Analyzing Side Channel Leakage of Masked Implementations with Stochastic Methods*

Kerstin Lemke-Rust and Christof Paar

Horst Görtz Institute for IT Security
Ruhr University Bochum
44780 Bochum, Germany
{lemke,cpaar}@crypto.rub.de

Abstract. Side channel cryptanalysis is a collective term for implementation attacks aiming at recovering secret or private keys from a cryptographic module by observing its physical leakage at run-time. Stochastic methods have already been introduced for first order differential side channel analysis. This contribution provides a compendium for the use of stochastic methods on masked implementations, i.e., on implementations that use internal random numbers in order to effectively prevent first order side channel attacks. Practical evidence is given that stochastic methods are also well suited for analyzing masked implementations, especially, as they are capable of combining several chosen components of different internal states for a multivariate side channel analysis.

Keywords: Side Channel Cryptanalysis, Stochastic Methods, Boolean Masking, Multivariate Side Channel Analysis, Higher-Order Side Channel Analysis.

1 Introduction

Traditionally, cryptanalysis employs mathematical tools to evaluate the security claims by cryptographers. However, once cryptographic schemes are implemented in integrated circuits such as smartcards, the resulting cryptographic implementation can no longer be solely seen as a mathematical object. Moreover and much more impressively, recent research results in the last decade have demonstrated that implementation attacks are a serious threat to cryptographic modules.

In reality, information flow also occurs on measurable physical channels such as the execution time, the power consumption, and the electromagnetic emanation of a cryptographic implementation. These channels are known as *side channels* in literature and may leak information on internal states of a cryptographic implementation at run-time. An adversary is said to be successful if side channel cryptanalysis yields to a critical entropy loss of a secret or private key used in a cryptographic implementation.

* Supported by the European Commission through the IST Contract IST-2002-507932 ECRYPT, the European Network of Excellence in Cryptology.

This contribution deals with the usage of instantaneous physical observables in the immediate vicinity of the cryptographic device, e.g., power consumption or electromagnetic radiation [9, 6]. The underlying working hypothesis for side channel cryptanalysis assumes that measurable observables depend on the internal state of a cryptographic algorithm whereby this dependency is specific for each implementation. This dependency can be predicted, e.g., by assuming a standard leakage model such as the Hamming weight or Hamming distance of internal states [2] at key recovery. More powerful attacks, however, aim at learning the exploitable side channel leakage prior to key recovery.

Such advanced attacks have been proposed by [4, 15]. They consist of two stages, a so-called *profiling stage* and a *key recovery stage*. Both stages use the same implementation, but usually differ in assumptions on the available knowledge. In the context of this contribution it is assumed that plaintext (or ciphertext), keys and masking values are known at the profiling stage. At key recovery only plaintext (or ciphertext) is assumed to be known to the adversary. Both approaches [4, 15] have in common that key recovery applies the maximum likelihood principle based on multivariate Gaussian probability densities that have been estimated during profiling. Reference [4] introduces template attacks that acquire empirical probability densities for each key dependency. For use with block ciphers, stochastic methods [15] approximate the probability density by considering selected basis functions of internal states for profiling, thus offering a significant reduction of required efforts. Additionally, by testing suitable basis functions a developer can gain insights in the nature of the side channel leakage. Reference [7] provides a performance analysis for templates and stochastic methods. It concludes that T-Test based templates are usually the method of choice, however, stochastic methods are of importance in case the adversary is limited at profiling. Limitations at profiling may be caused by a low number of measurements or by a high number of subkey dependencies for a given cipher.

In response to Differential Side Channel Analysis (DSCA) developers of cryptographic implementations may include randomization techniques such as secret splitting or masking schemes, e.g., [3, 5]. These randomization techniques shall prevent from predicting any relevant bit in any cycle of the implementation. As result, statistical tests using physical observables at *one* instant, i.e., first order side channel analysis, cannot be assumed to be successfully applied in key recovery. However, as already indicated in [9] high-order differential analysis can combine multiple instants from within a measurement trace. Previous work on second-order DSCA [11, 17] constructs a new leakage signal by multiplying (or subtracting) the observables at related time instants before statistics is applied. This reduction generally loses information if compared to a multivariate analysis. By assuming that the leakage signals follow the n-bit Hamming weight model [8] provided a derivation on the height of the expected second-order DPA signals and [14] uses predicted probability density functions to improve second-order power analysis. Further practical results for second- and higher-order DPA acting on the Hamming weight assumption are given by [13, 16]. Moreover, Template-enhanced DPA attacks were introduced in [1] to defeat masking under the assumption that

the adversary has access to an implementation with a biased random number generator during profiling. Reference [12] evaluates different types of template attacks attacks on masked implementations and concludes that a template based DPA attack leads to the best results.

This contribution elaborates stochastic methods for analyzing masked implementations and provides experimental case studies for a boolean masking scheme in Section 3.

2 Stochastic Methods on a Masked Implementation

2.1 Introduction to the Stochastic Model

The stochastic model [15] for non-masked implementations assumes that the adversary measures physical observables at time t in order to guess a subkey $k \in \{0,1\}^s$. The letter $x \in \{0,1\}^d$ denotes a known part of the data, i.e., plaintext or ciphertext, respectively. A physical observable $I_t(x,k)$ at time t is seen as a random variable

$$I_t(x,k) = h_t(x,k) + R_t \ . \tag{1}$$

The first summand $h_t(x,k)$ quantifies the deterministic part of the measurement as far it depends on x and k. The term R_t denotes a random variable that does not depend on x and k. R_t includes all kinds of noise as there are intrinsic and external noise, noise of the measurement apparatus and algorithmic noise that stems from deterministic contributions that do not depend on x and k. The random variable $I_t(x,k)$ is interpreted as 'displaced' noise R_t with mean displacement $h_t(x,k)$. Without loss of generality one may assume that $\mathbb{E}(R_t) = 0$ since otherwise one could replace $h_t(x,k)$ and R_t by $h_t(x,k) + \mathbb{E}(R_t)$ and $R_t - \mathbb{E}(R_t)$, respectively. It follows $\mathbb{E}(I_t(x,k)) = h_t(x,k)$. In this contribution, random variables are denoted with capital letters while their realizations, i.e. measured quantities, are denoted with the respective small letters.

Stochastic methods profile the real physical leakage by approximation within a suitable chosen vector subspace that is spanned by basis functions of one or several intermediate results of the cryptographic implementation. For practical purposes, it is important to note that the subkey space is chosen to be sufficiently small as at key recovery all subkey hypotheses have to be tested.

Example 1. For an unmasked implementation of AES, one may use the intermediate result $x_0 := S(x \oplus k)$, i.e. the 8-bit output of the AES S-box S given 8-bit data x and an 8-bit subkey k. A suitable vector subspace is spanned by the constant function 1 and the single bits of x_0. See [15] for more details.

In the stochastic model profiling aims to determine a function $h_t^*(\cdot,\cdot)$ that is 'close' to the unknown function $h_t(\cdot,\cdot)$. For simplicity, attention is restricted to the set of functions $\mathcal{F}_{u;t}$, that is a real vector subspace spanned by u known

functions $g_{tl}\colon \{0,1\}^d \times \{0,1\}^s \to \mathbb{R}$ for each instant t:

$$\mathcal{F}_{u;t} := \{h'\colon \{0,1\}^d \times \{0,1\}^s \to \mathbb{R} \mid \sum_{l=0}^{u-1} \beta_l g_{tl} \text{ with } \beta_l \in \mathbb{R}\} \tag{2}$$

One may assume that the functions g_{tl} are linearly independent so that $\mathcal{F}_{u;t}$ is isomorphic to $\mathbb{R}^u$.

2.2 The Stochastic Model for Masked Implementations

As already outlined in [15] the stochastic model of Section 2.1 can be generalized to the presence of masking. This generalized model assumes that the adversary measures physical observables at time t that additionally depend on a mask $y \in \{0,1\}^v$. A physical observable $I_t(x,y,k)$ at time t is seen as a random variable

$$I_t(x,y,k) = h_t(x,y,k) + R_t \tag{3}$$

The first summand $h_t(x,y,k)$ quantifies the deterministic part of the measurement as far it depends on x, y, and k. The term R_t denotes a random variable that does not depend on x, y, and k and fulfills $\mathbb{E}(R_t) = 0$.

For the approximation at masked implementation, a real vector subspace $\mathcal{F}_{u;t}$ is used that is spanned by u known functions $g_{tl}\colon \{0,1\}^d \times \{0,1\}^v \times \{0,1\}^s \to \mathbb{R}$ for each instant t:

$$\mathcal{F}_{u;t} := \{h'\colon \{0,1\}^d \times \{0,1\}^v \times \{0,1\}^s \to \mathbb{R} \mid \sum_{l=0}^{u-1} \beta_l g_{tl} \text{ with } \beta_l \in \mathbb{R}\} \tag{4}$$

The detailed algorithms for profiling masked implementations are provided in SubSect. 2.3. SubSect. 2.4 presents decision strategies for key recovery.

2.3 Profiling

Under the assumption that the adversary has access to the random numbers used for masking profiling works similar to [15]. In the context of this explanation, we assume that the adversary uses a known static key k for all N measurement vectors that are assumed to be available at the profiling stage. Alternatively, it is possible to use known variable keys instead that are loaded randomly for each measurement.

Profiling is split into up to three tasks:

1. Estimation of the deterministic part $h_t(x,y,k)$,
2. Selection of (relevant) instants for a multivariate characterization, and
3. Estimation of the multivariate noise.

This contribution considers the two main methods of the stochastic model [15]: the *minimum principle* and the *maximum likelihood principle*. Both differ in certain parts of the profiling and key recovery phase. While it is sufficient for the minimum principle to profile the deterministic side channel leakage and to select relevant instants, the maximum likelihood principle additionally requires an estimation of the multivariate noise. Table 1 summarizes the differences.

Table 1. Tasks for the minimum principle and the maximum likelihood principle. Note that the number of measurements N is split into two disjoint subsets N_1 and N_2 with $N_1 + N_2 = N$ for the maximum likelihood principle.

Method	Minimum Principle	Maximum Likelihood Principle
Estimation of the deterministic part	yes (N samples)	yes (N_1 samples)
Selection of instants	yes (N samples)	yes (N_1 samples)
Estimation of the noise	no	yes (N_2 samples)

Estimation of the deterministic part: For the following explanations a $p \times N$ matrix $\mathbf{I}$ is introduced that is defined by the N measurement vectors $\boldsymbol{i}_i \in \mathbb{R}^p$ with $1 \leq i \leq N$:

$$\mathbf{I} = \begin{pmatrix} i_{11}(x_1, y_1, k) & i_{12}(x_1, y_1, k) & \dots & i_{1p}(x_1, y_1, k) \\ i_{21}(x_2, y_2, k) & i_{22}(x_2, y_2, k) & \dots & i_{2p}(x_2, y_2, k) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ i_{N1}(x_N, y_N, k) & i_{N2}(x_N, y_N, k) & \dots & i_{Np}(x_N, y_N, k) \end{pmatrix} \tag{5}$$

The i-th row vector is the original i-th measurement vector. For each row vector there is an associated plaintext (or ciphertext) $x_i \in \{0,1\}^d$ and an associated mask $y_i \in \{0,1\}^v$. The j-th column vector of matrix $\mathbf{I}$ is $\boldsymbol{i_j}^{col} := (i_{1j}, i_{2j}, \dots, i_{Nj})^T$ and includes all measurement values for the same instant j. In the following, the notation $i_{ij}(x_i, y_i, k)$ is used instead of $i_t(x_i, y_i, k)$. Profiling of the deterministic part is done separately for each instant, i.e., for profiling purposes the column vector $\boldsymbol{i_j}^{col}$ is the starting point.

For each sampled instant $j \in \{1, \dots, p\}$, the adversary chooses u_j functions g_{jl} with $0 < l \leq u_j$ that span the vector subspace $\mathcal{F}_{u_j;j}$. In the presence of masking the vector subspace is reasonably spanned by two (or more) intermediate results that occur during computation. This yields a joint probability density of the side channel leakage at two (or more) intermediate results.

Example 2. One choice for profiling is to choose the n-bit intermediate results y and $x \oplus y \oplus k$ in case of a boolean masking scheme (see Section 3 for details).

One may define a $2n+1$ dimensional vector subspace that is spanned by the constant function 1 and the single bits of y and $x \oplus y \oplus k$.

The fitting problem to be solved is to find real valued coefficients $\boldsymbol{\beta_j} := (\beta_{j0}, \ldots, \beta_{j,u_j-1})^T$ such that the measurement quantities $i_{ij}(x_i, y_i, k)$ and controlled quantities $g_{j0}(x_i, y_i, k), \ldots, g_{j,u_j-1}(x_i, y_i, k)$ are linked by

$$i_{ij}(x_i, y_i, k) = \beta_{j0} + \sum_{l=1}^{u_j-1} \beta_{jl} g_{jl}(x_i, y_i, k) \; \forall i \in \{1, \ldots, N\}$$

in $\mathcal{F}_{u_j;j}$. This then yields an approximation on the deterministic part of the side channel for the instantiation of the stochastic variable I_j at sampled instant j given the controlled variables $g_{j0}(x, y, k), \ldots, g_{j,u_j-1}(x, y, k)$ with $g_{j0}(x, y, k)$ being the constant function 1. The coefficient β_{j0} gives the expectation value of the non-data dependent signal part. The coefficients β_{jl} with $l \neq 0$ are the data dependent signal portions, thereby indicating the points in time where exploitable side channel leakages occurs.

The coefficients $\boldsymbol{\beta_j}$ are approximated with least squares estimates. The $N \times u_j$ design matrix for this problem is

$$\mathbf{M} = \begin{pmatrix} 1 & g_{j1}(x_1, y_1, k) & g_{j2}(x_1, y_1, k) & \ldots & g_{j,u_j-1}(x_1, y_1, k) \\ 1 & g_{j1}(x_2, y_2, k) & g_{j2}(x_2, y_2, k) & \ldots & g_{j,u_j-1}(x_2, y_2, k) \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ 1 & g_{j1}(x_N, y_N, k) & g_{j2}(x_N, y_N, k) & \ldots & g_{j,u_j-1}(x_N, y_N, k) \end{pmatrix} . \tag{6}$$

The solution to the general linear least square problem is then given by

$$\boldsymbol{\beta_j^*} = \left(\mathbf{M}^T \mathbf{M}\right)^{-1} \mathbf{M}^T \boldsymbol{i_j}^{col} \tag{7}$$

provided that $\mathbf{M}^T\mathbf{M}$ is regular. As result, for each sampled instant $j \in \{1, \ldots, p\}$ the deterministic part is estimated by

$$\widetilde{h}_j^*(x, y, k) = \sum_{l=0}^{u_j-1} \beta_{jl}^* g_{jl}(x, y, k) \; . \tag{8}$$

Selection of (relevant) instants: It is highly desirable to sort out instants that do not contribute to the deterministic leakage in order to reduce noise as well as the dimension of the characterization problem for the maximum likelihood approach. Instant selection algorithms aim to find the points of interest j with significant contributing coefficients β_{jl}^* for $l > 0$. An appropriate method for selecting contributing instants is based on the T-Test and can be found in [7]. Previous algorithms based on the squared Euclidean norm $\|b\|^2 := \|(\beta_{j1}^*, \beta_{j2}^*, ..., \beta_{j,u_j-1}^*)\|^2 = \sum_{i=1}^{u_j-1} {\beta_{ji}^*}^2$ of contributing coefficients are given in [15].

Estimation of the multivariate noise: For the maximum likelihood principle one needs to determine a multivariate density. Therefore, it is assumed that the random vector $\boldsymbol{R_t}$ is jointly normally distributed with covariance matrix $\mathbf{C} = (c_{uv})_{1 \le u,v \le m}$, i.e. $c_{uv} := \mathbb{E}(R_{t_u} R_{t_v}) - \mathbb{E}(R_{t_u})\mathbb{E}(R_{t_v})$. For this task the adversary uses a complementary set that consists of $N_2 = N - N_1$ measurement curves to estimate the distribution of the m-dimensional random vector $\boldsymbol{R_t} = \boldsymbol{I_t}(X, Y, k) - \boldsymbol{h_t}(X, Y, k)$. Herein, the vector $\boldsymbol{t}$ stands for $(t_1, \ldots, t_m)$, i.e., the m selected instants. $\boldsymbol{R_t}$ and $\boldsymbol{I_t}$ denote the random vector $(R_{t_1}, \ldots, R_{t_m})$ and $(I_{t_1}, \ldots, I_{t_m})$, respectively. For the estimation of the covariance matrix $\mathbf{C}$, the adversary computes the N_2 vectors $\boldsymbol{z}_i := \boldsymbol{i_t}(x_i, y_i, k) - \widetilde{\boldsymbol{h}}^*_{\boldsymbol{t}}(x_i, y_i, k)$.

If the covariance matrix $\mathbf{C}$ is regular the random vector $\boldsymbol{R_t}$ is estimated to have the m-dimensional density $\widetilde{f}_{\mathbf{C}}$ with

$$\widetilde{f}_{\mathbf{C}} \colon \mathbb{R}^m \to \mathbb{R} \qquad \widetilde{f}_{\mathbf{C}}(\boldsymbol{z}) = \frac{1}{\sqrt{(2\pi)^m \det \mathbf{C}}} \exp\left(-\frac{1}{2}\boldsymbol{z}^T \mathbf{C}^{-1} \boldsymbol{z}\right) . \tag{9}$$

2.4 Key Recovery

Key recovery targets the same implementation that is now loaded with an unknown key k° and uses internally generated random numbers for masking. For the analysis, N_3 measurements are assumed to be available. In the key recovery phase, however, knowledge of the masks $y_1, \ldots, y_{N_3}$ cannot be assumed. Note that the measured quantities $\boldsymbol{i_t}(x_i, y_i, k^\circ)$ depend on two unknown values, the ever-changing value y_i and the fixed value k°.

From a logical point of view masking reduces the adversary's information. At a non-masked implementation the adversary is able to predict any intermediate result if the guessed subkey k is equal to k° [15]. Therefore the adversary is able to estimate on $\widetilde{\boldsymbol{h}}^*_{\boldsymbol{t}}(x_i, k)$ and to obtain a probability measure to indeed observe each subkey k. At a masked implementation the actual intermediate result has to be treated as an unknown number in the most general case. Instead of *one* intermediate result and therefore *one* estimated density a masked intermediate result can attain *all* outcomes $\widetilde{\boldsymbol{h}}^*_{\boldsymbol{t}}(x_i, 0, k)$ up to $\widetilde{\boldsymbol{h}}^*_{\boldsymbol{t}}(x_i, 2^v - 1, k)$. The best adversarial strategy is to use *all* possible outcomes weighted with the probability for each random number $y' \in \{0,1\}^v$. If these random numbers are unbiased and independent then $\mathbb{P}(y_i = y') = 2^{-v}$ for all $i \le N_3$ and $y' \in \{0,1\}^v$.

Minimum Principle: The adversary evaluates

$$\alpha_{MP}(x_1, \ldots, x_{N_3}; k) := \frac{1}{N_3} \sum_{i=1}^{N_3} \min_{y' \in \{0,1\}^v} \|\boldsymbol{i_t}(x_i, y_i, k^\circ) - \widetilde{\boldsymbol{h}}^*_{\boldsymbol{t}}(x_i, y', k)\|^2 \tag{10}$$

and decides for the subkey $k' \in \{0,1\}^s$ that minimizes $\alpha_{MP}(x_1, \ldots, x_{N_3}; k)$:

$$k' = \arg \min_{k \in \{0,1\}^s} \alpha_{MP}(x_1, \ldots, x_{N_3}; k) . \tag{11}$$

Equations (10) and (11) are referred to as the minimum principle in the presence of masking. Small values of the squared Euclidean norm $\|\boldsymbol{i_t}(x_i, y_i, k^\circ) - \widetilde{\boldsymbol{h}}_{\boldsymbol{t}}^*(x_i, y', k)\|^2$ indicate small deviations of the side channel leakage from the deterministic part and therefore enhance the probability for indeed observing the event of $y' = y_i$ and $k = k^\circ$. Accordingly, the guess of a subkey k is probably not correct if high values of the term $\|\boldsymbol{i_t}(x_i, y_i, k^\circ) - \widetilde{\boldsymbol{h}}_{\boldsymbol{t}}^*(x_i, y', k)\|^2$ are attained for all possible masks.

Maximum Likelihood Principle: The adversary hence decides for the subkey

$$k' = \arg \max_{k \in \{0,1\}^s} \alpha_{MLP}(x_1, \ldots, x_{N_3}; k) \tag{12}$$

that maximizes the term

$$\alpha_{MLP}(x_1, \ldots, x_{N_3}; k) := \prod_{i=1}^{N_3} \sum_{y' \in \{0,1\}^v} \mathbb{P}(y_i = y') \widetilde{f}_{\mathbf{C}} \left(\boldsymbol{i_t}(x_i, y_i, k^\circ) - \widetilde{\boldsymbol{h}}_{\boldsymbol{t}}^*(x_i, y', k) \right) \tag{13}$$

among all $k \in \{0,1\}^s$. The mixture of densities on the right-hand side of (13) also depends on the unknown random numbers $y_1, \ldots, y_{N_3}$.

Maximizing the term $\alpha_{MLP}(x_1, \ldots, x_{N_3}; k)$ in (13) is equivalent to maximizing $\ln(\alpha_{MLP}(x_1, \ldots, x_{N_3}; k))$. By setting $\boldsymbol{z}_{i,y',k} = \boldsymbol{i_t}(x_i, y_i, k^\circ) - \widetilde{\boldsymbol{h}}_{\boldsymbol{t}}^*(x_i, y', k)$ in the multivariate Gaussian density and neglecting constant factors of the Gaussian distribution in equation (9), $\ln(\alpha_{MLP}) := \ln(\alpha_{MLP}(x_1, \ldots, x_{N_3}; k))$ results in

$$\ln(\alpha_{MLP}) := \sum_{i=1}^{N_3} \ln \left(\sum_{y' \in \{0,1\}^v} \mathbb{P}(y_i = y') \cdot \exp\left(-\frac{1}{2} \cdot \boldsymbol{z}_{i,y',k}^T \mathbf{C}^{-1} \boldsymbol{z}_{i,y',k} \right) \right) . \tag{14}$$

For high values of N_3, using the term in (14) is the practical method of choice for guessing the subkey

$$k' = \arg \max_{k \in \{0,1\}^s} \ln(\alpha_{MLP}(x_1, \ldots, x_{N_3}; k)) . \tag{15}$$

3 Experimental Analysis of a Masked Implementation

Parameters with an impact on efficiency in side channel cryptanalysis include (i) the quantity of inherent leakage (chip dependent), (ii) the quality of the laboratory equipment (lab dependent), and (iii) the algorithms' ability to extract information (method dependent). Among them, this experimental analysis deals with the method dependent part. Experiments have been carried out with a standard microcontroller that is commercially available and does not incorporate any hardware side channel countermeasures. These experiments are used for a proof of concept of our proposed algorithms. The absolute success rates

given here are specific for our implementation. Moreover, improvements of the laboratory equipment used may be conceivable and the analysis of physically secured integrated circuits may require significantly enhanced efforts. For a fair comparison of different methods in the presence of masking, e.g., with the different types of template attacks in [12] it is absolutely necessary to use the same set of measurement vectors. The concrete physical leakage function of the cryptographic implementation has presumably also an impact on efficiency, i.e., some previously proposed methods may work nicely only if the physical leakage corresponds to the Hamming weight.

This section focuses on an application of the most general case for higher order analysis in the presence of masking, i.e., an implementation of boolean masking is considered that is typically the first step of, e.g., a masked AES or DES implementation. At a masked cryptographic implementation it is further necessary to switch the mask at non-linear or arithmetic operations which is not considered as part of this experimental analysis. Then the situation is even more favorable because of additional leakage signals, e.g., caused by S-Box processing or the use of a restricted form of masking.

The implementation done on an 8-bit microprocessor AT90S8515 proceeds as shown in Fig. 1. Note that there are different implementation choices of masking. Alternatively, it can be assumed that $x \oplus y$ is computed first before adding the key k, as, e.g., done in [11]. The motivation for the choice of this implementation in Fig. 1 is based on the fact that neither $x \oplus k$ nor $x \oplus y$ should be observable at a single point in time. Leakage on y, k, x, $y \oplus k$ and $y \oplus k \oplus x$, however, remains observable at single instants.

The physical channel used is the power consumption of the 8-bit microprocessor AT90S8515. By using the estimation of the deterministic part it was assured that first order differential analysis is prevented by verifying that leakage of the intermediate results $k \oplus x$ and $y \oplus x$ at single instants is negligible if any.

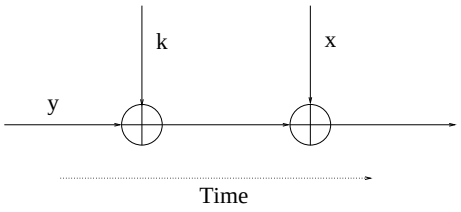

Fig. 1. Process of boolean masking

3.1 Profiling Phase

Concretely, four measurement series were recorded with $N = 10,000$ measurements each using different fixed keys. Further, one additional measurement series

includes $N = 20,000$ measurements with varying keys drawn randomly from a uniform distribution. All these series were used for profiling purposes.

For profiling a 'white-box' model is assumed, i.e., x, k, and y are known. Profiling applies the stochastic model at the two intermediate results $y \in \{0,1\}^8$ and $(y \oplus k \oplus x) \in \{0,1\}^8$. More concretely, the vector subspace is spanned by the constant function 1, the bits of y, and the bits of $x \oplus y \oplus k$ yielding an 17 dimensional vector subspace. In a first try, the deterministic part $h_j^*(x,y,k)$ was approximated by

$$\widetilde{h}_j^*(x,y,k) = \beta_{j0}^* \cdot 1 + \sum_{l=1}^{8} \beta_{jl}^* \cdot g_l(y) + \sum_{l=9}^{16} \beta_{jl}^* \cdot g_{l-8}(x \oplus y \oplus k) \; . \tag{16}$$

Herein, the function $g_l : \{0,1\}^8 \to \{0,1\}$ outputs the l-th bit of an 8-bit data item with a bit ordering from the most significant bit ($l = 1$) to the least significant bit ($l = 8$). The coefficients β_{jl}^* are determined by solving (7). As result, it turned out that leakage signals of y and $x \oplus y \oplus k$ are well separated in time. Therefore, it is appropriate to reduce the number of dimensions during profiling which helps in suppressing noise in the estimation process. For the refined application of (7), the chosen vector subspace depends on the time instant j, i.e., y is profiled if $0 < j \leq 2500$ and $x \oplus y \oplus k$ is profiled if $2500 < j \leq 10000$:

$$\widetilde{h}_j^*(x,y,k) = \begin{cases} \beta_{j0}^* \cdot 1 + \sum_{l=1}^{8} \beta_{jl}^* \cdot g_l(y) & \text{if } 0 < j \leq 2500 \\ \beta_{j0}^* \cdot 1 + \sum_{l=9}^{16} \beta_{jl}^* \cdot g_{l-8}(x \oplus y \oplus k) & \text{if } 2500 < j \leq 10000 \end{cases} \tag{17}$$

Accordingly, the coefficients β_{jl}^* are set to zero if not profiled in the given time frame:

$$\beta_{jl}^* := \begin{cases} 0 & \text{if } (9 \leq l \leq 16) \text{ and } (0 < j \leq 2500) \\ 0 & \text{if } (1 \leq l \leq 8) \text{ and } (2500 < j \leq 10000) \end{cases}$$

Fig. 2 shows the squared Euclidean norm of the data dependent coefficients β_{jl}^* with $l > 0$ as result of solving (7) given the vector subspaces of (17).

Profiling by using the vector subspace given in (17) was also applied to four measurement series with different fixed keys. By comparing experimental profiling results it turned out that the estimated coefficients β_{jl}^* differ significantly for different measurement series at a few instants (see Fig. 3). There are even leakage contributions that are completely suppressed in the series with varying keys, thus indicating that there is key dependent leakage which averages to zero if considering profiling based on varying keys. The selection of relevant instants for the series with varying keys yielded ten instants. For the series with fixed keys two additional signals were found and included in the set of selected instants.

According to Table 1, all N measurements were used for estimating β_{jl}^* for the minimum principle. Profiling for the maximum likelihood principle uses the setting $N_1 = N_2 = N/2$, i.e., half of the measurements were used for estimating β_{jl}^*. For the maximum likelihood principle, based on the selected instants, the covariance matrix for the multivariate Gaussian density was then computed as described in SubSect. 2.3.

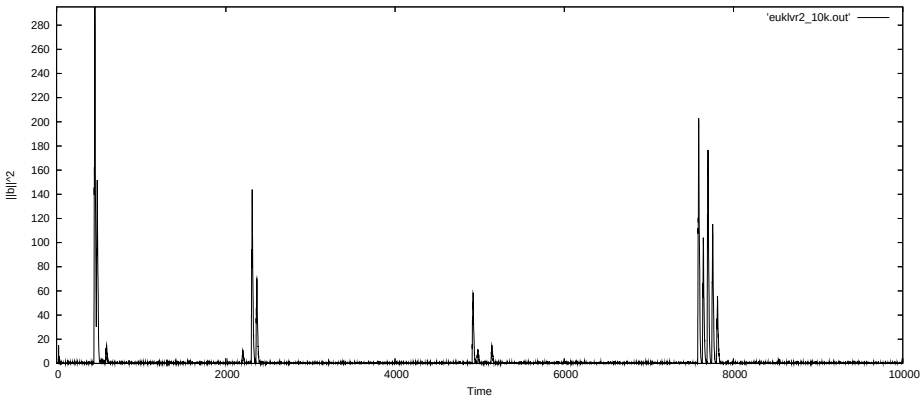

Fig. 2. Squared Euclidean norm $\|b\|^2 = \|(\beta^*_{j1}, \beta^*_{j2}, ..., \beta^*_{j16})\|^2$ of the bit depending coefficients as result of profiling according to equation (17) by using the measurement series with varying key data. Note that this computation includes two sets of basis functions in subsequent, but separated time frames. High values for $\|b\|^2$ indicate instants with significant deterministic side channel leakage.

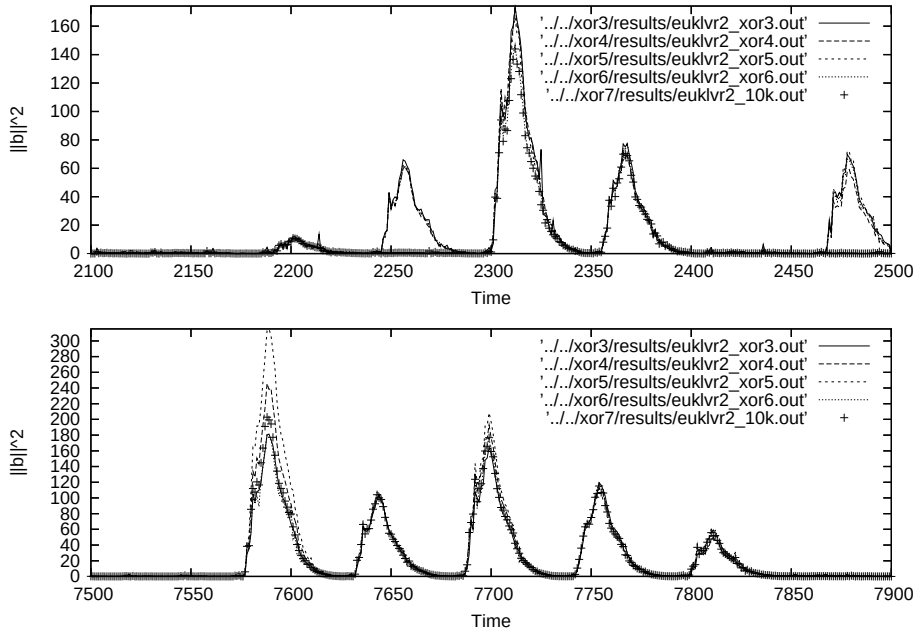

Fig. 3. Squared Euclidean norm $\|b\|^2 = \|(\beta^*_{j1}, \beta^*_{j2}, ..., \beta^*_{j16})\|^2$ of the bit depending coefficients after profiling for different series. In these time frames, e.g., in the upper plot around offset 2260 and 2480, profiling results with fixed keys give a signal part that is wiped out if profiling is done with varying keys. The lower plot shows resulting differences in the estimation for fixed keys. For each series, $10,000$ single measurements were used.

3.2 Key Recovery Phase

At key recovery, the adversary generally only knows x and aims at retrieving the fixed key k°. Key recovery was done at a series with fixed keys provided that the series assigned for profiling was different. As far as key recovery according to the maximum likelihood principle is concerned equations (14) and (15) were used.

Maximum Likelihood Principle: Results for 'varying key' profiling are based on a ten-dimensional covariance matrix and are summarized in Table 2.

Table 2. Success Rate (SR) that the correct key value is the best candidate as result of (14) by using N_3 randomly chosen measurements (100 repetitions). Profiling was done with variable keys with $N_1 = N_2 = 10000$.

N_3	SR Series no. 1	SR Series no. 2	SR Series no. 3	SR Series no. 4
10	8 %	9 %	5 %	6 %
20	20 %	24 %	14 %	13 %
30	26 %	33 %	22 %	24 %
50	53 %	53 %	36 %	34 %
100	82 %	78 %	55 %	62 %
200	92 %	95 %	79 %	82 %
400	98 %	99 %	93 %	96 %

Table 3. Profiling with fixed key for series no 2, no. 3, and no. 4 with $N_1 = N_2 = 5000$ and key recovery on series no 1. Success Rate (SR) that the correct key value is the best candidate as result of (14) by using N_3 randomly chosen measurements (100 repetitions).

N_3	SR Profiling Series no. 2	SR Profiling Series no. 3	SR Profiling Series no. 4
10	0 %	4 %	4 %
20	5 %	11 %	12 %
30	11 %	25 %	17 %
50	6 %	21 %	32 %
100	24 %	53 %	53 %
200	42 %	78 %	76 %
400	50 %	96 %	86 %

Note that in case of misses of the correct key value often a closely related key value differing only at one bit is obtained instead, especially at high values of N_3. Such an observation is reasonable, as differential side channel analysis on a boolean operation yields to related key hypotheses [10]. Results for 'fixed key' profiling can be found in Table 3 by using a twelve-dimensional covariance

matrix. Table 3 applies three different probability densities (obtained from series no. 2, 3, and 4) to series no. 1 yielding results of various quality. This is a clear indicator for a lack of symmetry at some instants. If comparing Table 2 with Table 3 the average success rate for key recovery is 69 % for 'varying key' profiling while it is 43 % for 'fixed key' profiling at $N_3 = 100$. Trial classifications on the profiling series themselves, however, yield success rates of 97 % at $N_3 = 100$. These results lead to two conclusions. First, it is indicated that profiling for all subkeys will clearly increase success rates and second, the use of a measurement series with varying keys is advantageous if profiling for all subkeys is not feasible, e.g., because of limitations at the profiling stage.

Maximum Likelihood Principle with Known Masking Values: Here, we consider an artificial case that key recovery can be done with known masking values, i.e., masking is completely ineffective. This might be a realistic case if the random number generator used for generating masking values is predictable, e.g., as result of physical modification or of special insights in the construction. The procedure for key recovery was modified in such a way that y is known and is equivalent to a first order side channel analysis. Results are presented in Table 4. For example, for $N_3 = 10$ the success rate to obtain the correct key value is 62.0 %. Among the key misses, a total amount of 25.5 % aggregates at eight related key values differing only by one bit from the correct key value. The security gain of masking in terms of N_3 can be quantified if comparing to Table 2, series no. 1. If considering success rates of about 90 %, N_3 is enlarged by roughly a factor of ten.

Table 4. Summarizing the results of key recovery with knowledge of masking values. Success Rate (SR) that the correct key value is the best candidate (1000 repetitions) on series no. 1.

N_3	SR (Known masking values)
2	8.8 %
3	17.2 %
5	31.3 %
7	46.0 %
10	62.0 %
20	89.1 %
30	97.3 %
50	99.5 %

Minimum Principle: For the application of the minimum principle, the series with varying keys was used for profiling and the choice of time instants was identical to the application of the maximum likelihood principle. The results give evidence that the minimum principle works in practice. Success rates of

42 % if $N_3 = 100$ and of 89 % if $N_3 = 400$ were obtained by using series no. 1 (see Table 5). These results can be compared with the column for series no. 1 in Table 2 and reveal a noticeable efficiency loss if compared to the maximum likelihood principle.

Table 5. Summarizing the results of the application of the minimum principle. Profiling was done on the series with varying keys and it was $N = 20,000$. Success Rate (SR) that the correct key value is the best candidate (100 repetitions) on series no. 1.

N_3	SR (Minimum Principle)
10	5 %
20	11 %
30	12 %
50	14 %
100	42 %
200	65 %
400	89 %

4 Conclusion

This contribution provides a compendium for the application of stochastic methods on masked implementations. Stochastic methods do not require any assumption of the physical leakage model, instead they provide an approximation of the side channel leakage in any given vector subspace as result of the profiling stage. Because of that, stochastic methods are indeed a practical alternative for analyzing masked cryptographic implementations and may be important for a developer of a masked cryptographic implementation in order to figure out remaining side channel leakage.

References

1. Dakshi Agrawal, Josyula R. Rao, Pankaj Rohatgi, and Kai Schramm. Templates as Master Keys. In Josyula R. Rao and Berk Sunar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2005*, volume 3659 of *LNCS*, pages 15–29. Springer, 2005.
2. Eric Brier, Christophe Clavier, and Francis Olivier. Correlation Power Analysis with a Leakage Model. In M. Joye and J.-J. Quisquater, editors, *Cryptographic Hardware and Embedded Systems - CHES 2004*, volume 3156 of *LNCS*, pages 16–29. Springer, 2004.
3. Suresh Chari, Charanjit S. Jutla, Josyula R. Rao, and Pankay Rohatgi. Towards Sound Approaches to Counteract Power-Analysis Attacks. In M. Wiener, editor, *Advances in Cryptology – CRYPTO '99*, volume 1666 of *LNCS*, pages 398–412. Springer, 1999.

4. Suresh Chari, Josyula R. Rao, and Pankaj Rohatgi. Template Attacks. In B. S. Kaliski, Ç Koç, and C. Paar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2002*, volume 2523 of *LNCS*, pages 13–28. Springer, 2003.
5. Jean-Sébastian Coron and Louis Goubin. On Boolean and Arithmetic Masking against Differential Power Analysis. In Ç.K. Koç and C. Paar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2000*, volume 1965 of *LNCS*, pages 231–237. Springer, 2000.
6. Karine Gandolfi, Christophe Mourtel, and Francis Olivier. Electromagnetic Analysis: Concrete Results. In Ç Koç, D. Naccache, and C. Paar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2001*, volume 2162 of *LNCS*, pages 251–261. Springer, 2001.
7. Benedikt Gierlichs, Kerstin Lemke-Rust, and Christof Paar. Templates vs. Stochastic Methods. In Louis Goubin and Mitsuru Matsui, editors, *Cryptographic Hardware and Embedded Systems - CHES 2006*, volume 4249 of *LNCS*, pages 15–29. Springer, 2006.
8. Marc Joye, Pascal Paillier, and Berry Schoenmakers. On Second-Order Differential Power Analysis. In Josyula R. Rao and Berk Sunar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2005*, volume 3659 of *LNCS*, pages 293–308. Springer, 2005.
9. Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential Power Analysis. In M. Wiener, editor, *Advances in Cryptology – CRYPTO '99*, volume 1666 of *LNCS*, pages 388–397. Springer, 1999.
10. Kerstin Lemke, Kai Schramm, and Christof Paar. DPA on n-Bit Sized Boolean and Arithmetic Operations and Its Application to IDEA, RC6, and the HMAC-Construction. In M. Joye and J.-J. Quisquater, editors, *Cryptographic Hardware and Embedded Systems - CHES 2004*, volume 3156 of *LNCS*, pages 205–219. Springer, 2004.
11. Thomas S. Messerges. Using Second-Order Power Analysis to Attack DPA Resistant Software. In Ç.K. Koç and C. Paar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2000*, volume 1965 of *LNCS*, pages 238–251. Springer, 2000.
12. Elisabeth Oswald and Stefan Mangard. Template Attacks on Masking – Resistance is Futile. In Masayuki Abe, editor, *Topics in Cryptology – CT-RSA 2007, The Cryptographers' Track at the RSA Conference 2007*, volume 4377 of *LNCS*, pages 243–256. Springer, 2006.
13. Elisabeth Oswald, Stefan Mangard, Christoph Herbst, and Stefan Tillich. Practical Second-Order DPA Attacks for Masked Smart Card Implementations of Block Ciphers. In David Pointcheval, editor, *Topics in Cryptology - CT-RSA 2006, The Cryptographers' Track at the RSA Conference 2006*, volume 3860 of *LNCS*, pages 192–207. Springer, 2006.
14. Eric Peeters, François-Xavier Standaert, Nicolas Donckers, and Jean-Jacques Quisquater. Improved Higher-Order Side-Channel Attacks with FPGA Experiments. In Josyula R. Rao and Berk Sunar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2005*, volume 3659 of *LNCS*, pages 309–323. Springer, 2005.
15. Werner Schindler, Kerstin Lemke, and Christof Paar. A Stochastic Model for Differential Side Channel Cryptanalysis. In Josyula R. Rao and Berk Sunar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2005*, volume 3659 of *LNCS*, pages 30–46. Springer, 2005.

16. Kai Schramm and Christof Paar. Higher Order Masking of the AES. In David Pointcheval, editor, *Topics in Cryptology - CT-RSA 2006, The Cryptographers' Track at the RSA Conference 2006*, volume 3860 of *LNCS*, pages 208–225. Springer, 2006.
17. Jason Waddle and David Wagner. Towards Efficient Second-Order Power Analysis. In M. Joye and J.-J. Quisquater, editors, *Cryptographic Hardware and Embedded Systems - CHES 2004*, volume 3156 of *LNCS*, pages 1–15. Springer, 2004.