paper is organized as follows. Section 2 introduces some preliminaries. Section 3 presents the side channel theoretical model considered in this paper. In Sections 4 and 5, we respectively analyze differential SCA and profiling SCA. Based on these analyses, Section 6 shows how to efficiently evaluate the success rate of the focused attacks. Finally an empirical validation is provided in Section 7 and concluding remarks are given in Section 8.

2 Preliminaries

The calligraphic letters, like $\mathcal{X}$, are used to denote finite sets (*e.g.* $\mathbb{F}_2^n$). The corresponding large letter X denotes a random variable over $\mathcal{X}$, while the lowercase letter x denotes a particular realization of X. The probability of an event ev is denoted by $\mathrm{P}[ev]$. In case X has a continuous distribution, the notation $x \mapsto P[X = x]$ is further used to denote the probability density function (pdf) of X. The expectation and the variance of a random variable X are respectively denoted by $\mathrm{E}[X]$ and $\mathrm{Var}[X]$. The covariance between two random variables X and Y is denoted by $\mathrm{Cov}[X, Y]$. The Gaussian distribution of dimension T with T-size expectation vector m and $T \times T$ covariance matrix $\varSigma$ is denoted by $\mathcal{N}(m, \varSigma)$, and the corresponding pdf is denoted by $\phi_{\varSigma,m}$. We recall that this pdf is defined for every $x \in \mathbb{R}^T$ as:

$$\phi_{\varSigma,m}(x) = \frac{1}{\sqrt{(2\pi)^T |\varSigma|}} \exp\left(-\frac{1}{2}(x - m)' \, \varSigma^{-1} \, (x - m) \right) ,$$

where $(x - m)'$ denotes the transpose of the vector $(x - m)$ and $|\varSigma|$ denotes the determinant of the matrix $\varSigma$. If the dimension T equals 1, then the Gaussian distribution is said to be *univariate* and the single element of the covariance matrix is the variance that is denoted by σ^2. If $T > 1$, the Gaussian distribution is said to be *multivariate*.

3 Side Channel Model

A formal modeling of side channel key recovery attacks has been initiated by Standaert *et al.* in [21]. The theoretical model introduced hereafter follows the outlines of their work.

3.1 Side Channel Key Recovery Attacks

Let E_{SK} be a cryptographic algorithm E parameterized by a secret key SK. Let K be a random variable representing a guessable part of SK. Let X be a random variable representing a part of a public value such as an input (resp. output) of E_{SK}. Let S be a random variable representing the result of an intermediate computation of E_{SK} that satisfies $S = \varphi(X, K)$ for a given function $\varphi : \mathcal{X} \times \mathcal{K} \to \mathcal{S}$. We denote by L the random variable that represents the side channel leakage generated by the computation (and/or the handling) of S on a physical implementation of E_{SK}. We shall further denote by $\mathrm{L}(s)$ the random variable $(\mathrm{L}|S = s)$.

A side channel key recovery attack targeting the signal S aims at recovering the value k^* taken by K on a given physical implementation of E_{SK}. For such a purpose, the attacker collects several, say N, leakage measurements $(\mathrm{l}_i)_i$ resulting from the computation of $\varphi(x_i, k^*)$ for N inputs $(x_i)_i$. Namely, the l_i's are realizations of the random variables $\mathrm{L}(\varphi(K, x_i))$ that are assumed to be mutually independent. Then, the attack makes use of a *distinguisher*, that is a function D which, from the leakage measurements vector $\mathbf{l} = (\mathrm{l}_1, \cdots, \mathrm{l}_N)$ and the corresponding inputs vector $\mathbf{x} = (x_1, \cdots, x_N)$, outputs a *distinguishing vector* $\mathbf{d} = (d_k)_{k \in \mathcal{K}}$. If the distinguisher is sound and if the leakage brings enough information on S, then $k^* = \mathrm{argmax}_{k \in \mathcal{K}} \, d_k$ holds with a non-negligible probability.

Finally, a side channel adversary can be defined as the composition of a distinguisher with a strategy to select the algorithm inputs *i.e.* the x_i values. These can be randomly drawn (in a known plaintext/ciphertext attack setting) or they can be chosen by the adversary (in a chosen plaintext attack setting). In this paper, we do not assume a specific strategy. Rather, we investigate the success rate of an attack according to the inputs vector $\mathbf{x}$.

3.2 Gaussian Leakage Model

In practice, the leakage measurements are composed of several samples, say T, corresponding to several successive instants in time. The leakage L can hence be modeled by a T-size random vector. In the *Gaussian leakage model*, the leakage $\mathrm{L}(s)$ resulting from the computation of any signal $s \in \mathcal{S}$ has a Gaussian distribution: $\mathrm{L}(s) \sim \mathcal{N}(m_s, \Sigma_s)$.

Remark 1. The Gaussian model assumption is both very usual in the side channel literature (see for instance [6, 15, 19, 21]) and fairly realistic in practice (see for instance [15, §4]).

For clarity and without ambiguity, we shall respectively denote by m_{x,k^*} and Σ_{x,k^*} the mean vector $m_{\varphi(x,k^*)}$ and the covariance matrix $\Sigma_{\varphi(x,k^*)}$.

3.3 Success Rate

The success rate is a classical metric in side channel analysis. Usually, a key recovery attack is considered successful if the distinguishing vector satisfies $k^* = \mathrm{argmax}_{k\in\mathcal{K}}\, d_k$. In [21], the authors propose to extend the notion of success rate to different orders. The o^{th} order success rate of a side channel attack using a distinguisher D and a public vector $\mathbf{x}$, and targeting a secret key k^* is defined as:

$$\text{Succ-}o^{\mathsf{D}}_{\mathbf{x},k^*} = \mathrm{P}\left[\left(\mathrm{l}_i \leftarrow \mathrm{L}\left(\varphi(k^*, x_i)\right)\right)_i \;;\; \mathbf{d} \leftarrow \mathsf{D}(\mathbf{x}, \mathbf{l}) \;:\; k^* \in \underset{k\in\mathcal{K}}{\text{argmax-}o}\; d_k\right] \;,$$

where argmax-$o_{k\in\mathcal{K}}$ d_k denotes the set of the o elements $k \in \mathcal{K}$ that maximize d_k. The notion of order is motivated by the fact that an attacker may perform an off-line exhaustive search after the side channel analysis. A o^{th} order success means that the attacker has at the most o key guesses to test after the attack in order to recover the correct one.

Remark 2. In [21], the authors also suggest to use another metric: the so-called *guessing entropy* [17, 5]. This one is defined as the expected rank of the good key guess in the distinguishing vector, namely it indicates the average number of key guesses to test after the side channel analysis. This notion is discussed in Appendix A where we show that it can be expressed with respect to the success rates at the different orders.

Our Approach. In order to determine the exact success rate of an attack, we must investigate the multivariate probability distribution of the distinguishing vector. This distribution can be expressed with respect to the inputs vector $\mathbf{x}$, the secret key k^* and the leakage distribution parameters $(m_s, \Sigma_s)_{s \in \mathcal{S}}$. In the rest of the paper, we will investigate the two main families of side channel analysis: differential SCA and profiling SCA. We will show that under the Gaussian assumption, the multivariate distribution of the distinguishing vector is (or at least can be precisely approximated by) a multivariate Gaussian distribution. This will enable us to show how the success rate of such attacks can be efficiently computed.

4 Differential Side Channel Analysis

4.1 Description

Differential side channel analysis uses correlation techniques as distinguisher. Several variants have been proposed in the literature [1, 4, 3, 12, 18]. In this paper, we focus on the Pearson correlation coefficient since it is the most widely used and seems to be the most efficient technique in practice [4]. Note that our analysis could be easily extended to other differential distinguishers that rely on correlation computations [1, 3, 12, 18]. The adversary is assumed to own a model of the side channel leakage that is a function $\mathsf{M} : \mathcal{X} \times \mathcal{K} \to \mathbb{R}$ such that $\mathsf{M}(x, k)$ is linearly related to the expectation of the leakage $\mathrm{L}(\varphi(x, k))$. The attack consists in estimating, for every key guess $k \in \mathcal{K}$, the linear correlation between the *prediction* $\mathsf{M}(X, k)$ (*i.e.* the predicted value of the leakage for the guess k) and the observable leakage $\mathrm{L}(\varphi(X, k^*))$. This correlation is estimated based on the prediction vector $\big(\mathsf{M}(x_1, k), \cdots, \mathsf{M}(x_N, k)\big)$ and the leakage measurements vector $\mathbf{l}$ by the following coefficient:

$$\rho_k = \frac{\frac{1}{N}\sum_i \Big(\mathsf{M}(x_i, k) - \frac{1}{N}\sum_j \mathsf{M}(x_j, k)\Big)\Big(\mathrm{l}_i - \frac{1}{N}\sum_j \mathrm{l}_j\Big)}{\sqrt{\frac{1}{N}\sum_i \Big(\mathsf{M}(x_i, k) - \frac{1}{N}\sum_j \mathsf{M}(x_j, k)\Big)^2}\sqrt{\frac{1}{N}\sum_i \Big(\mathrm{l}_i - \frac{1}{N}\sum_j \mathrm{l}_j\Big)^2}} \ . \tag{1}$$

If the model is sound, the prediction vector for the correct key guess is significantly correlated to the leakage measurements vector. As a result, for N large enough, ρ_k is expected to be maximal for $k = k^*$.

Since the correlation distinguisher takes as input a set of 1-size leakage measurements, we investigate hereafter the distribution of this distinguisher in the univariate Gaussian model.

4.2 Distinguisher Distribution

Let us first denote by τ_x the occurrence ratio of an element $x \in \mathcal{X}$ through the inputs vector $\mathbf{x}$, *i.e.* :

$$\tau_x = \frac{|\{i; x_i = x\}|}{N} \ . \tag{2}$$

We shall further denote by $\overline{\mathsf{M}}_k$ and $\widehat{\sigma}_k$ the mean and the standard deviation of the prediction vector $\left(\mathsf{M}(x_i,k)\right)_i$, namely:

$$\overline{\mathsf{M}}_k = \sum_{x\in\mathcal{X}} \tau_x \mathsf{M}(x,k) \quad \text{and} \quad \widehat{\sigma}_k^2 = \sum_{x\in\mathcal{X}} \tau_x \left(\mathsf{M}(x,k) - \overline{\mathsf{M}}_k\right)^2 \ .$$

Instead of focusing on ρ_k, we focus in the sequel on the following coefficient:

$$\dot{\rho}_k = \frac{1}{\widehat{\sigma}_k N} \sum_{i=1}^{N} \left(\mathsf{M}(x_i,k) - \overline{\mathsf{M}}_k\right) \mathrm{l}_i \ . \tag{3}$$

The distribution of $(\dot{\rho}_k)_{k\in\mathcal{K}}$ is indeed more convenient to analyze than the one of $(\rho_k)_{k\in\mathcal{K}}$. Moreover, one can verify that the ratio $\dot{\rho}_k/\rho_k$ equals the standard deviation of the leakage measurement vector $\mathbf{l}$. Consequently, $\dot{\rho}_k/\rho_k$ is positive and constant with respect to the key guess k. As a result, argmax-$o_{k\in\mathcal{K}}\ \rho_k =$ argmax-$o_{k\in\mathcal{K}}\ \dot{\rho}_k$ holds for every k and hence, the success rate of the attack is fully determined by the distribution of the vector $(\dot{\rho}_k)_{k\in\mathcal{K}}$. The next proposition provides us with the exact distribution of this vector.

Proposition 1. *The vector $(\dot{\rho}_k)_{k\in\mathcal{K}}$ has a multivariate Gaussian distribution whose expectation satisfies for every $k \in \mathcal{K}$:*

$$\mathrm{E}\left[\dot{\rho}_k\right] = \frac{1}{\widehat{\sigma}_k} \sum_{x\in\mathcal{X}} \tau_x \left(\mathsf{M}(x,k) - \overline{\mathsf{M}}_k\right) m_{x,k^*} \ , \tag{4}$$

and whose covariance satisfies for every $(k_1,k_2) \in \mathcal{K}^2$:

$$\mathrm{Cov}\left[\dot{\rho}_{k_1}, \dot{\rho}_{k_2}\right] = \frac{1}{N\widehat{\sigma}_{k_1}\widehat{\sigma}_{k_2}} \sum_{x\in\mathcal{X}} \tau_x \left(\mathsf{M}(x,k_1) - \overline{\mathsf{M}}_{k_1}\right)\left(\mathsf{M}(x,k_2) - \overline{\mathsf{M}}_{k_2}\right) \sigma_{x,k^*}^2 \ . \tag{5}$$

Proof. Since the l_i's are drawn from Gaussian distributions $\mathcal{N}\left(m_{x_i,k^*}, \sigma_{x_i,k^*}\right)$ and since the vector $(\dot{\rho}_k)_{k\in\mathcal{K}}$ is a linear transformation of $\mathbf{l}$, one deduces that $(\dot{\rho}_k)_{k\in\mathcal{K}}$ has a multivariate Gaussian distribution.

Now, for every $x \in \mathcal{X}$, we have $N\tau_x$ elements among the x_i's that are equal to x. This, together with (3) immediately leads to (4). Then, the mutual independence of the l_i's and the bilinearity of the covariance imply (5). ◇

Proposition 1 gives the exact distribution of the distinguishing vector $(\dot{\rho}_k)_{k\in\mathcal{K}}$. This makes it possible to precisely compute the success rate of a differential SCA that involves the Pearson correlation coefficient (see Section 6).

If the model is sound, namely if $\mathsf{M}(x,k)$ is linearly related to $m_{x,k}$, then (4) implies that the expectation of $\dot{\rho}_k$ is maximal for the good key guess $k = k^*$ which shows the soundness of the attack.

From (4) we see that the distinguishing vector expectation does not depend on the leakage variance nor on the number of leakage measurements. Conversely, (5) shows that the covariance matrix depends on these parameters. If the leakage variance is multiplied by a factor λ then so does the covariance matrix. And if the number of measurements is multiplied by a factor λ then the covariance matrix is multiplied by $1/\lambda$. As a result, if the leakage variance is increased by a given factor, the number of leakage measurements must also be increased by the same factor to keep unchanged the distinguisher distribution and hence the attack success rate.

Another interesting observation is that the distribution of $(\dot{\rho}_k)_{k\in\mathcal{K}}$ does not fully depend on the inputs vector $\mathbf{x}$ but only on the different ratios τ_x's. A usual choice, for a chosen plaintext differential SCA, is to set these ratios at $\tau_x = 1/|\mathcal{X}|$. For a known plaintext/ciphertext differential SCA, assuming that the x_i's are uniformly drawn, we further have $\tau_x \approx 1/|\mathcal{X}|$ for N large enough. We investigate this setting hereafter.

Uniform Setting. We investigate here the setting where the x_i's are chosen such that $\tau_x = 1/|\mathcal{X}|$ holds for every x. We further assume that the target signal S can be expressed as $S = \psi(X \oplus K)$ where ψ is a balanced function (*i.e.* the cardinal of $\psi^{-1}(s)$ is constant for every $s \in \mathcal{S}$).

In the uniform setting, the previous study can be simplified. In this setting, the mean and the standard deviation of the prediction vector are constant with respect to k^*. Indeed, for every $k \in \mathcal{K}$, we have $\overline{\mathsf{M}}_k = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} \mathsf{M}(s)$ and $\widehat{\sigma}_k = \sqrt{\frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\left(\mathsf{M}(s) - \overline{\mathsf{M}}\right)^2}$. Hence, we can focus on the following coefficient:

$$\ddot{\rho}_k = \frac{1}{N}\sum_{i=1}^{N} \mathsf{M}(x_i, k)\mathrm{l}_i \ . \tag{6}$$

Once again $\ddot{\rho}_k/\rho_k$ is positive and constant with respect to k which implies that focusing on ρ_k instead of $\ddot{\rho}_k$ does not affect the success rate of the attack. The following corollary gives the distribution of $(\ddot{\rho}_k)_{k\in\mathcal{K}}$.

Corollary 1. *The vector $(\ddot{\rho}_k)_{k\in\mathcal{K}}$ has a multivariate Gaussian distribution whose expectation satisfies for every $k \in \mathcal{K}$:*

$$\mathrm{E}\left[\ddot{\rho}_k\right] = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} \mathsf{M}(x, k) m_{x,k^*} \ , \tag{7}$$

and whose covariance satisfies for every $(k_1, k_2) \in \mathcal{K}^2$:

$$\mathrm{Cov}\left[\ddot{\rho}_{k_1}, \ddot{\rho}_{k_2}\right] = \frac{1}{N|\mathcal{X}|}\sum_{x\in\mathcal{X}} \mathsf{M}(x, k_1)\mathsf{M}(x, k_2)\sigma^2_{x,k^*} \ . \tag{8}$$

Proof. Corollary 1 straightforwardly holds from Proposition 1 by setting $\overline{\mathsf{M}}_k$ to 0 and $\widehat{\sigma}_k$ to 1. ◇

An interesting property of the uniform setting is stressed in the following proposition.

Proposition 2. *Let $(d_k)_{k\in\mathcal{K}}$ and $(d'_k)_{k\in\mathcal{K}}$ be the distributions of the vector $(\ddot{\rho}_k)_{k\in\mathcal{K}}$ for two secret keys $k_1^* \in \mathcal{K}$ and $k_2^* \in \mathcal{K}$ respectively. In the uniform setting, the distributions $(d_{k\oplus k_1^*})_{k\in\mathcal{K}}$ and $(d'_{k\oplus k_2^*})_{k\in\mathcal{K}}$ are indentical.*

Proof. In the uniform setting, we have $\mathsf{M}(x,k) = \mathsf{M}(\psi(x \oplus k))$ and $m_{x,k^*} = m_{\psi(x\oplus k^*)}$. Hence, from (7) we get $\mathrm{E}\left[d_{k\oplus k_1^*}\right] = \mathrm{E}[d'_{k\oplus k_2^*}]$ for every $k \in \mathcal{K}$ and from (8) we get $\mathrm{Cov}\left[d_{k_1\oplus k_1^*}, d_{k_2\oplus k_1^*}\right] = \mathrm{Cov}[d'_{k_1\oplus k_2^*}, d'_{k_2\oplus k_2^*}]$ for every $(k_1,k_2) \in \mathcal{K}^2$. Finally, since $(d_k)_{k\in\mathcal{K}}$ and $(d'_k)_{k\in\mathcal{K}}$ are both Gaussian then they are identical. $\diamond$

Proposition 2 shows that the vector $(\ddot{\rho}_{k\oplus k^*})_{k\in\mathcal{K}}$ has the same distribution for every k^*. Moreover, the event $k^* \in \text{argmax-}o_{k\in\mathcal{K}}\ \ddot{\rho}_k$ can be rewritten as $0 \in \text{argmax-}o_{k\in\mathcal{K}}\ \ddot{\rho}_{k\oplus k^*}$. Since the distribution of $(\ddot{\rho}_{k\oplus k^*})_{k\in\mathcal{K}}$ is independent of k^*, we get that, in the uniform setting, the success rate is constant with respect to k^*. Therefore, one only needs to analyze the distribution of $(\ddot{\rho}_{k\oplus k^*})_{k\in\mathcal{K}}$ for a given secret key (*e.g.* for $k^* = 0$) to get the distribution and the success rate of $(\ddot{\rho}_k)_{k\in\mathcal{K}}$ for any secret key k^*.

5 Profiling Side Channel Analysis

5.1 Description

Profiling Side Channel Analysis assumes an adversary that owns a *profile* of the side channel leakage (also called *template* in the literature from the initial work of Chari *et al.* [6]). More precisely, the adversary owns an estimation of the pdf $\mathrm{l} \mapsto \mathrm{P}\left[\mathrm{L} = \mathrm{l} | S = s\right]$ for every $s \in \mathcal{S}$. In practice, this estimation is obtained through a *profiling phase* on a physical implementation identical to the targeted one (except the secret key) and that is under the attacker control (see for instance [2,6,14,19]).

The attack consists in estimating the likelihood of a key guess k, *i.e.* the probability that K is equal to k, given the leakage measurements vector $\mathbf{l}$ and the inputs vector $\mathbf{x}$. Assuming that K is uniformly distributed (which is very usual in cryptography), it can be checked that this probability satisfies:

$$\mathrm{P}\left[K = k | (\mathbf{l}, \mathbf{x})\right] = \alpha \prod_{i=1}^{N} \mathrm{P}\left[\mathrm{L} = \mathrm{l}_i | S = \varphi(x_i, k)\right] \ , \tag{9}$$

where α denotes a value constant with respect to k.

In the Gaussian model, the leakage pdf $\mathrm{l} \mapsto \mathrm{P}\left[\mathrm{L} = \mathrm{l} | S = s\right]$ is the Gaussian pdf ϕ_{Σ_s, m_s}. Estimating such a pdf amounts to estimate the parameters (m_s, Σ_s) for every $s \in \mathcal{S}$. In the sequel, we shall denote these estimations by $\widehat{m}_s$ and $\widehat{\Sigma}_s$. For clarity and without ambiguity, the parameters $\widehat{m}_{\varphi(x,k)}$ and $\widehat{\Sigma}_{\varphi(x,k)}$ are further denoted by $\widehat{m}_{x,k}$ and by $\widehat{\Sigma}_{x,k}$.

For computational reasons, one usually processes the logarithm of the estimated likelihood and averages it on the number of leakage measurements. Moreover, since α is constant with respect to k, it is usually ignored. On the whole, one computes the *log-likelihood* $\mathcal{L}_k = \frac{1}{N}\log(\mathrm{P}\left[K = k|(\mathbf{l}, \mathbf{x})\right]/\alpha)$. In the Gaussian model, $\mathcal{L}_k$ satisfies:

$$\mathcal{L}_k = \frac{1}{2N}\sum_{i=1}^{N}\Big(\log\big((2\pi)^T|\widehat{\Sigma}_{x_i,k}|\big) - (\mathrm{l}_i - \widehat{m}_{x_i,k})'\,\widehat{\Sigma}_{x_i,k}^{-1}\,(\mathrm{l}_i - \widehat{m}_{x_i,k})\Big)\ . \qquad (10)$$

In the next section, we investigate the distribution of the log-lekelihood distinguisher under the Gaussian model assumption.

5.2 Distinguisher Distribution

Let us first introduce few notations. The element of the i^{th} row and of the j^{th} column of a matrix A is denoted by $A[i,j]$ while the i^{th} element of a vector V is denoted by $V[i]$. A' denotes the transpose of a matrix (or a vector) A. The notation $\|\cdot\|$ is used to denote the Euclidian norm while the notation $\|\cdot\|_{hs}$ refers to the Hilbert-Schmidt matrix norm defined by $\|A\|_{hs} = \sqrt{\sum_{i,j} A[i,j]^2}$. We shall further denote by A^2 the product $A'A$ and by $A^{-1/2}$ any matrix satisfying $(A^{-1/2})'A^{-1/2} = A$ (*e.g.* the Cholesky decomposition matrix). Finally the trace of A is denoted by $\mathrm{Tr}(A)$.

The next proposition provides a precise approximation of the distribution of the likelihood vector $(\mathcal{L}_k)_{k\in\mathcal{K}}$ (the proof is given in Appendix B).

Proposition 3. *The distribution of the vector $(\mathcal{L}_k)_{k\in\mathcal{K}}$ tends toward a multivariate Gaussian distribution as N grows. Moreover, for every $k \in \mathcal{K}$, the expectation of $\mathcal{L}_k$ satisfies:*

$$\mathrm{E}\left[\mathcal{L}_k\right] = \frac{1}{2}\sum_{x\in\mathcal{X}}\tau_x\Big(\log\big(2\pi|\widehat{\Sigma}_{x,k}|\big) - \Big\|\widehat{\Sigma}_{x,k}^{-1/2}\,(m_{x,k^*} - \widehat{m}_{x,k})\Big\|^2 - \mathrm{Tr}\Big(\widehat{\Sigma}_{x,k}^{-1/2}\,\Sigma_{x,k^*}\,(\widehat{\Sigma}_{x,k}^{-1/2})'\Big)\Big)\ , \qquad (11)$$

and for every $(k_1, k_2) \in \mathcal{K}^2$, the covariance between $\mathcal{L}_{k_1}$ and $\mathcal{L}_{k_2}$ satisfies:

$$\mathrm{Cov}\left[\mathcal{L}_{k_1}, \mathcal{L}_{k_2}\right] = \frac{1}{N}\sum_{x\in\mathcal{X}}\tau_x\Big(\ \frac{1}{2}\Big\|\widehat{\Sigma}_{x,k_1}^{-1/2}\,\Sigma_{x,k^*}\,(\widehat{\Sigma}_{x,k_2}^{-1/2})'\Big\|_{hs}^2 + \ (m_{x,k^*} - \widehat{m}_{x,k_1})'\,\widehat{\Sigma}_{x,k_1}^{-1}\,\Sigma_{x,k^*}\,\widehat{\Sigma}_{x,k_2}^{-1}\,(m_{x,k^*} - \widehat{m}_{x,k_2})\ \Big)\ . \qquad (12)$$

Proposition 3 gives an approximation of the distribution of the log-likelihood vector $(\mathcal{L}_k)_{k\in\mathcal{K}}$ which becomes quickly tight as N grows (see Appendix C). As shown in Section 6, this approximation can be used to estimate the success rate of profiling SCA. The computational cost of (11) and (12) is $O(|\mathcal{X}|T^3)$ where T denotes the leakage dimension. The total cost of computing the distribution parameters is hence $O(|\mathcal{K}|^2|\mathcal{X}|T^3)$. This may be prohibitive if the leakage dimension is high. However, the leakage dimension can be reduced by pre-processing the leakage measurements [2, 20]. In practice, $T = 3$ is often sufficient to catch most of the side channel information [2, 20].

In order to simplify our analysis, let us make the following assumption.

Assumption 1 (Constant Covariance Assumption) *The covariance matrix Σ_s is the same for all signals $s \in \mathcal{S}$.*

Remark 3. This assumption is quite usual in the literature (see for instance [19, 11, 21]). The noise in the leakage is indeed often independent of the target signal. This is especially true if most of the noise amount is produced by a noise generator (independent of the target algorithm) as a countermeasure to side channel analysis.

Observing the expectation of $\mathcal{L}_k$ (11), one identifies three sums. The first one and the third one only involve the leakage covariance matrices and/or their estimations. Therefore, under the constant covariance assumption, these sums are constant with respect to k and hence, they provide no discrimination between the different key candidates. Actually, only the second sum in (11) provides some discrimination which depends on the leakage means $m_{x,k}$ (corresponding to the different processed signals $s = \varphi(x, k)$). If these means are clearly dissociated and if their estimations $\widehat{m}_{x,k}$ are precise, then the second sum is around zero for and only for the good key guess k^*. As a result, the expectation of $\mathcal{L}_k$ is maximized for the good key guess $k = k^*$ which illustrates the attack soundness.

From (11) and (12) we also see that, unlike for differential SCA, increasing the number of leakage measurements and increasing the leakage variance do not have a complementary effect on the distinguisher distribution. However, it has a complementary effect on the success rate: if the leakage covariance matrix is multiplied by a factor λ (and assuming that its estimation is also multiplied by λ) then the attacker must multiply the number of measurements by a factor λ in order to keep the success rate constant. This fact is formally demonstrated in Appendix D. We hence remark (according to the analysis in Section 4.2) that Differential SCA and Profiling SCA are affected in the same way by the leakage noise increase. Besides, when the leakage noise is amplified, the ratio between the efficiencies[1] of both attacks remains constant.

As final remark, let us mention that Proposition 2 also applies to the log-likelihood vector $(\mathcal{L}_k)_{k\in\mathcal{K}}$. Besides, in the uniform setting (see Section 4.2), the success rate of the profiling SCA is also constant with respect to k^*.

6 Success Rate Evaluation

In accordance with the analyses of Sections 4.2 and 5.2, we assume that the distribution of the distinguishing vector $\mathbf{d} = (d_k)_{k\in\mathcal{K}}$ is a multivariate Gaussian $\mathcal{N}(m_{\mathbf{d}}, \Sigma_{\mathbf{d}})$. In this section we present two approaches to compute the success rate of a side channel key recovery attack, once the parameter of this distribution have been determined.

In the first approach, we show that the success rate can be expressed as a sum of Gaussian cumulative distribution functions (cdf). It can hence be estimated by

[1] By efficiency, we mean the required number of leakage measurements to succeed the attack (with high probability).

numerically computing these cdf. The second approach consists in simulating the multivariate Gaussian vector $\mathbf{d}$ several times in order to get a precise estimation of the success rate.

6.1 Numerical Computation

We show hereafter that the success rate can be expressed as a sum of Gaussian cdf. For this purpose, we need to introduce the *comparison vector* that is the $(|\mathcal{K}|-1)$-size vector $\mathbf{c} = (c_k)_{k\in\mathcal{K}/\{k^*\}}$ defined for every $k \in \mathcal{K}/\{k^*\}$ by:

$$c_k = d_{k^*} - d_k \ . \tag{13}$$

If all the coordinates of this vector are positive then the attack succeeds in isolating the good key guess as first candidate. If n coordinates are negative then the attack rates the good key guess as the $(n+1)^{\text{th}}$ candidate; in other words, it succeeds at the $(n+1)^{\text{th}}$ order. The comparison vector is a linear transformation of the distinguishing vector by a $((|\mathcal{K}|-1)\times|\mathcal{K}|)$-matrix P whose expression straightforwardly follows from (13). This implies that the comparison vector has a multivariate Gaussian distribution $\mathcal{N}(m_{\mathbf{c}}, \varSigma_{\mathbf{c}})$ where $m_{\mathbf{c}} = Pm_{\mathbf{d}}$ and $\varSigma_{\mathbf{c}} = P\varSigma_{\mathbf{d}}P'$.

Let $\alpha \subseteq \{1,\cdots,|\mathcal{K}|-1\}$ be a set of indices and let I_α and S_α be the $(|\mathcal{K}|-1)$-size vectors defined by:

$$I_\alpha[i] = \begin{cases} -\infty & \text{if } i \in \alpha \\ 0 & \text{if } i \notin \alpha \end{cases} \quad \text{and} \quad S_\alpha[i] = \begin{cases} 0 & \text{if } i \in \alpha \\ +\infty & \text{if } i \notin \alpha \end{cases} \ .$$

The vector $\mathbf{c}$ has exactly n negative coordinates if and only if there exists a set α of cardinal n s.t. $I_\alpha < \mathbf{c} < S_\alpha$. Since the intervals $([I_\alpha, S_\alpha])_\alpha$ are disjoints, the probability that exactly n coordinates of $\mathbf{c}$ be negative can be written as:

$$p_n = \sum_{\alpha;|\alpha|=n} \mathrm{P}\left[I_\alpha \le \mathbf{c} \le S_\alpha\right] \ . \tag{14}$$

The o$^{\text{th}}$ order success rate equals the sum $p_0 + p_1 + \cdots + p_{o-1}$ which from (14) gives:

$$\text{Succ-}o = \sum_{\alpha;|\alpha|<o} \mathrm{P}\left[I_\alpha \le \mathbf{c} \le S_\alpha\right] = \sum_{\alpha;|\alpha|<o} \varPhi_{m_{\mathbf{c}},\varSigma_{\mathbf{c}}}(I_\alpha, S_\alpha) \ , \tag{15}$$

where $\varPhi_{m,\varSigma}$ denotes the Gaussian cdf that satisfies $\varPhi_{m,\varSigma} : (a,b) \mapsto \int_a^b \phi_{m,\varSigma}(x)\,dx$.

Relation (15) shows that the o^{th} order success rate can be computed by performing $\sum_{i<o}\binom{|\mathcal{K}|-1}{i}$ multivariate Gaussian cdf calculations (on $(|\mathcal{K}|-1)$-size Gaussian vectors). The numerical computation of multivariate Gaussian cdf is a classical issue in statistics. Some solutions exist (see for instance [9, 10]) that can be used to precisely compute the success rate according to (15).

This approach has some drawbacks. Firstly, the numerical computations of Gaussian cdf may be difficult with covariance matrices having particular forms and/or quite high dimensions. For instance it requires that the covariance matrix is not singular, which is not always the case in our context. Yet another drawback of this approach is that the computation of high order success rates requires an important number of Gaussian cdf computations. Regarding these issues, a possible alternative is presented in the next section.

6.2 Gaussian simulation

Another possibility to compute the success rate is to perform a Gaussian simulation. The principle is to simulate several times the distribution $\mathcal{N}(m_{\mathbf{d}}, \Sigma_{\mathbf{d}})$. This amounts to randomly pick up several distinguishing vectors each one corresponding to the result of an attack. The success rate is estimated based on these different results. In other words this approach works as an attack simulation but instead of performing the attack several times, we perform several Gaussian random vectors simulation which is clearly more efficient especially when the number of leakage measurements is high and/or the leakage dimension is high. Another advantage of this approach is that the success rate at the different orders as well as the guessing entropy (see Appendix A) can all be computed using the same simulated distinguishing vectors. Finally Gaussian simulation is sound even when the covariance matrix is singular which may happen in our context.

7 Empirical Validation

In order to empirically validate the theoretical analyses conducted in the previous sections we performed several simulations. We chose $S = X \oplus K$ as target signal where X and K are 8 bits variables. The leakage means $(m_s)_{s \in \mathcal{S}}$ and the leakage covariance matrix Σ (assumed constant for the different signals $s \in \mathcal{S}$) were drawn with random coefficients. Their dimensions were set to 1 for differential SCA, and to 3 for profiling SCA (this is a typical dimension when subspace-based profiling is involved [2, 20]). The attacker model/estimations were first assumed to be exact (*i.e.* $\mathsf{M}(s) = m_s$, $\widehat{m}_s = m_s$ and $\widehat{\Sigma} = \Sigma$) and then assumed to be slightly erroneous (by inserting random errors).

On the one hand, the success rate was estimated empirically by simulating the attack. Namely, the leakage measurements l_i corresponding to random inputs x_i were randomly picked up according to the leakage parameters $(m_{x_i,k^*}, \Sigma_{x_i,k^*})$. The attack was performed several times (few thousands) on such simulated measurements in order to obtain an empirical success rate. On the other hand, the success rate was estimated using our approach. We computed the distinguishing vector expectation and covariance matrix (such as described in Sections 4.2 and 5.2) according to the leakage parameters and assuming $\tau_x = 1/256$ for every x. Then we performed Gaussian simulations (see Section 6.2) to get an estimation of the success rate.

As expected, for differential SCA, the different success rates obtained with our approach always match almost perfectly the success rates obtained by attack simulations. For profiling SCA, the success rates obtained with our approach also match quite well the success rates obtained by attack simulations. The precision

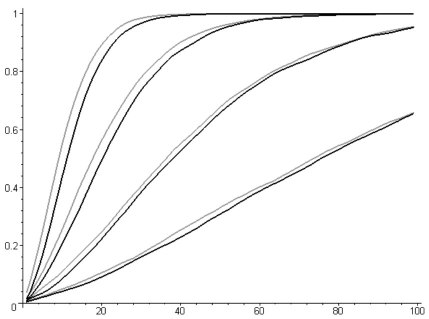

Fig. 1. Success rates of different profiling SCA attacks over an increasing number of leakage measurements.

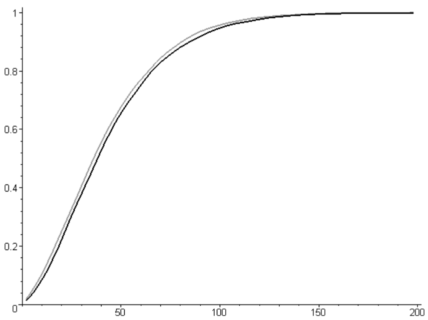

Fig. 2. Success rates of a profiling SCA attack over an increasing number of leakage measurements.

of this matching depends on the number of leakage measurements required for the attack to succeed (with a high probability). When this number is quite low (*i.e.* around few hundreds), our estimation slightly overvalues the real success rate. This overvaluation becomes less marked as the number of required leakage measurements increases. As an illustration the success rate of four attacks requiring different amounts of leakage measurements is plotted Figure 1. Success rates obtained by attack simulation are plotted in black while the corresponding ones obtained with our approach are plotted in grey. The convergence can be clearly observed. Figure 2 shows both success rates for an attack requiring around 200 leakage measurements. When moving up to 500 required leakage measurements, the curves completely mix up.

The different empirical results that we obtained have demonstrated the soundness of our theoretical analysis. They also show that the approximation $\tau_x \approx 1/|\mathcal{X}|$ is sound when the x_i's are randomly drawn (*i.e.* in a known plaintext/ciphertext attack setting).

8 Conclusion and Open Issues

In this paper, we have investigated the issue of evaluating the success rate of side channel analysis in the Gaussian leakage model. For the two main families of SCA, namely differential SCA and profiling SCA, we have shown that the distinguishing vector resulting from the attack has (or at least quickly tends towards) a multivariate Gaussian distribution. This allowed us to exhibit an efficient way to compute the success rate of such an attack according to the number of leakage measurements and to the leakage distribution parameters. Finally, our analysis was validated empirically by a large number of attack simulations.

Our analysis stresses several interesting open issues. Future works could focus on chosen plaintext attacks and investigate how the choice of the target inputs may affect the success rate of an attack. Another interesting issue is the tolerance for a distinguisher to the error on the leakage model. How does an error on the attacker model/estimations affect the success rate of the attack ? Finally, extension of our analysis to protected implementations (for instance by masking techniques) would be of great interest to quantify their security.

Acknowledgements

I would like to thank Emmanuel Prouff for his reviews and for his suggestions that substantially improved the paper. I am also grateful to Jean-Sébastien Coron, Emmanuelle Dottax and Christophe Giraud for their reviews and helpful comments.

References

1. M.-L. Akkar, R. Bévan, P. Dischamp, and D. Moyart. Power Analysis, What is Now Possible. In T. Okamoto, editor, *ASIACRYPT 2000*, volume 1976 of *LNCS*, pages 489–502. Springer, 2000.
2. C. Archambeau, E. Peeters, F.-X. Standaert, and J.-J. Quisquater. Template Attacks in Principal Subspaces. In L. Goubin and M. Matsui, editors. *CHES 2006*, volume 4249 of *LNCS*, pages 1–14. Springer, 2006.
3. R. Bévan and E. Knudsen. Ways to Enhance Power Analysis. In P. Lee and C. Lim, editors, *ICISC 2002*, volume 2587 of *LNCS*, pages 327–342. Springer, 2002.
4. E. Brier, C. Clavier, and F. Olivier. Correlation Power Analysis with a Leakage Model. In M. Joye and J.-J. Quisquater, editors, *CHES 2004*, volume 3156 of *LNCS*, pages 16–29. Springer, 2004.
5. C. Cachin. *Entropy Measures and Unconditional Security in Cryptography.* PhD thesis, 1997.
6. S. Chari, J. Rao, and P. Rohatgi. Template Attacks. In B. Kaliski Jr., Ç. Koç, and C. Paar, editors, *CHES 2002*, volume 2523 of *LNCS*, pages 13–29. Springer, 2002.
7. C. Clavier, J.-S. Coron, and N. Dabbous. Differential Power Analysis in the Presence of Hardware Countermeasures. In Ç. Koç and C. Paar, editors, *CHES 2000*, volume 1965 of *LNCS*, pages 252–263. Springer, 2000.
8. K. Gandolfi, C. Mourtel, and F. Olivier. Electromagnetic Analysis: Concrete Results. In Ç. Koç, D. Naccache, and C. Paar, editors, *CHES 2001*, volume 2162 of *LNCS*, pages 251–261. Springer, 2001.
9. A. Genz. Numerical Computation of Multivariate Normal Probabilities. *Journal of Computational and Graphical Statistics*, 1:141–149, 1992.
10. A. Genz. Comparison of Methods for the Computation of Multivariate Normal Probabilities. *Computing Science and Statistics*, 25:400–405, 1993.
11. B. Gierlichs, K. Lemke-Rust, and C. Paar. Templates vs. Stochastic Methods. In L. Goubin and M. Matsui, editors. *CHES 2006*, volume 4249 of *LNCS*, pages 15–29. Springer, 2006.
12. P. Kocher, J. Jaffe, and B. Jun. Differential Power Analysis. In M. Wiener, editor, *CRYPTO '99*, volume 1666 of *LNCS*, pages 388–397. Springer, 1999.
13. P. Kocher. Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. In N. Koblitz, editor, *CRYPTO '96*, volume 1109 of *LNCS*, pages 104–113. Springer, 1996.

14. K. Lemke-Rust and C. Paar. Gaussian Mixture Models for Higher-Order Side Channel Analysis. In P. Paillier and I. Verbauwhede, editors, *CHES 2007*, volume 4727 of *LNCS*, pages 14–27. Springer, 2007.
15. S. Mangard, E. Oswald, and T. Popp. *Power Analysis Attacks – Revealing the Secrets of Smartcards.* Springer, 2007.
16. S. Mangard. Hardware Countermeasures against DPA – A Statistical Analysis of Their Effectiveness. In T. Okamoto, editor, *CT-RSA 2004*, volume 2964 of *LNCS*, pages 222–235. Springer, 2004.
17. J. Massey. Guessing and Entropy. *IEEE ISIT*, page 204, 1994.
18. T. Messerges, E. Dabbish, and R. Sloan. Investigations of Power Analysis Attacks on Smartcards. In *the USENIX Workshop on Smartcard Technology (Smartcard '99)*, pages 151–161, 1999.
19. W. Schindler, K. Lemke, and C. Paar. A Stochastic Model for Differential Side Channel Cryptanalysis. In J. Rao and B. Sunar, editors, *CHES 2005*, volume 3659 of *LNCS*. Springer, 2005.
20. F.-X. Standaert, and C. Archambeau. Using Subspace-Based Template Attacks to Compare and Combine Power and Electromagnetic Information Leakages. To Appear in *CHES 2008*.
21. F.-X. Standaert, T. G. Malkin, and M. Yung. A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. Cryptology ePrint Archive, Report 2006/139, 2006. http://eprint.iacr.org/.
22. F.-X. Standaert, E. Peeters, G. Rouvroy, and J.-J. Quisquater. An Overview of Power Analysis Attacks against Field Programmable Gate Arrays. *IEEE*, 94(2):383–394, 2006.

A Guessing Entropy

The guessing entropy [17, 5] is defined as the expected number of key guesses to test before recovering a target key value. As pointed out in [21], the guessing entropy is relevant in the context of side channel analysis since it indicates the average workload to perform after side channel analysis. Let $\mathrm{rank}_k(\mathbf{d})$ denote the index $i \in \{1, \cdots, |\mathcal{K}|\}$ such that d_k is the i^{th} higher element of $\mathbf{d}$. The guessing entropy of a side channel attack using a distinguisher D and a public vector $\mathbf{x}$, and targeting a secret key k^* is formally defined as:

$$\mathrm{GE}^{\mathsf{D}}_{\mathbf{x},k^*} = \mathrm{E}\left[\left(\mathrm{l}_i \leftarrow \mathrm{L}\left(\varphi(k^*, x_i)\right)\right)_i \; ; \; \mathbf{d} \leftarrow \mathsf{D}(\mathbf{x}, \mathbf{l}) \; : \; \mathrm{rank}_{k^*}(\mathbf{d})\right] \; . \tag{16}$$

The guessing entropy is related to the success rate of every order. In fact, the correct key guess is rated at the o^{th} rank in the distinguishing vector if and only it is rated among the o first candidates but it is not rated among the $o-1$ first candidates. As a result, the probability that the correct key guess be rated at the o^{th} rank satisfies for every o: $\mathrm{P}\left[\mathrm{rank}_{k^*}(\mathbf{d}) = o\right] = \text{Succ-}o - \text{Succ-}(o-1)$, where Succ-0 is naturally defined at zero. This brings to the following relation:

$$\mathrm{GE} \;=\; \sum_{o=1}^{|\mathcal{K}|} o \; \mathrm{P}\left[\mathrm{rank}_{k^*}(\mathbf{d}) = o\right] \;=\; |\mathcal{K}| - \sum_{o=1}^{|\mathcal{K}|-1} \text{Succ-}o \; . \tag{17}$$

B Proof of Proposition 3

The proof of Proposition 3 makes use of the following lemma.

Lemma 1. *Let X be a T-size random vector having a Gaussian distribution $\mathcal{N}(0, \Sigma)$. Let A_1 and A_2 be two $(T \times T)$-matrices and let m_1 and m_2 be two T-size vectors. Let Q_1 and Q_2 be two quadratic forms defined, for $j = 1, 2$, by $Q_j = (X + m_j)' A_j^2 (X + m_j)$. For $j = 1, 2$, the expectation of Q_j satisfies:*

$$\mathrm{E}[Q_j] = \|A_j\, m_j\|^2 + \mathrm{Tr}(A_j\, \Sigma\, A_j') \ . \tag{18}$$

And the covariance of Q_1 and Q_2 satisfies:

$$\mathrm{Cov}[Q_1, Q_2] = 2\, \|A_1\, \Sigma\, A_2'\|_{hs}^2 + 4\ m_1'\, A_1^2\, \Sigma\, A_2^2\, m_2 \ . \tag{19}$$

Proof. We have $Q_j = \sum_{i=1}^{T} (A_j\, (X + m_j))[i]^2$ which leads to:

$$\begin{aligned} \mathrm{E}[Q_j] &= \sum_{i=1}^{T} \mathrm{E}\left[(A_j\, (X + m_j))[i]^2\right] && (20) \\ &= \sum_{i=1}^{T} \mathrm{E}[(A_j\, (X + m_j))[i]]^2 + \sum_{i=1}^{T} \mathrm{Var}[(A_j\, (X + m_j))[i]] \ , && (21) \end{aligned}$$

since $\mathrm{E}\left[Y^2\right] = \mathrm{Var}[Y] + \mathrm{E}[Y]^2$ holds for every random variable Y. From $(X + m_j) \sim \mathcal{N}(m_j, \Sigma)$ we have $A_j\, (X + m_j) \sim \mathcal{N}\left(A_j\, m_j, A_j\, \Sigma\, A_j'\right)$ which directly yields (18).

The quadratic form Q_j can be rewritten as $Q_j = (A_j\, X)^2 + (A_j\, m_j)^2 + 2 m_j'\, A_j^2\, X$ for $j = 1, 2$. By bilinearity, $\mathrm{Cov}[Q_1, Q_2]$ satisfies:

$$\begin{aligned} \mathrm{Cov}[Q_1, Q_2] = \mathrm{Cov}&\left[(A_1\, X)^2,\ (A_2\, X)^2\right] \\ &+ 2\ \mathrm{Cov}\left[(A_1\, X)^2,\ m_2'\, A_2\, X\right] + 2\ \mathrm{Cov}\left[(A_2\, X)^2,\ m_1'\, A_1\, X\right] \\ &+ 4\ \mathrm{Cov}[m_1'\, A\, X,\ m_2'\, A\, X] \ . \end{aligned} \tag{22}$$

We claim the three following relations:

$$\mathrm{Cov}\left[(A_1\, X)^2, (A_2\, X)^2\right] = 2\, \|A_1\, \Sigma\, A_2'\|_{hs}^2 \ , \tag{23}$$

$$\mathrm{Cov}\left[(A_1\, X)^2, m_2'\, A_2^2\, X\right] = \mathrm{Cov}\left[(A_2\, X)^2, m_1'\, A_1^2\, X\right] = 0 \ , \tag{24}$$

$$\mathrm{Cov}\left[m_1'\, A_1^2\, X,\ m_2'\, A_2^2\, X\right] = m_1'\, A_1^2\, \Sigma\, A_2^2\, m_2 \ . \tag{25}$$

These relations together with (22) result in (19) and state the correctness of Lemma 1. Relation (25) straightforwardly holds from the bilinearity of the covariance and by symmetry of A_1^2 (*i.e.* $(A_1^2)' = A_1^2$). Relations (23) and (24) are stated hereafter.

First, let us show (23). The covariance between $(A_1\,X)^2$ and $(A_2\,X)^2$ can be rewritten as:

$$\begin{aligned}\mathrm{Cov}\left[(A_1\,X)^2,(A_2\,X)^2\right] &= \sum_{i,j}\mathrm{Cov}\left[(A_1\,X)[i]^2,(A_2\,X)[j]^2\right]\\ &= \sum_{i,j}\left(\mathrm{E}\left[(A_1\,X)[i]^2(A_2\,X)[j]^2\right]-\mathrm{E}\left[(A_1\,X)[i]^2\right]\mathrm{E}\left[(A_2\,X)[j]^2\right]\right)\ . \qquad (26)\end{aligned}$$

Since the expectations of $A_1\,X$ and $A_2\,X$ equal zero, the expectation of the product $(A_1\,X)[i]^2(A_2\,X)[j]^2$ is the Gaussian forth order moment that is known to satisfy:

$$\begin{aligned}\mathrm{E}\left[(A_1\,X)[i]^2(A_2\,X)[j]^2\right] = \mathrm{E}\left[(A_1\,X)[i]^2\right]\mathrm{E}\left[(A_2\,X)[j]^2\right]\\ +2\ \mathrm{Cov}\left[(A_1\,X)[i],(A_2\,X)[j]\right]^2\ . \qquad (27)\end{aligned}$$

Hence, (26) gives:

$$\mathrm{Cov}\left[(A_1\,X)^2,(A_2\,X)^2\right] = 2\sum_{i,j}\mathrm{Cov}\left[(A_1\,X)[i],(A_2\,X)[j]\right]^2\ . \qquad (28)$$

Since we have $\mathrm{Cov}\left[(A_1\,X)[i],(A_2\,X)[j]\right] = (A_1\,\Sigma\,A_2')[i,j]$, one deduces that (28) finally results in (23).

We now show the correctness of (24). We have:

$$\mathrm{Cov}\left[(A_1\,X)^2,\ m_2'\,A_2^2\,X\right] = \sum_{i}\mathrm{Cov}\left[(A_1\,X)[i]^2,\ m_2'\,A_2^2\,X\right]\ . \qquad (29)$$

Since X has a zero mean, every term of the previous sum is a Gaussian third order moment and is hence equal to zero. This way, we get (24). ◇

We give hereafter the proof of Proposition 3.

Proof. (Proposition 3) Since the l_i's are independently drawn from Gaussian distributions $\mathcal{N}(m_{x_i,k^*},\Sigma_{x_i,k^*})$ and since, for every x, there is a ratio τ_x of the x_i's that equal x, Relation (10) and Lemma 1 directly lead to (11) and (12).

Now, $(\mathcal{L}_k)_{k\in\mathcal{K}}$ can be expressed as a linear transformation of the vector $\sum_{i=1}^N l_i$ and of the vector $\left(\sum_{i=1}^N l_i[j_1]l_i[j_2]\right)_{1\leq j_1,j_2\leq T}$. The first one has a multivariate Gaussian distribution and, from the multivariate central limit theorem, the second one tends toward a multivariate Gaussian distribution as N grows. Hence $(\mathcal{L}_k)_{k\in\mathcal{K}}$ tends toward a multivariate Gaussian distribution as N grows. ◇

C Convergence of the Log-Likelihood Distribution

According to (10), the log-likelihood $\mathcal{L}_k$ can be expressed as the sum of $|\mathcal{X}|$ values $\mathcal{L}_{k,x}$ that are defined by:

$$\mathcal{L}_{k,x} = \frac{\tau_x}{2} \log\left((2\pi)^T |\widehat{\Sigma}_{x,k}|\right) - \frac{1}{2N} \sum_{\substack{i=1 \\ x_i=x}}^{N} (\mathrm{l}_i - \widehat{m}_{x,k})' \, \widehat{\Sigma}_{x,k}^{-1} (\mathrm{l}_i - \widehat{m}_{x,k}) \; . \tag{30}$$

The first term is constant and the second term is a sum of $N\tau_x$ elements of the form $X' A^2 X$ where A is the matrix $\widehat{\Sigma}_{x,k}^{-1}$ and X is a Gaussian random variable $\mathcal{N}(m_{x,k^*} - \widehat{m}_{x,k}, \Sigma_{x,k^*})$. The distribution of such a sum is given in the following lemma. At first, let us recall that the chi-square distribution with n degrees of freedom $\chi^2(n)$ is the distribution obtained by summing n independent $\mathcal{N}(0,1)$-distributed random variables.

Lemma 2. *Let $(X_j)_j$ be n independent T-size random vectors having a Gaussian distribution $\mathcal{N}(m, \Sigma)$, let A be a $(T \times T)$-matrix and let $(Q_j)_j$ be the quadratic forms defined as $Q_j = X_j' A^2 X_j$. The sum of the Q_j satisfies:*

$$\sum_{j=1}^{n} Q_j = \beta + G + \sum_{i=1}^{T} \alpha_i C_i \; , \tag{31}$$

where $\beta = n(A \cdot m)^2$, $\alpha_i = (A \Sigma A')[i,i]$, G is an univariate Gaussian random variable, C_i are T chi-square random variables with n degrees of freedom.

Proof. For $j = 1, 2$, we have $Q_j = (A X_j)^2$. Denoting by $\overline{X}_j$ the centered random variable $X_j - m$, we get $Q_j = (A m)^2 + 2 \, m' A^2 \overline{X}_j + (A \overline{X}_j)^2$ and hence, $\sum_j Q_j = \beta + 2 \sum_j m' A^2 \overline{X}_j + \sum_j \sum_i (A \overline{X}_j)[i]^2$.

After denoting $2 \sum_j m' A^2 \overline{X}_j$ by G and $\frac{1}{\alpha_i} \sum_j (A \overline{X}_j)[i]^2$ by C_i, we get (31). Now, G is Gaussian since it is defined as a sum of Gaussian random variables. Moreover, the covariance matrix of $A X_j$ being equal to $A \Sigma A'$, we have, for every j: $\alpha_i = \mathrm{Var}\left[(A \overline{X}_j)[i]\right]$. This implies that $\frac{1}{\sqrt{\alpha_i}} (A \overline{X}_j)[i]$ is $\mathcal{N}(0,1)$-distributed for every j, hence by definition C_i is $\chi^2(n)$-distributed. ◇

A chi-square distribution with n degrees of freedom quickly tends towards a Gaussian distribution as n grows. A rule of thumb in probability theory is to consider the approximation $\chi^2(n) \approx \mathcal{N}(n, 2n)$ quite reasonable for $n \geq 30$. From Lemma 2, $\mathcal{L}_{k,x}$ is a sum between a constant, a Gaussian random variable and T chi-square random variables with $N\tau_x$ degrees of freedom. Therefore, for $N\tau_x$ large enough, we can consider that $\mathcal{L}_{k,x}$ has a Gaussian distribution. If this holds for every $x \in \mathcal{X}$ then the distribution of $\mathcal{L}_k$ can fairly be approximated by a Gaussian.

D Profiling SCA – Number of Leakage Measurements *vs.* Leakage Variance

Let us denotes the leakage covariance matrix by Σ and its estimation by $\widehat{\Sigma}$. Under the constant covariance assumption, (11) and (12) can be rewritten as:

$$\mathrm{E}\left[\mathcal{L}_k\right] = C_1 - \frac{1}{2} \sum_{x \in \mathcal{X}} \tau_x \left\| \widehat{\Sigma}^{-1/2} \left(m_{x,k^*} - \widehat{m}_{x,k}\right) \right\|^2 \ , \tag{32}$$

and

$$\mathrm{Cov}\left[\mathcal{L}_{k_1}, \mathcal{L}_{k_2}\right] = C_2 + \frac{1}{N} \sum_{x \in \mathcal{X}} \tau_x \ \left(m_{x,k^*} - \widehat{m}_{x,k_1}\right)' \, \widehat{\Sigma}^{-1} \, \Sigma \, \widehat{\Sigma}^{-1} \left(m_{x,k^*} - \widehat{m}_{x,k_2}\right) \ , \tag{33}$$

where C_1 and C_2 are some values constant with respect to k that satisfy $C_1 = \log\left(2\pi|\widehat{\Sigma}|\right) + \mathrm{Tr}\Big(\widehat{\Sigma}^{-1/2}\, \Sigma \left(\widehat{\Sigma}^{-1/2}\right)'\Big)$ and $C_2 = \frac{1}{2N} \left\| \widehat{\Sigma}^{-1/2}\, \Sigma\, (\widehat{\Sigma}^{-1/2})' \right\|_{hs}^2$.

We show in Section 6.1 that the success rate depends of the distribution of the comparison vector $\mathbf{c} = (c_k)_{k \in \mathcal{K}/\{k^*\}}$ that is defined, for Profiling SCA, by $c_k = \mathcal{L}_{k^*} - \mathcal{L}_k$ for every $k \in \mathcal{K}$. Assuming $(\mathcal{L}_k)_{k \in \mathcal{K}}$ Gaussian, $\mathbf{c}$ has a Gaussian distribution whose parameters satisfies:

$$\mathrm{E}\left[c_k\right] = \mathrm{E}\left[\mathcal{L}_{k^*}\right] - \mathrm{E}\left[\mathcal{L}_k\right] \ , \tag{34}$$

and

$$\mathrm{Cov}\left[c_{k_1}, c_{k_2}\right] = \mathrm{Var}\left[\mathcal{L}_{k^*}\right] + \mathrm{Cov}\left[\mathcal{L}_{k_1}, \mathcal{L}_{k_2}\right] - \mathrm{Cov}\left[\mathcal{L}_{k^*}, \mathcal{L}_{k_1}\right] - \mathrm{Cov}\left[\mathcal{L}_{k^*}, \mathcal{L}_{k_2}\right] \ . \tag{35}$$

From these expressions, we can see that the constant terms C_1 and C_2 of (32) and (33) cancel each other out in the expectation and the covariance matrix of $\mathbf{c}$. It thus appears that multiplying the leakage covariance matrix by a factor λ (and assuming that its estimation is also multiplied by λ) results in the multiplication of $m_{\mathbf{c}}$ and $\Sigma_{\mathbf{c}}$ by $1/\lambda$ while multiplying the number of leakage measurements by λ results in the multiplication of $\Sigma_{\mathbf{c}}$ by $1/\lambda$.

One can verify that the Gaussian cdf satisfies for every (a, b):

$$\Phi_{m/\lambda, \Sigma/\lambda^2}(a, b) = \Phi_{m, \Sigma}(\lambda a, \lambda b) \ . \tag{36}$$

As shown in Section 6.1, the success rate can be expressed as a sum of cdf $\Phi_{m_{\mathbf{c}}, \Sigma_{\mathbf{c}}}$ with inputs in $\{0, +\infty, -\infty\}^{|\mathcal{K}|-1}$. One thus deduces from (36) that multiplying $m_{\mathbf{c}}$ by $1/\lambda$ and $\Sigma_{\mathbf{c}}$ by $1/\lambda^2$ keeps the success rate unchanged. Hence we obtain that multiplying the leakage covariance matrix and multiplying the number of leakage measurements have complementary effects on the success rate of Profiling SCA.