A New Difference Method for Side-Channel Analysis with High-Dimensional Leakage Models

Annelie Heuser[1,4], Michael Kasper[2,4],
Werner Schindler[3,4], and Marc Stöttinger[1,4]

[1] Darmstadt University of Technology, Germany
{Heuser,Stoettinger}@iss.tu-darmstadt.de
[2] Fraunhofer Institute for Secure Information Technology (SIT), Germany
Michael.Kasper@sit.fraunhofer.de
[3] Bundesamt für Sicherheit in der Informationstechnik (BSI), Germany
Werner.Schindler@bsi.bund.de
[4] Center for Advanced Security Research Darmstadt (CASED), Germany

Abstract. The goal of the DPA contest v2 (2009 – 2010) was to find the most efficient side-channel attack against a particular unprotected AES-128 hardware implementation. In this paper we discuss two problems of general importance that affect the success rate of profiling based attacks, and we provide effective solutions. First, we consider the impact of temperature variations on the power consumption, which causes a so-called *drifting offset*. To cope with this problem we introduce a new method called *Offset Tolerant Method (OTM)* and adjust OTM to the stochastic approach (SA-OTM). The second important issue of this paper concerns the choice of an appropriate leakage model as this determines the success rate of SA and SA-OTM. Experiments with *high-dimensional* leakage models show that the overall leakage is not only caused by independent transitions of bit lines. Compared to the formely best submitted attack of the DPA contest v2 the combination of SA-OTM with high-dimensional leakage models reduces the required number of power traces to 50%.

Keywords: Side-Channel Analysis, Stochastic Approach, Environmental Influences, Drifting Offset, High-dimensional Leakage Models.

1 Introduction

For more than a decade side-channel analysis has been an important field of research in both academia and industry. Usually these attacks apply mathematical techniques, e.g., statistical methods, to exploit compromising side-channel leakage (e.g., runtime behavior, power consumption or electromagnetic emanation), which is emitted during the regular execution of a cryptographic algorithm. Power attacks can be divided into non-profiled and profiled methods. Prominent representatives of non-profiled side-channel attacks are Differential Power Analysis [11], Correlation Power Analysis [2], and Mutual Information Analysis [6]. These attacks try to recover the secret information without a preceding profiling

O. Dunkelman (Ed.): CT-RSA 2012, LNCS 7178, pp. 365–382, 2012.

phase. Profiled side-channel attacks, such as template attacks [3] or the stochastic approach [17], have the potential to be much more powerful and efficient. In a profiling based attack an adversary (attacker, designer, evaluator) uses a training device to characterize the leakage of a cryptographic implementation by creating templates or by developing a well-fitted leakage model. Then he tries to recover the key from the target device, using the knowledge from the profiling phase.

Generally speaking, measurements performed by different laboratories are often difficult to compare due to different acquisition platform sensitivities, different implementations of cryptographic algorithms, noise and other environmental influences. The organizers of the DPA contest v2 [4] provided measurement traces that allow a fair comparison of several side-channel attacks. We decided to apply the stochastic approach.

In this contribution we deal with the two important problems that may affect the success rate of profiling based attacks. First, we highlight difficulties that arise from environmental influences during the acquisition phase. Motivated by the DPA contest v2 measurements we investigate the impact of temperature variations. In fact, variations of the environmental temperature as well as temperature variations inside the device may change the (average) level of power consumption and thus the level of electrical current and voltage consumption. We denote this unexpected phenomenon as *drifting offset*. The origin of temperature variations, their impact on the power and current consumption, and possible preventive measures are discussed in Sect. 2. In Sect. 4 we introduce a new algorithmic method, which we denote as *Offset Tolerant Method* (OTM) and integrate it into the stochastic approach, abbreviated by SA-OTM.

Second, we consider the precise representation of the compromising side-channel leakage by suitable leakage models. Profiling-based attacks are very powerful and effective tools but their efficiency strongly depends on the suitability of the applied leakage model. As stated in [13] several formal works assume that independent transitions of bit lines imply independent contributions of side-channel leakage. If this assumption is valid a leakage model that only considers the input/output bits of the SBox separately will be sufficient. However, Renauld et al. [13] uses Mutual Information Analysis as an information theoretic metric [18] to show that this assumption may not always be valid in practice. With regard to this observation we apply different *high-dimensional* leakage models, which represent the individual leakage of each bit line as well as the leakage caused by the combination of several bit lines. Referred to the DPA contest v2 the combination of SA-OTM with high-dimensional leakage models results in the best success rates.

2 Extrinsic and Environmental Influences on Side-Channel Evaluation Process

It is well-known that extrinsic and environmental influences as temperature, cosmic radiation and terrestrial radiation have an impact on the design in terms of

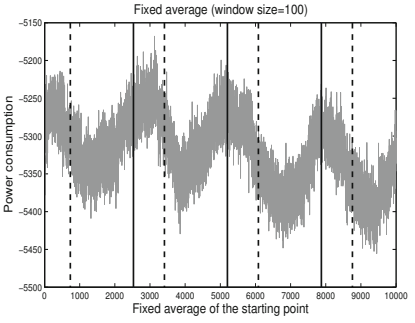

Fig. 1. Fixed average of the starting point; points in time: solid line 04:05 pm (starting time + 24h), dotted line 0:00 (midnight + 24h)

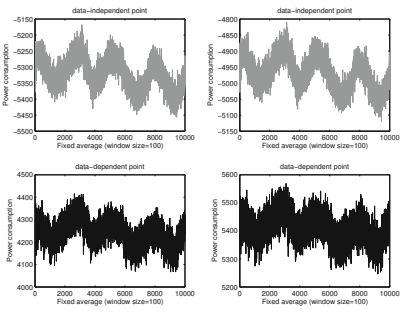

Fig. 2. Drifting offset at data-dependent and data-independent time points

the reliability and dependability of the integrated circuits functionality. However, the relevance of these phenomena in security analysis and, in particular in the side-channel analysis, have not been subject of public discussions yet. Usually, it is tacitly assumed that the power traces are recorded under constant environmental conditions. Neither temperature changes nor explicit influences caused by variations in the temperature of a system state are considered. We found out that the power curves of the DPA contest v2 [4] show a *drifting offset*, which might result from temperature variations. The template base of the DPA contest v2 consists of 1.000.000 traces, which were recorded during approximately 3 days and 19 hours. To give evidence for the drifting offset we selected particular time instants for all power traces and calculated the mean value over non-overlapping sets of 100 subsequent traces, which gave 10.000 mean values. Figure. 1 shows the mean power values at the starting point of the power traces. The dotted line corresponds to the beginning of a new day, and the solid line marks a 24 hour cycle. Figure. 1 illustrates the correspondence of the mean power consumption to the diurnal rhythm. Figure 2 shows the drifting offset for data-independent time points where no encryption is performed (gray) and for data-dependent points (black). Obviously, the drifting offset is larger than (average) effects that stem from data-dependent computations.

In order to confirm that the environmental temperature is the true reason for the existence of drifting offsets we performed own measurements on the SASEBO-GII platform. We simulated environmental temperature variations by mounting a peltier element and a cooling / heating system on the surface of the target FPGA. Figure 3 shows the voltage drop over a measurement shunt V_M for a single time instant. This reveals the direct relation between the environmental temperature and the power consumption of a device, which is proportionally bounded to the measured voltage drop of the shunt in the ground line. Note that the voltage drop at trace $\approx$ 7500 results from the activation of the peltier

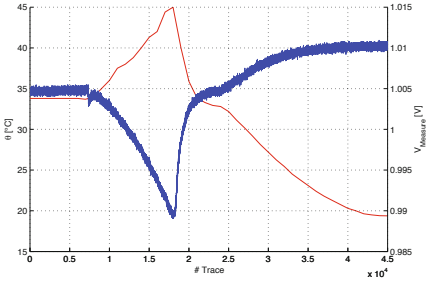

Fig. 3. Dependency between environmental temperature and power consumption. The thin red line represents the temperature while the thick blue line stands for the measured voltage drop.

element. Physical coherencies and possible preventive measures are discussed in the following.

3 Impact of Environmental Conditions

In the present section we analyze the impact of environmental conditions on the power consumption. We focus on the material specific temperature coefficient, denoted by α_{θ_0}, and on the impact to the characteristic ohmic resistance $\vartheta(\theta_0)$ of a target circuit. Eq. (1) provides a (linearized) formula that expresses the impact of the difference between the actual temperature θ and a reference temperature θ_0 on an ohmic resistor

$$R(\theta) = \vartheta(\theta_0) \cdot (1 + \alpha_{\theta_0} \cdot (\theta - \theta_0)) \,. \tag{1}$$

A measurement circuit usually consists of a target circuit (e.g., an FPGA configured with the cryptographic 'target' implementation) and a set of further electronic board components. The measurement circuit is usually realized by an ohmic shunt, which is chained between the target and a stable power supply. The voltage drop over this shunt V_M is used to calculate the power consumption of the target. The voltage divider (2) provides a simplified model for the relation between the supply voltage V_{cc} of the target device and V_M,

$$V_M = \frac{R_{board}}{R_{board} + R_{target}} \cdot V_{cc} \,, \tag{2}$$

where R_{target} denotes the ohmic resistance of target under attack and R_{board} denotes the overall resistance of all ohmic components of the above mentioned measurement circuit. Substituting Eq. (1) into the voltage divider for V_M gives Eq. (3)

$$V_M = \frac{V_{cc}}{1 + \left(\frac{\vartheta_{target}(\theta_{0,target})}{\vartheta_{board}(\theta_{0,board})}\right) \cdot \left(\frac{1+\alpha_{\theta_{0,target}} \cdot (\theta_{target} - \theta_{0,target})}{1+\alpha_{\theta_{0,board}} \cdot (\theta_{board} - \theta_{0,board})}\right)} \,. \tag{3}$$

Further experiments with the SASEBO-G II FPGA evaluation board verified that the impact of $\alpha_{\theta_{0,board}}$ is much smaller than the impact of $\alpha_{\theta_{0,target}}$. Consequently, the power consumption is more affected by temperature variations on the FPGA than by the temperature variations of the shunt and other board components. Hence Eq. (3) may be simplified to

$$V_M = \frac{V_{cc}}{1 + \left(\frac{\vartheta_{target}(\theta_{0,target})}{\vartheta_{board}(\theta_{0,S})}\right) \cdot \left(1 + \alpha_{\theta_{0,target}} \cdot \underbrace{(\theta_{target} - \theta_{0,target})}_{\substack{heating>0,\\ cooling<0}}\right)}. \tag{4}$$

For the SASEBO-G II FPGA evaluation board (as for similar evaluation boards) V_M increases significantly if the target FPGA is cooled down and decreases significantly if the FPGA is heated, which coincides with the expositions in Sect. 2.

Preventing drifting offsets by providing constant environmental conditions. An intuitive and natural method to prevent drifting offsets is to keep the temperature of both the device and the environment constant. The environmental temperature can be controlled when using a heating cabinet or a climatic chamber. The device may be preheated or cooled during the measurements in order to stabilize the temperature of the device.

However, these measures reduce the thermal effects only to a certain level. Since the thermal processes are very slow and the response time is very long it is yet difficult to control them precisely. Thus, the temperature gradient between the device and the environment should to be stable for a certain time interval. This time interval is certainly shorter than the full profiling measurement period. Moreover, the adversary may not have unlimited access to the target device so that these measures may not always be possible. This raises several questions for further research: If an attacker is able to learn on a training device of the same type under stable environmental conditions, can he also ensure constant conditions during the attack? Is it possible to enforce identical conditions in different situations? If not: how can unstable environmental conditions be handled efficiently?

4 A Novel Method for Effective Offset Elimination

Like for other attacks that (maybe implicitly) consider the average power consumption, e.g., template attacks, the efficiency of the stochastic approach may decrease significantly in presence of drifting offsets (Figs. 1 and 2). In the light of Sects. 2 and 3 we may assume that the offset drifts slowly. An intuitive approach to eliminate drifting offsets is to consider differences of consecutive power traces in place of the power traces themselves. This is an *Offset Tolerant Method* (OTM), and we adjust this method to the stochastic approach (SA), abbreviated by SA-OTM. This may sound simple, however, it will turn out later that several mathematical difficulties have to be overcome.

4.1 The 'Normal' Stochastic Approach: A Brief Summary

The 'normal' stochastic approach is an established, effective method in profiled power analysis, which combines engineer's knowledge and expertise with advanced stochastic methods [17, 7, 16, 10, 8]. In this subsection we summarize its central steps. In Subsection 4.2 we will refer to this description, and we work out the differences to SA-OTM. Principal component analysis (PCA) is well-known in the context of template attacks [1]. Below we adjust PCA to the stochastic approach. We begin with some notations.

Notation 1. We denote subkeys by $k \in \{0,1\}^s$ while $x \in \{0,1\}^p$ stands for (the relevant part of) the plaintext or ciphertext, respectively (typically, 8 or 16 bits). Random variables are denoted by capital letters, realizations thereof, i.e. values taken on by these random variables, by the corresponding small letters. Vectors are written in bold, e.g., $\mathbf{t}$ stands for $(t_1, \ldots, t_m)$, and $\mathbf{R_t}$ denotes the random vector $(R_{t_1}, \ldots, R_{t_m})$. Accordingly, $\mathbf{I_t}(x,k)$, $\mathbf{i_t}(x,k)$, $\mathbf{h}^*_{\mathbf{t};k}(x,k)$ etc. while $\sim$ indicates estimates. We write $\mathrm{diag_n}(\mathrm{b}_1, \ldots, \mathrm{b_n})$ for a diagonal $n \times n$ square matrix with diagonal elements $b_1, \ldots, b_n$, and $\mathcal{N}_n(\mu, F)$ denotes an n-dimensional normal distribution with mean vector μ and covariance matrix F. Finally, $f_F(\cdot)$ denotes the density of $N(0, F)$.

The stochastic approach refers to the mathematical model

$$I_t(x,k) = h_t(x,k) + R_t \tag{5}$$

where t denotes a time instant. The power consumption $i_t(x,k)$ is interpreted as a realization of a random variable $I_t(x,k)$ whose (unknown) distribution depends on the pair (x,k). The leakage function $h_{t;k}(x,k)$ quantifies its deterministic part, which depends on x and k, while R_t denotes the noise. W.l.o.g. we may assume $E(R_t) = 0$. Note that both the leakage function $h_t(\cdot,\cdot)$ and the distribution of the noise are unknown and thus have to be estimated.

Profiling Phase. Let $t \in \{t_1, \ldots, t_m\}$ and $k \in \{0,1\}^s$ be fixed for the moment. We view the restricted function $h_{t;k} \colon \{0,1\}^p \times \{k\} \to \mathbb{R}$, $h_{t;k}(x,k) := h_t(x,k)$ as an element of the 2^p-dimensional real vector space $\mathcal{F}_k := \{h' \colon \{0,1\}^p \to \mathbb{R}\}$. Basis functions $g_{0,j;k}(\cdot,k) = 1$ (constant function), $\ldots, g_{u-1,t;k}(\cdot,k)$ shall be selected under consideration of the concrete implementation, since they shall capture the relevant source of side-channel leakage (cf. e.g., [8] and Sect. 5). The SA does not aim at the exact function $h_{t;k}(\cdot,k)$ itself but at its best approximator $h^*_{t;k}(\cdot,k)$ in $\mathcal{F}_{u,t;k}$, the subspace which is spanned by $g_{0,j;k}(\cdot,k), \ldots, g_{u-1,t;k}(\cdot,k)$. Using the power measurements $i_t(x_1,k), \ldots, i_t(x_{N_1},k) \in \mathbb{R}$ the least square estimate $\widetilde{h}^*_{t;k}(\cdot,k)$ of $h_{t;k}(\cdot,k)$ is determined. Let

$$A := \begin{pmatrix} g_{0,t;k}(x_1,k) & \ldots & g_{u-1,t;k}(x_1,k) \\ \vdots & & \vdots \\ g_{0,t;k}(x_{N_1},k) & \ldots & g_{u-1,t;k}(x_{N_1},k) \end{pmatrix}. \tag{6}$$

If A^TA is regular (usual case) the normal equation $A^TA\mathbf{b} = A^T\mathbf{i}_t$ has unique solution

$$\widetilde{\mathbf{b}}^* = (A^TA)^{-1}A^T\mathbf{i}_t, \qquad \text{with} \quad \widetilde{\mathbf{b}}^* := (\widetilde{\beta}_0^*, ..., \widetilde{\beta}_{u-1}^*), \quad \text{and} \tag{7}$$

$$\widetilde{h}_{t;k}^*(\cdot,k) = \sum_{j=0}^{u-1} \widetilde{\beta}_{j,t;k}^* g_{j,t;k}(\cdot,k) \quad \text{(least square estimate of } h_{t;k}^*(\cdot,k)\text{)}. \tag{8}$$

The coefficients $\tilde{\beta}_{0,t;k}^*, \ldots, \tilde{\beta}_{u-1,t;k}^*$ are called β-characteristic.

In the second profiling step the covariance matrix C of the noise vector $\mathbf{R_t}$ is estimated, finally yielding a density for the random vector $\mathbf{I_t}(x,k)$. From an information theoretic point of view it seems to be advisable to consider as many time instants $t_1 < \cdots < t_m$ as possible. Unfortunately, then the covariance matrix C is often 'almost' singular so that even moderate estimation errors in $\widetilde{C}$ may amplify drastically in $\widetilde{C}^{-1}$ (needed to calculate $f_C(\cdot)$), and matrix inversion becomes an ill-posed numerical problem. Since C and its estimate $\widetilde{C}$ are symmetric positive semi-definite matrices an orthogonal matrix $P \in O(m)$ exists, for which $P^T\widetilde{C}P = \widetilde{D}_m$ with $\widetilde{D}_m = \mathrm{diag_m}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_\mathrm{m})$. The diagonal elements $\widetilde{\lambda}_1 \geq \cdots \geq \widetilde{\lambda}_m \geq 0$ (eigenvalues of $\widetilde{C}$), and the j^{th} column v_j of P is an eigenvector of $\widetilde{C}$ to eigenvalue $\widetilde{\lambda}_j$ (main axis transformation). If the first s eigenvalues are considerably larger than the others, i.e. $\widetilde{\lambda}_{s+1} \ll \widetilde{\lambda}_s$, we concentrate on that subspace of $\mathbb{R}^\mathrm{m}$, which is spanned by the eigenvectors $v_1, \ldots, v_s$. More precisely, if P_s denotes the $(m \times s)$-matrix with columns $v_1, \ldots, v_s$ then

$$P_s^T\widetilde{C}P_s = \widetilde{D}_s \quad \text{with } \widetilde{D}_s = \mathrm{diag_s}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_\mathrm{s}) \quad \text{(PCA)}. \tag{9}$$

If the random vector Y is $\mathcal{N}_m(0,C)$-distributed then P_s^TY is $\mathcal{N}_s(0, P_s^TCP_s)$-distributed ([9]), i.e. has the s-dimensional normal density f_{D_s}. For large m it is not advisable to calculate P_s and $\widetilde{D}_s$ via main axis transformation of $\widetilde{C}$. Instead, one should apply the singular value decomposition [9] as it is numerically more stable.

Attack Phase. In the attack phase the adversary performs N_3 measurements at the target device and obtains power vectors $\mathbf{i_t}(x_1, k^\dagger), \ldots, \mathbf{i_t}(x_{N_3}, k^\dagger)$ with the unknown subkey $k^\dagger$. The adversary decides for that subkey candidate $k^* \in \{0,1\}^s$ that maximizes

$$\prod_{l=1}^{N_3} f_{\widetilde{D}_s}\left(P_s^T\left(\mathbf{i_t}(x_l,k^\dagger) - \widetilde{\mathbf{h}}_{\mathbf{t};k}^*(x_l,k^*)\right)\right). \tag{10}$$

4.2 SA-OTM: A New Variant of SA

In the following we assume that the power traces are labelled in the same order as they have been recorded, and that the data-independent offset drifts slowly. We denote the offsets at time t by $\tau_{t;1}, \tau_{t;2}, \ldots$ where the second index indicates

the number of the power trace. In particular, $\tau_{t;l} - \tau_{t;l+1} \approx 0$ for all $l \geq 1$. For 'normal' SA $\widetilde{\beta}^*_{0,t,k}$ estimates the average power consumption in the profiling phase. Note that this average might differ from the corresponding value within the attack phase. Moreover, regarding the measurements of the DPA contest v2 the ratio $|\widetilde{\beta}^*_{0,t,k}|/\sum_{j=1}^{8}|\widetilde{\beta}^*_{j,t,k}| \approx 70$, and hence even moderate relative differences in $\widetilde{\beta}^*_{0,t,k}$ might have considerable impact on the attack efficiency. We refine (5) and get

$$I_t(x_l, k) = h_t(x_l, k) + \tau_{t;l} + R_t\,. \tag{11}$$

In particular, $I_t(x_l, k) \sim N(h_t(x_l, k) + \tau_{t;l}, \sigma^2)$. Of course, if $\tau_{t;l} = 0$ for all power traces (11) reduces to (5). SA-OTM applies to the enhanced mathematical model (11).

SA-OTM: Profiling Phase

Estimation of $h^{*\circ}_{t,k}$ and of the β-characteristic Since the drifting offset $\tau_{t;l}$ only affects the coefficient $\beta_{0,t;k}$ in contrast to 'normal' SA we do not aim at $h^*_{t,k}(\cdot, k)$ but at $h^{*\circ}_{t,k}(\cdot, k) := \sum_{j=1}^{u-1}\beta^*_{j,t;k}g_{j,t;k}(\cdot, k)$. In place of $\mathcal{F}_{u,t;k}$ we consider the subspace

$$\mathcal{F}^{\circ}_{u,t;k} := \{h' \colon \{0,1\}^p \times \{k\} \to \mathbb{R} \mid \mathrm{h'} = \sum_{\mathrm{j=1}}^{\mathrm{u-1}} \beta'_{\mathrm{j,t,k}}\mathrm{g_{j,t;k}} \text{ with } \beta'_{\mathrm{j,t,k}} \in \mathbb{R}\}, \tag{12}$$

i.e., we neglect the first basis vector $g_{0,t;k} = 1$. The straight-forward approach is to proceed as in 'normal' SA, by simply cancelling the first column of matrix A (Eq. (6)).

Alternatively, one may consider differences of consecutive power measurements. More precisely, for $l = 1, \ldots, N_1 - 1$ let $d_{j,t,k}(x_l, x_{l+1}, k) := g_{j,t;k}(x_l, k) - g_{j,t;k}(x_{l+1}, k)$ and $di_t(x_l, x_{l+1}, k) := i_t(x_l, k) - i_t(x_{l+1}, k)$. Further, we define the $(N_1 - 1)$-dimensional vector $\mathbf{\Delta i}_t := (di_t(x_l, x_2, k), \ldots, di_t(x_{N_1-1}, x_{N_1}, k))$ and

$$A^{\circ} := \begin{pmatrix} d_{1,t;k}(x_1, x_2, k) & \ldots & d_{u-1,t;k}(x_1, x_2, k) \\ \vdots & \ddots & \vdots \\ d_{1,t;k}(x_{N_1-1}, x_{N_1}, k) & \ldots & d_{u-1,t;k}(x_{N_1-1}, x_{N_1}, k) \end{pmatrix}. \tag{13}$$

If the $(u-1 \times u-1)$ dimensional matrix product $(A^{\circ T}A^{\circ})$ is regular then in analogy to (7) and (8) we obtain

$$\widetilde{\mathbf{b}}^{*\circ} = (A^{\circ T}A^{\circ})^{-1}A^{\circ T}\mathbf{\Delta i}_t \qquad \text{with} \quad \widetilde{\mathbf{b}}^{*\circ} := (\widetilde{\beta}^*_1, ..., \widetilde{\beta}^*_{u-1}), \text{ and} \tag{14}$$

$$\widetilde{h}^{*\circ}_{t;k}(\cdot, k) = \sum_{j=1}^{u-1}\widetilde{\beta}^*_{j,t,k}g_{j,t;k}(\cdot, k) \quad \text{(least square estimate of } h^{*\circ}_{t;k}(\cdot, k))\,. \tag{15}$$

For infinite sample size N_1 the estimates $\widetilde{\beta}^*_{1,t;k}, \ldots, \widetilde{\beta}^*_{u-1,t;k}$ from both estimation methods ('straight-forward', 'difference method') converge to the exact parameter values $\beta^*_{1,t;k}, \ldots, \beta^*_{u-1,t;k}$. For the power traces from the DPA contest v2 the

difference method turned out to be more efficient (higher rate of convergence), which should be due to the fact that $\beta^*_{0,t;k}$ clearly dominates the other coefficients. Note that in the first profiling step it is a (reasonable) option to use differences of power traces while it is unavoidable in the second profiling step and in the attack phase.

Estimation of the Distribution of $\mathbf{R_t}$ *and PCA.* Since the offsets $\tau_{t;l}$ are unknown, we apply OTM. In fact, since $\tau_{t;l} - \tau_{t;l+1} \approx 0$ and

$$\begin{aligned}&(\mathbf{I_t}(x_l,k) - \mathbf{I_t}(x_{l+1},k)) - (\mathbf{h}^{*\circ}_{\mathbf{t}}(x_l,k) - \mathbf{h}^{*\circ}_{\mathbf{t}}(x_{l+1},k)) \approx \\ &\mathbf{I_t}(x_l,k) - \mathbf{h_t}(x_l,k) - \tau_{t;l} - (\mathbf{I_t}(x_{l+1},k) - \mathbf{h_t}(x_{l+1},k) - \tau_{t;l+1}) \sim \mathcal{N}(0,2C).\end{aligned} \tag{16}$$

Consequently, we go for an estimate of $2C$ instead of C. Now let $\mathbf{w}_{\mathbf{t},l;k} := \mathbf{i_t}(x_l,k) - \widetilde{\mathbf{h}}^{*\circ}_{\mathbf{t}}(x_l,k)$. Then

$$\begin{aligned}\widetilde{2C} &:= \frac{1}{N_2-1}\widetilde{M}^{\circ T}\widetilde{M}^{\circ} \quad \text{with the } (m \times (N_2-1))\text{-matrix} \\ \widetilde{M}^{\circ T} &:= (\mathbf{w}_{\mathbf{t},1;k} - \mathbf{w}_{\mathbf{t},2;k}, \ldots, \mathbf{w}_{\mathbf{t},N_2-1;k} - \mathbf{w}_{\mathbf{t},N_2;k})\end{aligned} \tag{17}$$

provides an estimate for $2C$. We point out that the columns of M° are not independent. However, let M°_{ev} and M°_{odd} denote the submatrices of M°, which consist of the columns with even indices or of the columns with odd indices, respectively. For odd N_2

$$\frac{1}{N_2-1}\widetilde{M}^{\circ T}\widetilde{M}^{\circ} = \frac{1}{2}\left(\frac{2}{N_2-1}\widetilde{M}^{\circ T}_{\text{ev}}\widetilde{M}^{\circ}_{\text{ev}} + \frac{2}{N_2-1}\widetilde{M}^{\circ T}_{\text{odd}}\widetilde{M}^{\circ}_{\text{odd}}\right). \tag{18}$$

Both submatrices have independent columns, which yield estimates for $2C$ (analogously to the SA case). We point out that $\widetilde{M}^\circ_{\text{ev}}$ and $\widetilde{M}^\circ_{\text{odd}}$ are only weakly correlated since the l^{th} row of $\widetilde{M}^\circ$ is only correlated to rows $(l-1)$ and l. The matrices C and $2C$ have the same eigenspaces and thus the same transformation matrix P_s (cf. Eq. (9)). Applying the singular value decomposition to $\widetilde{M}^\circ$ yields P_s as well as estimates $\widetilde{2D}_s = 2\widetilde{D}_s$ and $\widetilde{D}_s$ for $2D_s$ and D_s, respectively.

SA-OTM: Attack Phase. We assume that the attacker has recorded N_3 measurement vectors $\mathbf{i_t}(x_1,k^\dagger),\ldots,\mathbf{i_t}(x_{N_3},k^\dagger)$ from a target device with a secret (unknown) subkey $k^\dagger$. As for SA the attacker applies a maximum likelihood estimation rule but for SA-OTM the situation becomes more complicated (Theorem 1).

Notation 2. If $F_1,\ldots,F_r$ are matrices with the same number of columns then $RV(F_1,\ldots,F_r)$ denotes the block matrix whose first rows are given by F_1, the next rows by F_2 etc.

Theorem 1. *For $l = 1, \ldots, N_3$ let $\mathbf{W}_{\mathbf{t},l;k} := \mathbf{I}_{\mathbf{t}}(x_l, k) - \mathbf{h}^{*\circ}_{\mathbf{t},k}(x_l, k)$. Then the $s(N_3 - 1)$-dimensional random vector*

$$\begin{aligned}\mathbf{W}_{\mathbf{t};k} &:= RV\big(P_s^T(\mathbf{W}_{\mathbf{t},1;k} - \mathbf{W}_{\mathbf{t},2;k}), P_s^T(\mathbf{W}_{\mathbf{t},2;k} - \mathbf{W}_{\mathbf{t},3;k}), \ldots, P_s^T(\mathbf{W}_{\mathbf{t},N_3-1;k} - \mathbf{W}_{\mathbf{t},N_3;k})\big)\\ &\sim N(RV(P_s^T(\tau_{\mathbf{t};1} - \tau_{\mathbf{t};2}), P_s^T(\tau_{\mathbf{t};2} - \tau_{\mathbf{t};3}), \ldots, P_s^T(\tau_{\mathbf{t};N_3-1} - \tau_{\mathbf{t};N_3})), G(D_s))\\ &\approx N(0, G(D_s)) \qquad (19)\end{aligned}$$

with the $(s(N_3 - 1) \times s(N_3 - 1))$-dimensional block tridiagonal matrix

$$G(D_s) = \begin{pmatrix} 2D_s & -D_s & & & \\ -D_s & 2D_s & -D_s & & \\ & -D_s & 2D_s & -D_s & \\ & & \ddots & \ddots & \ddots \\ & 0 & & -D_s & 2D_s & -D_s \\ & & & & -D_s & 2D_s \end{pmatrix}.$$

Proof. Since the random vectors $\mathbf{W}_{t,1;k}, \ldots, \mathbf{W}_{t,N_3;k}$ are independent the random vector

$$\mathbf{V} := RV(P_s^T(\mathbf{W}_{t,1;k}), P_s^T(\mathbf{W}_{t,2;k}), \ldots, P_s^T(\mathbf{W}_{t,N_3;k}))$$

is $\mathcal{N}\left(RV(P_s^T(\tau_{t;1}), P_s^T(\tau_{t;2}), \ldots, P_s^T(\tau_{t;N_3})), \hat{D}\right)$-distributed where $\hat{D}$ stands for the $(sN_3 \times sN_3)$-dimensional block diagonal matrix whose N_3 diagonal blocks equal D_s. We conclude $\mathbf{W}_{t;k} = L(V)$ where $L\colon \mathbb{R}^{sN_3} \to \mathbb{R}^{s(N_3-1)}$ denotes the linear mapping

$$L(RV(z_1, \ldots, z_{N_3})) := RV(z_1 - z_2, z_2 - z_3, \ldots, z_{N_3-1} - z_{N_3}).$$

By [9] (3.31) we have

$$L(\mathbf{V}) \sim \mathcal{N}(RV(P_s^T(\tau_{\mathbf{t};j} - \tau_{\mathbf{t};2}), P_s^T(\tau_{\mathbf{t};2} - \tau_{\mathbf{t};3}), \ldots, P_s^T(\tau_{\mathbf{t};N_3-1} - \tau_{\mathbf{t};N_3}), L\hat{D}L^T).$$

A careful computation verifies $L\hat{D}L^T = G(D_s)$, which proves the first assertion of Theorem 1. Since L is linear the second assertion follows from the assumption that the differences $\tau_{\mathbf{t};1} - \tau_{\mathbf{t};2}, \tau_{\mathbf{t};2} - \tau_{\mathbf{t};3}, \ldots, \tau_{\mathbf{t};N_3-1} - \tau_{\mathbf{t};N_3} \approx 0$. □

If the vector space $\mathcal{F}^{\circ}_{u,t;k}$ catches the relevant parts of the leakage then $\mathbf{h}^{*\circ}_{\mathbf{t};k}(x_l, k) - \mathbf{h}^{*\circ}_{\mathbf{t};k}(x_{l+1}, k) \approx \mathbf{h}^{*}_{\mathbf{t};k}(x_l, k) - \tau_{\mathbf{t};l} - \mathbf{h}^{*}_{\mathbf{t};k}(x_{l+1}, k) + \tau_{\mathbf{t};l+1}$, which motivates the following maximum likelihood decision rule. The adversary decides for the subkey $k \in \{0,1\}^s$, which maximizes $f_{G(\widetilde{D}_s)}$, or equivalently minimizes

$$({\mathbf{w}'_{\mathbf{t};k}}^T G(\widetilde{D}_s)^{-1} \mathbf{w}'_{\mathbf{t};k}) \text{ with} \qquad (20)$$

$$\mathbf{w}'_{\mathbf{t};k} := RV\Big(P_s^T(\mathbf{w}'_{\mathbf{t},1;k} - \mathbf{w}'_{\mathbf{t},2;k}), P_s^T(\mathbf{w}'_{\mathbf{t},2;k} - \mathbf{w}'_{\mathbf{t},3;k}), \ldots, P_s^T(\mathbf{w}'_{\mathbf{t},N_3-1;k} - \mathbf{w}'_{\mathbf{t},N_3;k})\Big)$$

and $\mathbf{w}'_{\mathbf{t},l;k} := \mathbf{i}_{\mathbf{t}}(x_l, k^{\dagger}) - \widetilde{\mathbf{h}}^{*\circ}_{\mathbf{t},k}(x_l, k)$ while $G(\widetilde{D}_s)$ is the estimate of $G(D_s)$.

Remark 1. Eq. (20) can be evaluated without inverting $G(\widetilde{D}_s)^{-1}$. Instead, one first solves the matrix-vector equation $G(\widetilde{D}_s)\mathbf{v} = \mathbf{w}'_{\mathbf{t};k}$ first, for which efficient numerical algorithms exist (e.g., iterative Krylov Methods [20]). Finally, one computes ${\mathbf{w}'}^T_{\mathbf{t};k}\mathbf{v}$. We point out that also these calculations could be saved by cancelling every second component in $\mathbf{w}'_{\mathbf{t};k}$ (at cost of doubling the number of attack traces!). As a compromise between efficiency and computational workload one might cancel every α^{th} component of $\mathbf{w}'_{\mathbf{t};k}$, which results a block diagonal matrix with $\frac{N_3}{\alpha}$ matrices G_l for $1 \leq l \leq \frac{N_3}{\alpha}$ and $\dim(G_l) \leq s \cdot (\alpha - 1) \ll \dim G(\widetilde{D}_s)$ in its diagonal. Here one 'wastes' $\frac{N_3}{\alpha}$ power traces for the sake of faster calculation. This method is of particular interest in context of the DPA contest v2 since the contest rules demand the continued evaluation of an attack for increasing sets of power traces. In our experiments (Sect. 6) we used $\alpha = 200$, without claiming that the choice of α is optimal.

5 On the Selection of Stochastic Leakage Models

The approximator of the leakage function $h_{t;k}$ (i.e., $h^*_{t;k}$ or $h^{*\circ}_{t;k}$) is close only for an appropriate subspace $\mathcal{F}_{u,t;k}$. The appropriateness depends on the leakage model and thus on the concrete subspace. In this section we consider different subspaces that may be used for attacks on the last round of an AES-128 hardware implementation. In [10] a 9-dimensional subspace $\mathcal{F}_{9,t;k}$ was investigated in detail (SA). The selection of $\mathcal{F}_{9,t;k}$ is reasonable if (one assumes that) the side-channel leakage is only caused by the sum of the individual transitions on all bit lines. High-dimensional subspaces also capture effects that arise from interactions between the transitions on two or more bit lines. Such effects occur due to properties of internal circuit structures, e.g., propagation glitches or cross-talk phenomena during the metastable phase of the registers, which is a well-known problem in CMOS VLSI Circuit Design [12, 5]. In Sect. 6 we consider $\mathcal{F}^{\circ}_{u,t;k}$ for $u \in \{9, 37, 93, 163, 219, 247, 255, 256\}$. Recall that $\dim \mathcal{F}^{\circ}_{u,t;k} = u - 1$. Of course, also high-dimensional subspaces keep the regression linear.

The possibility of applying high-dimensional subspaces was already pointed out in [17, 16]. In [13], Renauld et al. analyzed the information theoretic impact of high-dimensional leakage models on the mutual information. To the best of the authors' knowledge very high-dimensional subspaces have not been evaluated in concrete attacks yet.

5.1 High-Dimensional Subspaces for SA-OTM

With regard to an ordinary hamming distance model we first consider the 8-dimensional subspace $\mathcal{F}^{\circ}_{9,t;k}$ which exploits the corresponding intermediate value of the 9^{th} round XORed with the 10^{th} round key. More precisely, we select the following basis vectors

$$g_{j,t;k_{(y)}}((x_{(z)},x_{(y)}),k_{(y)}) = (\underbrace{(x_{(z)} \oplus S^{-1}(x_{(y)} \oplus k_{(y)}))_j}_{:=(\phi(x_{(z)},x_{(y)},k_{(y)}))_j := \hat{g}_{j,t;k}} - 2^{-1}) \qquad (21)$$

for $j = 1, \ldots, 8$.

The subtrahend 2^{-1} ensures $E_X(g_{j,t,k_{(}y)}(X, k_{(y)})) = 0$ for independent and uniformly distributed random variables $X_{(y)}$ and $X_{(z)}$, a reasonable model for two ciphertext bytes. The indices y and z are chosen according to the distance model of the AES design (cf. [10,8]).

Moreover, we consider high-dimensional subspaces. To simplify we introduce new notation. First, $\mathcal{B}_1 := \{g_{1,t;k_{(y)}}, \ldots, g_{8,t;k_{(y)}}\}$ collects all basis vectors from Eq. (21), which capture the contribution of the individual bit lines. Moreover, for $2 \leq i \leq 8$ the set

$$\mathcal{B}_i := \{\hat{g}_{j_1,t;k_{(y)}} \cdots \hat{g}_{j_i,t;k_{(y)}} - 2^{-i} \mid 1 \leq j_1 < \ldots < j_i \leq 8\} \qquad (22)$$

contains all unordered i-fold products of elements in $\mathcal{B}_1$ minus 2^{-i}. (A typical element in $\mathcal{B}_2$ is $\hat{g}_{4,t;k_{(y)}} \cdot \hat{g}_{7,t;k_{(y)}} - 2^{-2}$.) The subtrahend 2^{-i} ensures the zero-mean property for all elements of $\mathcal{B}_i$. Table 1 provides the basis vectors for all relevant subspaces (the elements of the sets in the second column).

Table 1. Set of basis functions for each subspace

$\dim(\mathcal{F}^{\circ}_{u,t;k})$ $(= u-1)$	Set of basis functions
8	$\mathcal{B}_1$
36	$\mathcal{B}_1 \cup \mathcal{B}_2$
92	$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$
162	$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4$
218	$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5$
246	$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5 \cup \mathcal{B}_6$
254	$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5 \cup \mathcal{B}_6 \cup \mathcal{B}_7$
255	$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5 \cup \mathcal{B}_6 \cup \mathcal{B}_7 \cup \mathcal{B}_8$

5.2 Leakage Models for the Stochastic Approach

As pointed out in Subsect. 4.2 the subspaces for SA-OTM are similar to the subspaces for SA, just the first basis vector $g_{0,t;k_{(y)}}$ is omitted. In particular, the subspace $\mathcal{F}_{9,t;k}$ is spanned by

$$g_{0,t;k_{(y)}}((x_{(z)},x_{(y)}),k_{(y)}) = 1 \qquad (23)$$
$$g_{j,t;k_{(y)}}((x_{(z)},x_{(y)}),k_{(y)}) = ((x_{(z)} \oplus S^{-1}(x_{(y)} \oplus k_{(y)}))_j - 2^{-1}) \quad \text{for } j = 1, \ldots, 8\,.$$

We define $\mathcal{B}_0 := \{g_{0,t;k_{(y)}}\}$, and the construction of high-dimensional subspaces follows analogously to Tab. 1 with additional basis vector $\mathcal{B}_0$.

5.3 Symmetry

References [10,16] consider leakage models with symmetries (for SA). In fact, the basis vectors $g_{j;t;k_{(y)}}$ from Subsect. 5.1 and 5.2 can be expressed by a composition of a key-independent function $\overline{g}_{j,t}\colon \{0,1\}^8 \to \mathbb{R}$ with the mapping $\phi\colon \{0,1\}^8 \times \{0,1\}^8 \times \{0,1\}^8 \to \{0,1\}^8$, given by

$$\phi(x_{(z)}, x_{(y)}, k_{(y)}) := (x_{(z)} \oplus S^{-1}(x_{(y)} \oplus k_{(y)}))\,.$$

This essentially reduces the argument of $g_{j;t;k_{(y)}}$ from 24 to 8 bits. If the leakage function $h_{t,k_{(y)}}((x_{(z)}, x_{(y)}), k_{(y)})$ also depends on its arguments only through $\phi(x_{(z)}, x_{(y)}, k_{(y)})$ one can compute $h^*_{t,k'_{(y)}}((\cdot,\cdot), k'_{(y)})$ for each $k'_{(y)}$ if $h^*_{t,k_{(y)}}((\cdot,\cdot), k_{(y)})$ is known for arbitrary subkey $k_{(y)}$. In particular, for uniformly distributed $(X_{(y)}, X_{(z)})$ for each $j < u$

$$\beta_{j,t;k'_{(y)}} \equiv \beta_{j,t} \quad \text{for all } k'_{(y)} \in \{0,1\}^8. \tag{24}$$

Reference [8] explains how to verify, resp. to falsify, whether symmetry assumptions are indeed valid, and a symmetry metric is introduced. Eq. (24) says that the coefficients $\beta_{j,t,\cdot}$ are identical for all admissible subkeys. This property allows to use all 1000.000 power traces of the template base (though belonging to different (sub)keys) jointly in a single least square estimation process. This gives more stable results, and (for each key byte) profiling step 1 has to be carried out only once.

6 Experimental Analysis

In this section we compare the efficiency of SA-OTM and of SA on basis of the DPA contest v2 power traces. We apply the leakage models from Sect. 5. The DPA contest v2 provides two data bases: A template base with 1.000.000 power traces (to develop an attack), and a public base, which contains power traces for 32 fixed keys, 20.000 traces for each key (to test the attack). The organizers of the contest evaluated the submitted attacks on a (non-public) private base to avoid "biased" attacks. The measurements were recorded on the SASEBO-G II FPGA-evaluation board [14] using a Virtex 5 FPGA [19]. Each encryption (AES with 128 bit keys) takes 10 clock cycles, and the SBOX realization is based on a composite field [15]. In analogy to the DPA contest v2 we calculate the partial success rate (PSR) and the global success rate (GSR) to compare the efficiency of the particular attacks. The PSR is the probability that the correct subkey is ranked first among all possible subkeys, while GSR denotes the probability that the complete key is ranked first. We are mainly interested in the minimum number of power traces for which the PSR is stable above 80% (i.e. the 'worst' byte is stable in $> 80\%$ of the experiments), and in the minimum number of power traces for which the GSR is stable above 80% ($\rightarrow$ evaluation criteria for the DPA contest v2).

As already mentioned the template base consists of 1.000.000 traces, i.e. for each subkey $\approx$ 4.000 traces. This number is too small for a sufficiently precise estimation of the β-characteristic for each key. However, due to the symmetry properties of the attacked implementation (cf. Subsect. 5.3) we could circumvent this problem. Accordingly, we computed the coefficients $\widetilde{\beta}^*_{j,t}$ (cf. Eq. (24)) on basis of all 1.000.000 power traces.

The application of PCA to the Covariance matrices $\widetilde{C}$ and $2\widetilde{C}$ showed that the first eigenvalue $\widetilde{\lambda}_1$ is at least 20 times larger than the other eigenvalues $\widetilde{\lambda}_2, \ldots, \widetilde{\lambda}_m$. Consequently, we selected $s = 1$, and hence P_s is an $(m \times 1)$ matrix (cf. Eq. (9)). For the evaluation of the PSR and the GSR we used the power

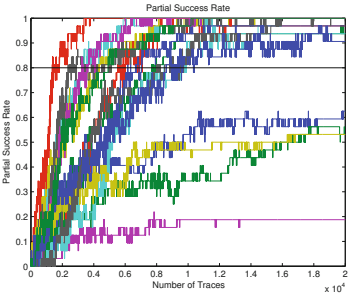

Fig. 4. Partial success rate: SA with a 9-dim. model

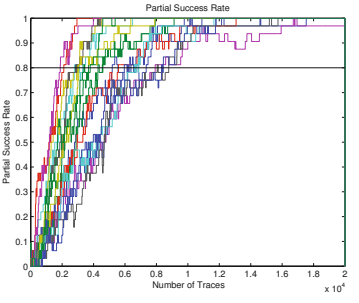

Fig. 5. Partial success rate: SA-OTM with a 8-dim. model

traces from the public base. Figure 4 depicts the PSR for the SA with the 9-dimensional leakage model from Eq. (23). Each curve corresponds to one of the 16 subkeys. Figure 5 shows the PSR for SA-OTM with the 8-dimensional leakage model (e.g., Eq. (21)). All bytes achieve the 80% threshold, and except for one subkey, even the 100% threshold. Figure 6 and 7 depict the PSR for SA and

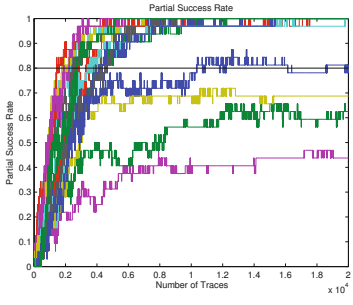

Fig. 6. Partial success rate: SA with a 37-dim. model

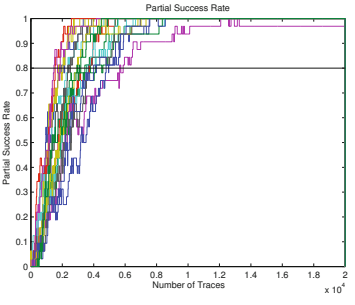

Fig. 7. Partial success rate: SA-OTM with a 36-dim. model

SA-OTM using the 37-dimensional and 36-dimensional model, which capture the individual leakage of each bit line *and* the leakage caused by the interaction between two arbitrary bit lines. Evidently, these leakage models describe the existing leakage more precisely. However, for SA the PSR criterion fails due to the same 4 bytes. Compared to the 8-dimensional leakage model for SA-OTM the minimum number of traces with stable PSR $> 80\%$ drops down from 8781 to 5876. The 93- and the 92-dimensional leakage model additionally capture the

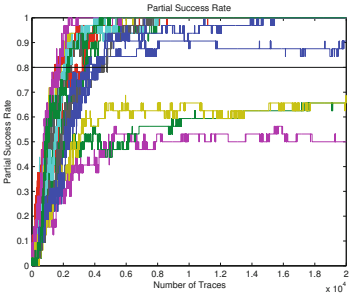

Fig. 8. Partial success rate: SA with a 93-dim. model

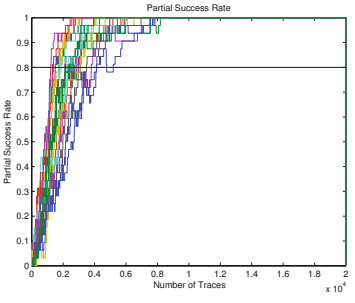

Fig. 9. Partial success rate: SA-OTM with a 92-dim. model

leakage that arises from the combinations of three bit lines. Experimental results are depicted in Figure 8 and Figure 9. The 93-dimensional model (SA) improves the PSR, but 3 bytes still do not reach 80% PSR. SA-OTM reaches the PSR stable above 80% after 5195 traces. The significant improvement of the PSR for specific bytes does not necessarily imply that the drifting offset only influences those bytes. It rather underlines that not all subkey bytes 'leak' in the same way. One might conjecture that those subkey bytes, which have less influence on the overall leakage are more affected by the drifting offset than the others. Figure 10 shows the GSR. SA-OTM requires about 6734 traces to achieve a GSR stable $> 80\%$. A GSR of 100% is only archived for SA-OTM with the 92-dimensional leakage model.

The best attack that was submitted during the contest achieves a PSR stable above 80% for 5.890 traces and a stable GSR $> 80\%$ for 7.061 traces. SA-OTM with the 92-dimensional model outperforms these benchmarks. Moreover, we computed the success rates of SA-OTM for the 162-, 218, 246-, 254-, 255- dimensional model, which increased this success rate further. These results indicate that the 218-dimensional subspace $\mathcal{F}^{\circ}_{219,t;k}$, which is spanned by the basis vectors in $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5$, seems to essentially capture the leakage. Tab. 2 contains all results for the public base and the private base (as far as known). For the private base a third contest criterion, the maximal partial guessing entropy below 10 (max PGE below 10) [18], is listed.

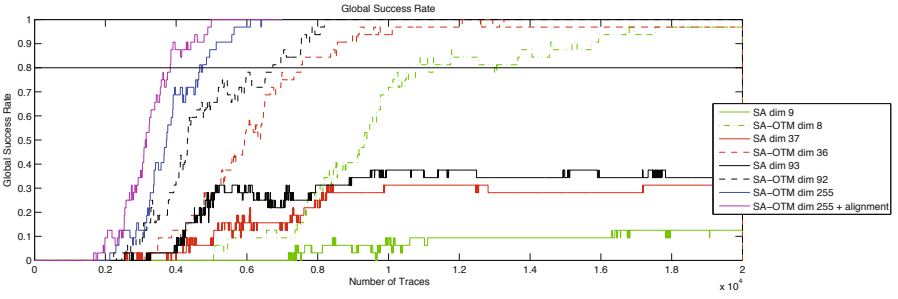

Fig. 10. Global success rate of SA and SA-OTM for different models

Remark 2. Following the expositions in Sect. 2 one might simply try to combine 'normal' SA with vertical trace alignment. However, as explained below also in combination with vertical alignment SA-OTM remains more efficient than SA.

Vertical Trace Alignment. We combined SA and SA-OTM with vertical trace alignment, a well-known method in power analysis. We 'normalized' each measurement trace to mean zero (over the whole trace), i.e. we computed *aligned* $i_{\mathbf{t}}(x_l, k) = i_{\mathbf{t}}(x_l, k) - \text{mean}(i_{\mathbf{t}}(x_l, k))$, with $\mathbf{t}$ ranging over the complete trace. However, even then SA with the 93-dimensional model did not exceed PSR $> 80\%$ for all bytes.

Since SA-OTM itself solves the problem with the drifting offset our goal here was to reduce the impact of outliers. Apart from the drifting offset, Figure 1 displays a few extreme values that might be caused by such outliers, which result from measurement errors or any other interference during the acquisition.

Table 2. Success rate (PSR stable $> 80\%$ / GSR stable $> 80\%$), and max PGE below 10 for the private base

Attack	Public Base	Private Base[1]
SA-OTM: dim 8	(8781 / 13020)	unknown
SA-OTM: dim 36	(5876 / 7533)	unknown
SA-OTM: dim 92	(5195 / 6734)	(4358 / 5571), 1.894
SA-OTM: dim 162	(4353 / 6144)	unknown
SA-OTM: dim 218	(3552 / 4564)	unknown
SA-OTM: dim 246	(3769 / 4691)	unknown
SA-OTM: dim 254	(3720 / 4740)	unknown
SA-OTM: dim 255	(3718 / 4748)	unknown
SA-OTM: dim 255 incl. alignment	(2682 / 3836)	(2748 / 3589), 1.356
Best submitted attack during the first & second period	unknown	(5890 / 7061), 2.767

[1] See `http://www.dpacontest.org/v2/hall_of_fame.php` for the results on the private base.

Alternatively, one could also try to identify and omit the outliers. SA-OTM with the 255-dimensional leakage model and vertical alignment SA-OTM achieves a PSR stable $> 80\%$ within 2748 traces and a GSR stable above 80% within 3589 traces (private base). These results reduce the required number of traces to 50% compared to the best submitted attack during the contest, cf. Tab. 2.

Further Work / Open Problems Our analysis raises several questions. Can the drifting offset be effectively prevented in practice? What is the smallest subspace that captures all relevant parts of the compromising leakage? Do different types of implementations demand different subspaces? Another ambitious topic for future work could be an automatized search for optimal (high-dimensional) subspaces, which finally might yield to appropriate basis vector selection methods.

7 Conclusion

In this contribution we investigated two fundamental problems that may affect the efficiency of profiling based attacks, and we developed efficient solutions. Drifting offsets (caused by temperature variations) cause difficulties for attacks, which consider (implicitly or explicitly) the average power consumption (typically profiling based attacks). We introduced a new method, denoted as the Offset Tolerant Method (OTM), which considers differences of consecutive pairs of power traces. We adjusted OTM to the stochastic approach (SA), abbreviated by SA-OTM. In presence of a drifting offset SA-OTM turned out to be clearly more efficient than SA, even in combination with vertical trace alignment.

We further addressed the problem of how to select suitable leakage models, which shall represent the compromising leakage as precise as possible. Our results show that leakage may also arise from the interaction of several bit lines. This effect can only be captured by high-dimensional leakage models. Combining these two improvements we achieved the best results of all participants of the DPA contest v2. Further research work might consider open problems formulated at the end of Sect. 6 or concentrate on improvements of SA-OTM, maybe in combination with alternative dimension reduction techniques.

Acknowledgment. The work presented in this contribution was supported by the German Federal Ministry of Education and Research (BMBF) in the project *Resist* through grant number 01IS10027A. We thank Christian Brandt for the heating cabinet framework.

References

1. Archambeau, C., Peeters, E., Standaert, F.X., Quisquater, J.J.: Template Attacks in Principal Subspaces. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 1–14. Springer, Heidelberg (2006)

2. Brier, E., Clavier, C., Olivier, F.: Correlation Power Analysis with a Leakage Model. In: Joye, M., Quisquater, J.J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 16–29. Springer, Heidelberg (2004)
3. Chari, S., Rao, J.R., Rohatgi, P.: Template Attacks. In: Kaliski Jr., B.S., Koç, C., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 13–28. Springer, Heidelberg (2003)
4. DPA contest v2, http://www.dpacontest.org/
5. Eo, Y., Eisenstadt, W., Jeong, J.Y., Kwon, O.K.: A new on-chip interconnect crosstalk model and experimental verification for CMOS VLSI circuit design. IEEE Transactions on Electron Devices 47(1), 129–140 (2000)
6. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual Information Analysis. In: Oswald, E., Rohatgi, P. (eds.) CHES 2008. LNCS, vol. 5154, pp. 426–442. Springer, Heidelberg (2008)
7. Gierlichs, B., Lemke-Rust, K., Paar, C.: Templates vs. Stochastic methods. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 15–29. Springer, Heidelberg (2006)
8. Heuser, A., Kasper, M., Schinder, W., Stöttinger, M.: How a Symmetry Metric Assists Side-Channel Evaluation - A Novel Model Verification Method for Power Analysis. In: 14th Euromicro Conference on Digital System Design Architectures, Methods and Tools (DSD 2011). IEEE (2011)
9. Kardaun, O.: Classical Methods of Statistics. Springer, Heidelberg (2005)
10. Kasper, M., Schindler, W., Stöttinger, M.: A Stochastic Method for Security Evaluation of Cryptographic FPGA Implementations. In: FPT 2010, pp. 146–154. IEEE Press (2010)
11. Kocher, P.C., Jaffe, J., Jun, B.: Differential Power Analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999)
12. Nieuwland, A.K., Katoch, A., Meijer, M.: Reducing Cross-Talk Induced Power Consumption and Delay. In: Macii, E., Koufopavlou, O.G., Paliouras, V. (eds.) PATMOS 2004. LNCS, vol. 3254, pp. 179–188. Springer, Heidelberg (2004)
13. Renauld, M., Standaert, F.X., Veyrat-Charvillion, N., Kamel, D., Flandre, D.: A Formal Study of Power Variability Issues and Side-Channel Attacks for Nanoscale Devices. In: Paterson, K.G. (ed.) EUROCRYPT 2011. LNCS, vol. 6632, pp. 109–128. Springer, Heidelberg (2011)
14. SASEBO GII, http://www.rcis.aist.go.jp/special/SASEBO/SASEBO-GII-en.html
15. Satoh, A., Morioka, S., Takano, K., Munetoh, S.: A Compact Rijndael Hardware Architecture with S-Box Optimization. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 239–254. Springer, Heidelberg (2001)
16. Schindler, W.: Advanced Stochastic Methods in Side Channel Analysis on Block Ciphers in the Presence of Masking. Math. Crypt. 2, 291–310 (2008)
17. Schindler, W., Lemke, K., Paar, C.: A Stochastic Model for Differential Side Channel Cryptanalysis. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 30–46. Springer, Heidelberg (2005)
18. Standaert, F.X., Malkin, T., Yung, M.: A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 443–461. Springer, Heidelberg (2009)
19. Virtex-5 FPGA User Guide (2010)
20. Vorst, H.A.V.D.: Iterative Krylov Methods for Large Linear Systems. Cambridge University Press, Cambridge (2003)