SNR to Success Rate: Reaching the Limit of Non-Profiling DPA

Suvadeep Hajra
Dept. of Computer Science & Engg.
Indian Institute of Technology, Kharagpur, India
suvadeep.hajra@gmail.com

Debdeep Mukhopadhyay
Dept. of Computer Science & Engg.
Indian Institute of Technology, Kharagpur, India.
debdeep.mukhopadhyay@gmail.com

***Abstract*—Profiling power attacks like Template attack and Stochastic attack optimizes their performance by jointly evaluating the leakages of multiple sample points. However, such multivariate approaches are rare among non-profiling DPA attacks, since integration of the leakage of a higher Signal-to-Noise Ratio (SNR) sample point with the leakages of lower SNR sample points might result in a decrease in the overall performance. We study the issue of optimally combining the leakages of multiple sample points using a linear function in great details. In this work, our contributions are three-fold: 1) we first derive a relation between the success rate of a CPA attack and the SNR of the power traces, 2) we introduce a multivariate leakage model for Virtex-5 FPGA device, and 3) using the proposed multivariate leakage model, we derive the linear Finite Impulse Response (FIR) filter coefficients which maximizes the SNR of the output leakage, thus optimizes the success rate of the CPA attacks in a non-profiling setup.**

I. Introduction

Differential Power Analysis (DPA) [20] has been proven to be an extremely lethal tool for side-channel analysis. It is highly effective in finding the secret key of a secure device by analysing the power traces of the device, even without knowing the implementation details. One of its strengths comes from its ability to exploit minute data-dependency of leakage by accumulating them over a large number of power traces. Since power traces are the scarce resource, to reduce the number of required power traces for a successful DPA attack, or to increase the success rate of a DPA attack using a limited number of power traces has been in the focus of DPA literature since its introduction.

The success rate of the DPA [23] attacks is largely influenced by the Signal-to-Noise Ratio (SNR) [23] of the power traces. As a consequence, in many applications, Power Analysis attacks are either preceded by various pre-processing techniques like integration [23], PCA [6] for the reduction of noise in the power traces or followed by some post-processing techniques like averaging [9], [5], [6], filtering [24] for the reduction of the effect of noise on the outputs of the distinguisher. These techniques attempt to improve the performance of the DPA attacks directly or indirectly by extracting information from multiple sample points. Some of these techniques like PCA are based on some implicit assumptions, thus optimally applicable to some specific scenarios only, while others deploy some heuristic methods (please refer to Sec. II-D). In this article, we take a model based approach to increase the effectiveness of DPA attacks by combining the leakages of multiple sample points.

Various profiling attacks like Template attack [8] and Stochastic attack [28] provide optimal performance by jointly evaluating the leakages at multiple sample points. However, they use a separate profiling step for approximating the multivariate leakage distribution [30] of the power traces. The profiling step requires a large number of power traces to estimate the multivariate leakage distribution with sufficient accuracy. Moreover in most of the cases, it needs the knowledge of the secret key which may not be available in many attacking scenarios.

Principal Component Analysis (PCA) has been introduced as a tool to reduce the size of the sample points in Template attacks [4]. Later in [29], PCA is used as a distinguisher. Recently in [6], Batina et al. have presented it as a preprocessing tool for reducing noise in a non-profiling setup. However, it performs better under the assumption that the data-dependent variations is larger than the noise variations. Unfortunately, in side-channel analysis, this assumption does not hold always. Though in [6], Batina et al. have proposed a new distinguisher based on some empirical observation, the performance of such distinguisher is far from being optimal.

Contributions: In this paper, we have studied how to maximize the success rate of a DPA attack by combining the leakages of multiple sample points. We have explored two possible ways of combining: a) combine the leakages of multiple sample points first and then apply a univariate distinguisher on the combined leakage, and b) apply a univariate distinguisher on multiple sample points independently and then combine their outputs. We have further shown that in certain cases both the approaches are equivalent in terms of the success rate of the attack. Next, have devised an optimal way of combining the leakages of multiple sample points using the following three steps:

1) We derive an exact relation between the SNR of the power traces and the success rate of a CPA attack which is the strongest form of DPA in some applications. Thanks to the relation, maximization of the success rate by combining leakages of multiple sample points becomes equivalent to the maximization of the SNR of the combined leakage.
2) We introduce a multivariate leakage model by extending

the conventional leakage model for multiple sample points for Virtex-5 FPGA device. The proposed multivariate leakage model enables us to determine the power of the data dependant signal of a sample point without knowing the correct key.
3) We derive a linear FIR filter which, when applied to the power traces, maximizes the SNR of its output. The derivation does not require the knowledge of the secret key, thus can be used in non-profiling DPA attacks. We also study how the derived linear FIR can be made more resistant to estimation error and computationally more efficient.

We have also experimentally verified our methods on various noisy scenarios.

Rest of the paper is organized in this way: Section II describes the background of DPA along with the necessary notations used in the work. In Section III, a relation between the success rate of CPA and the SNR of the power traces has been derived. Section IV has extended the conventional leakage model over multiple sample points which results into a multivariate leakage model. In Section V, an expression has been derived to compute the coefficients of the linear FIR filter which optimizes the SNR of its output. The optimum filter has been further approximated for making it more resistant to estimation error and computationally more efficient. Section VI describes several new basis systems for applying the approximated optimum filter to make the approximation more accurate. In Section VII, the improvements in the performance of CPA using the proposed filtering techniques have been experimentally verified for various scenarios. Section VIII verifies the optimality of the proposed pre-processing techniques. Finally, conclusion has been drawn in Section IX.

II. Preliminaries

A. Notations

For the rest of the paper, we will use a calligraphic letter like $\mathcal{X}$ to denote a finite set. The corresponding capital and small letter, X and x, are used to denote a random variable over the set and a particular element of it respectively. $E[X]$, σ_X and $Var(X)$ are used to denote mean, standard deviation and variance of the random variable X respectively. We also denote by $Cov(X,Y)$ and $Corr(X,Y)$, the covariance and the Pearson's correlation coefficient between the random variables X and Y respectively. The vector $\{x_0,\cdots,x_k\}$ is denoted by $\{x_i\}_{0\leq i\leq k}$. Alternatively, it is also denoted by a letter in bold like $\mathbf{x}$. For convenience, sometimes we use μ_X to denote the mean of the random variable X. Gaussian distribution with mean m and standard deviation σ is represented by $\mathrm{N}(m,\sigma)$. $\mathbf{x}'$ denotes the Hermitian transpose of the vector or matrix $\mathbf{x}$.

B. Differential Power Analysis

We will mainly follow the formalisation of Differential Power Analysis by Standaert et al. in [30]. It is briefly described below.

Generally, the DPA attacks have two parts. In the first part, a Device Under Test (DUT) is under the control of the attacker. The attacker collects the leakage L_{t^*} at time instant t^* due to the manipulation of some intermediate key-dependent variable $S=F_{k^*}(X)$ by executing the DUT repeatedly, say q times, for q different inputs. S is commonly referred to as *target* and $F_{k^*}:\mathcal{X}\rightarrow\mathcal{S}$ be a function of a known part of the plaintext $x\in\mathcal{X}$. F_{k^*} is determined by both the algorithm and a small part of the secret key referred to as the subkey $k^*\in\mathcal{K}$. The leakage L_{t^*} satisfies

$$L_{t^*}=\tilde{\Psi}(S)+N \tag{1}$$

where the function $\tilde{\Psi}:\mathcal{S}\rightarrow\mathbb{R}$ maps the target S to the deterministic part of the leakage and $N\sim\mathrm{N}(\mu_N,\sigma_N)$ accounts for the independent Gaussian noise. At the end, the attacker collects q measurement curves $\mathbf{l}_{t^*}=\{l_{t^*}^0,\cdots,l_{t^*}^{q-1}\}$ corresponding to the execution of q plaintexts $\mathbf{x}=\{x_0,\cdots,x_{q-1}\}$.

In the second part, the attacker chooses a suitable prediction model $\Psi:\mathcal{S}\rightarrow\mathbb{R}$ and compute the predicted leakage represented by the random variable P_k using $P_k=\Psi(S_k)=\Psi(F_k(X))$ for each key hypothesis $k\in\mathcal{K}$. If Ψ is a good approximation for $\tilde{\Psi}$, the leakage L_{t^*} is strongly dependent on the correct predicted leakage P_{k^*}. However, since $F_{k^*}(X)$ and $F_k(X)$ are almost independent for $k^*\neq k$, L_{t^*} is independent of the prediction variable P_k. Then, a statistical tool D is used to detect this dependence between the actual leakage and the predicted leakage for the correct key. The theoretical distinguisher is given by $\mathbf{D}=\{d_k\}_{k\in\mathcal{K}}=\{\mathrm{D}(L_{t^*},P_k)\}_{k\in\mathcal{K}}=\{\mathrm{D}(\tilde{\Psi}(F_{k^*}(X))+N,\Psi(F_k(X)))\}_{k\in\mathcal{K}}$. The theoretical first order success rate ($1OSR$) [30] of the attack is given by $Pr(k^*=argmax_{k\in\mathcal{K}}d_k)$. However in practice, the random variables X, L_{t^*} and N are estimated by the vector $\mathbf{x}$, $\mathbf{l}_{t^*}$ and $\mathbf{n}=\{n_0,\cdots,n_{q-1}\}$ respectively. Thus, the practical distinguisher is given by $\hat{\mathbf{D}}=\{\hat{d}_k\}_{k\in\mathcal{K}}=\{\hat{\mathrm{D}}(\mathbf{l}_{t^*},\Psi(F_k(\mathbf{x})))\}_{k\in\mathcal{K}}=\{\hat{\mathrm{D}}(\tilde{\Psi}(F_{k^*}(\mathbf{x}))+\mathbf{n},\Psi(F_k(\mathbf{x})))\}_{k\in\mathcal{K}}$ and the practical $1OSR$ of the attack is given by $Pr(k^*=argmax_{k\in\mathcal{K}}\hat{d}_k)$.

C. Correlation Power Analysis with a model

When the hardware leakage behavior follows a well known leakage model like Hamming weight model or Hamming distance model, some known prediction model Ψ closely approximates $\tilde{\Psi}$ i.e. $\tilde{\Psi}(s)\approx a\cdot\Psi(s)$ holds for some real constant a and for all $s\in\mathcal{S}$. Then, Eq. 1 can be approximated [7] as

$$L_{t^*}=a\cdot\Psi(S)+N \tag{2}$$

Under the above equation, the relation between the actual leakage L_{t^*} and the predicted leakage for the correct key $P_{k^*}=\Psi(S)$ (since $S=S_{k^*}$) becomes linear. In Correlation Power Analysis (CPA) [7], Pearson's correlation is used to detect the linearity by computing

$$\begin{aligned}\rho_k&=Corr(\hat{L}_{t^*},\Psi(F_k(\hat{X})))\\&=Corr(\hat{L}_{t^*},\hat{P}_k)\\&=\frac{Cov(\hat{L}_{t^*},\hat{P}_k)}{\hat{\sigma}_{L_{t^*}}\hat{\sigma}_k}\end{aligned} \tag{3}$$

for all $k \in \mathcal{K}$ where P_k and σ_k denotes $\Psi(S_k)$ and σ_{P_k} respectively. Since, Pearson's correlation detects the linear relation between two variables, it performs better than other attacks like Mutual Information Analysis (MIA) [15], Difference of Mean (DoM) [20]. When the hardware leakage model is not sufficiently known, 'generic' attacks like MIA perform better than CPA. In the rest of the paper, we will consider only the scenarios where the hardware follows a well known leakage behavior.

D. Multivariate DPA

In practical attacks, multiple leakage samples at discrete time instants are collected during the encryptions or decryptions. As a result, the leakage $\mathbf{L}$ is a vector of T random variables $\{L_0, \cdots, L_{T-1}\}$ where L_t represents the leakage of time instant (sample point) t for $0 \leq t < T$. One snapshot of $\mathbf{L}$ denoted by $\mathbf{l} = \{l_0, \cdots, l_{T-1}\}$ is referred to as a trace or power trace. In that case, a univariate distinguisher is applied on each of the sample points independently and then the attacker chooses the best result among those.

While all the profiling attacks like Template attack [8] and Stochastic attack [28] optimizes their performance by considering the multivariate leakage distribution of the power traces, combining the leakages of multiple sample points is rare in non-profiling DPA. Though a few distinguishers like MIA can be extended as a multivariate distinguisher, most of them are not easily extendable for multivariate DPA. Even though they can be extended for multivariate DPA, they do not always improve the performance of the attacks. Instead, such multivariate approaches mainly exist in the forms of various pre-processing techniques like PCA [29], [6], integration [23] and filtering [24]. However, they are either heuristic in nature or based on some assumption. Moreover, to the best of the authors knowledge, there is no such techniques which optimally combines the leakages of multiple sample points. In the paper, we investigate the possibility of the combining leakages of multiple sample points in a way that optimizes the success rate.

As shown in Fig. 1, there are two alternative approaches to combine the leakages of multiple sample points: 1) combine the leakages of multiple sample points first using a function $g : \mathbb{R}^T \to \mathbb{R}$ and then apply a univariate distinguisher on the resultant leakage $g(\mathbf{L})$ (as shown in Fig. 1a), and 2) apply the univariate distinguisher D on all the sample points independently resulting in $|\mathcal{K}|$ vectors $\{d_k(t)\}_{0 \leq t < T}$ for each $k \in \mathcal{K}$ and then apply the function g to generate the final distinguisher $\{\tilde{d}_k\}_{k \in \mathcal{K}}$ having $\tilde{d}_k = g(\{d_k(t)\}_{0 \leq t < T})$ (as shown in Fig. 1b).

Interestingly, if we consider Pearson's correlation (as in CPA) as the univariate distinguisher and restrict the function g to the space of linear functions, then the above two approaches are equivalent. To see it, let us denote the Pearson's correlation at sample point t for key guess k by $\rho_k(t)$. Since, g is a $T \times 1$ linear mapping, $g(y_0, ..., y_{T-1})$ can be represented as an inner product of the vector $\{y_0, ..., y_{T-1}\}$ and the real coefficient vector $\{h_0, ..., h_{T-1}\}$. Hence, the output for the key guess k obtained in the second approach

$$\begin{aligned}
\tilde{d}_k &= g(\{\hat{d}_k(t)\}_{0 \leq t < T}) \\
&= \sum_{t=0}^{T-1} h_t \rho_k(t) \\
&= \sum_{t=0}^{T-1} \frac{h_t Cov(\hat{L}_t, \hat{P}_k)}{\hat{\sigma}_{L_t} \hat{\sigma}_k} \\
&= \sum_{t=0}^{T-1} \frac{Cov(h_t \hat{L}_t / \hat{\sigma}_{L_t}, \hat{P}_k)}{\hat{\sigma}_k} \\
&= \frac{Cov(\sum_{t=0}^{T-1} h_t \hat{L}_t / \hat{\sigma}_{L_t}, \hat{P}_k)}{\hat{\sigma}_k} \\
&= \frac{Cov(\tilde{g}(\hat{\mathbf{L}}), \hat{P}_k)}{\hat{\sigma}_k} \\
&= Corr(\tilde{g}(\hat{\mathbf{L}}), \hat{P}_k) \hat{\sigma}_{\tilde{g}(\mathbf{L})}
\end{aligned}$$

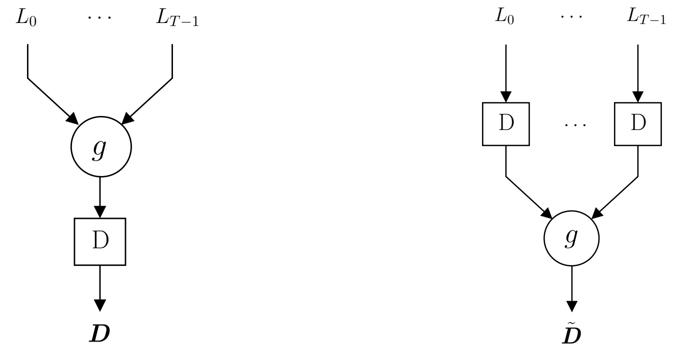
where $\tilde{g}$ be a $T \times 1$ linear mapping with coefficient vector $\{h_t / \hat{\sigma}_{L_t}\}_{t=0}^{T-1}$. Since $\hat{\sigma}_{\tilde{g}(\mathbf{L})}$ does not influence the success rate of a univariate distinguisher, for each linear function g in the second approach, there exist a linear function $\tilde{g}$ in the first approach which results in the same success rate. In other words, optimization of the success rate in the first approach also optimizes the success rate in the second approach and vice-versa. Since the first approach is computationally more efficient, we consider the first approach in the rest of the paper.

In the next section, we derive a relation between the success rate of CPA and the SNR of the power traces.

III. Deriving the Success Rate

The first attempt to estimate the number of traces required to achieved a level of success rate from the value of correlation coefficient has been made in [22]. However, it assumes that the distribution of the wrong keys to have zero mean. It also considers the distributions of the correlation coefficients for different keys as independent. Later in [27], Rivain proposed to evaluate the 'exact' success rate of a side channel analysis from the leakage distribution and the algorithmic parameters using the widely admitted Gaussian leakage model. In [13], Fei at al. has established an analytical relationship among the success rate of a mono-bit DPA attack, the side channel characteristic and the algorithm characteristic of an implementation. Their approach has been extended for multi-bits CPA in [31].

In this section, we follow an alternative way to derive the success rate of CPA on the leakage of a sample point which follows Eq. 2. As stated in [13], there are two sources of randomness in the output of a distinguisher. The first source is the randomness in the inputs and the second source is the noise (part of the leakage independent of the target S). In [13], Fei et al. have assumed the distribution of the input plaintext to be uniformly random while in [31], Thillard et al. assumed that the frequency of every plaintext to be same. In our derivation, we let the parameters to be the functions of input distribution, since the input distribution is known. Moreover when plaintext distribution is uniformly random, the parameters converge to some global value for sufficiently large number of plaintexts.



(a) Approach 1: Combining is done before applying the distinguisher.

(b) Approach 2: Combining is done after applying the distinguisher.

Fig. 1: Alternative approaches for combining leakages of multiple sample points.

Here we derive the success rate by considering a known distribution $\mathscr{D}$ for the input random variable X. For convenience, we neglect the sub-script t^* of the leakage L_{t^*} in Eq. 2.

As in [27], we define the occurrence ratio of $x \in \mathcal{X}$ in the input vector $\mathbf{x}$ as

$$r_x = \frac{|\{i|x_i = x\}|}{q} \tag{4}$$

and $\bar{P}_k$ as $\sum_{x\in\mathcal{X}} r_x P_k(x)$. For simplicity, we also focus on the distribution of the following coefficient

$$\dot{\rho}_k = \frac{1}{q\hat{\sigma}_k}\sum_{i=0}^{q-1}(P_k(x_i) - \bar{P}_k)l^i_{t^*} \tag{5}$$

Note that as argued in [27], by replacing ρ_k by $\dot{\rho}_k$, success rate remains unchanged in a univariate attack. The distribution of $\dot{\rho}_k$ is given by the following proposition.

Proposition 1: [27] The vector of coefficients $\{\dot{\rho}_k\}_{k\in\mathcal{K}}$ has multivariate Gaussian distribution for the input distribution $\mathscr{D}$ with mean vector $\mu_{\dot{\rho}}$ having elements

$$E[\dot{\rho}_k] = \frac{1}{\hat{\sigma}_k}\sum_{x\in\mathcal{X}} r_x(P_k(x) - \bar{P}_k)E[L|x] \tag{6}$$

for all $k \in \mathcal{K}$ and with covariance matrix $\mathbf{\Sigma}_{\dot{\rho}}$ having elements

$$\begin{aligned} Cov(\dot{\rho}_{k_1}, \dot{\rho}_{k_2}) = &\frac{1}{q\hat{\sigma}_{k_1}\hat{\sigma}_{k_1}}\sum_{x\in\mathcal{X}} r_x(P_{k_1}(x) - \bar{P}_{k_1})\times \\ &(P_{k_2}(x) - \bar{P}_{k_2})Var(L|x) \end{aligned} \tag{7}$$

for all $(k_1, k_2) \in \mathcal{K}^2$.

Applying the above proposition on the leakage L which follows the leakage model given in Eq. 2, we state and prove the following result about the distribution of the comparison vector (as defined in [27]) $\{\Delta\dot{\rho}_k\}_{k\in\mathcal{K}\setminus\{k^*\}} = \{\dot{\rho}_{k^*} - \dot{\rho}_k\}_{k\in\mathcal{K}\setminus\{k^*\}} = \mathbf{\Delta}\dot{\rho}$:

Corollary 1: The comparison vector $\mathbf{\Delta}\dot{\rho}$ has a multivariate Gaussian distribution with mean vector $\mu_{\mathbf{\Delta}\dot{\rho}}$ having elements

$$E[\Delta\dot{\rho}_k] = a \cdot Cov(\mathbf{\Delta P}_k, \mathbf{P}_{k^*}) \tag{8}$$

for all $k \in \mathcal{K}\setminus\{k^*\}$ where $\mathbf{\Delta P}_k = \{\frac{P_{k^*}(x)}{\hat{\sigma}_{k^*}} - \frac{P_k(x)}{\hat{\sigma}_k}\}_{x\in\mathcal{X}}$ and $\mathbf{P}_{k^*} = \{P_{k^*}(x)\}_{x\in\mathcal{X}}$. The distribution of the vector has a covariance matrix $\mathbf{\Sigma}_{\mathbf{\Delta}\dot{\rho}}$ having elements

$$Cov(\Delta\dot{\rho}_{k_1}, \Delta\dot{\rho}_{k_2}) = \frac{\sigma_N^2}{q}Cov(\mathbf{\Delta P}_{k_1}, \mathbf{\Delta P}_{k_2}) \tag{9}$$

for all $(k_1, k_2) \in (\mathcal{K}\setminus\{k^*\})^2$ where $\mathbf{\Delta P}_k$ is defined as before.

Proof: From the definition of $\Delta\dot{\rho}_k$ and Eq. 5, we get

$$\begin{aligned} \Delta\dot{\rho}_k &= \dot{\rho}_{k^*} - \dot{\rho}_k \\ &= \frac{1}{q\hat{\sigma}_{k^*}}\sum_{i=0}^{q-1}(P_{k^*}(x_i) - \bar{P}_{k^*})l^i_{t^*} - \frac{1}{q\hat{\sigma}_k}\sum_{i=0}^{q-1}(P_k(x_i) - \bar{P}_k)l^i_{t^*} \\ &= \frac{1}{q}\sum_{i=0}^{q-1}(\Delta P_k(x_i) - \Delta\bar{P}_k)l^i_{t^*} \end{aligned}$$

where $\Delta P_k(x_i)$ and $\Delta\bar{P}_k$ denote $\frac{P_{k^*}(x_i)}{\hat{\sigma}_{k^*}} - \frac{P_k(x_i)}{\hat{\sigma}_k}$ and $\frac{\bar{P}_{k^*}}{\hat{\sigma}_{k^*}} - \frac{\bar{P}_k}{\hat{\sigma}_k}$ respectively. Applying Proposition 1 on the above expression of $\Delta\dot{\rho}_k$ and using Eq. 2, we get

$$\begin{aligned} E[\Delta\dot{\rho}_k] &= \sum_{x\in\mathcal{X}} r_x(\Delta P_k(x) - \Delta\bar{P}_k)E[L|x] \\ &= \sum_{x\in\mathcal{X}} r_x(\Delta P_k(x) - \Delta\bar{P}_k)(a \cdot P_{k^*}(x) + \mu_N) \\ &= \sum_{x\in\mathcal{X}} r_x(\Delta P_k(x) - \Delta\bar{P}_k)(a \cdot P_{k^*}(x)) \\ &= a \cdot Cov(\mathbf{\Delta P}_k, \mathbf{P}_{k^*}) \end{aligned}$$

Similarly,

$$\begin{aligned} Cov(\Delta\dot{\rho}_{k_1}, \Delta\dot{\rho}_{k_2}) = &\frac{1}{q}\sum_{x\in\mathcal{X}} r_x(\Delta P_{k_1}(x) - \Delta\bar{P}_{k_1})\times \\ &(\Delta P_{k_2}(x) - \Delta\bar{P}_{k_2})Var(L|x) \\ = &\frac{\sigma_N^2}{q}Cov(\mathbf{\Delta P}_{k_1}, \mathbf{\Delta P}_{k_2}) \end{aligned}$$

■

For a successful attack, the condition $\{\Delta\dot{\rho}_k\}_{k\in\mathcal{K}\setminus\{k^*\}} > \mathbf{0}$ holds where $\mathbf{0}$ is a zero vector of size $|\mathcal{K}| - 1$ and $\mathbf{v}_1 > \mathbf{v}_2$ implies each element of $\mathbf{v}_1$ is greater than the corresponding element of $\mathbf{v}_2$. Thus the first order success rate can be given by the term $Pr(\{\Delta\dot{\rho}_k\}_{k\in\mathcal{K}\setminus\{k^*\}} > \mathbf{0})$. We mention by passing that for a negative value of a in Eq. 2, one would expect a negative correlation for the correct key and thus the definition of success rate should be changed accordingly. For the time being we assume a positive correlation for the correct key and state Proposition 2. Without the loss of generality, we also assume that the distribution of $\boldsymbol{\Delta}\dot{\rho} = \{\Delta\dot{\rho}_k\}_{k\in\mathcal{K}\setminus\{k^*\}}$ is non-degenerative [1].

Proposition 2: The first order success rate ($1OSR^{\mathscr{D}}$) of CPA for the input distribution $\mathscr{D}$ is given by

$$1OSR^{\mathscr{D}} = \Phi_{\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\dot{\rho}}}(\mu_{\boldsymbol{\Delta}\dot{\rho}}) \tag{10}$$

where $\Phi_{\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\dot{\rho}}}$ be the cdf of a multivariate normal distribution with $|\mathcal{K}|-1$ dimensional zero mean vector and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\dot{\rho}}$.

Proof: Since, $\boldsymbol{\Delta}\dot{\rho}$ follows the multivariate normal distribution with mean $\mu_{\boldsymbol{\Delta}\dot{\rho}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\dot{\rho}}$, the first order success rate is given by

$$\begin{aligned}
1OSR^{\mathscr{D}} &= Pr(\boldsymbol{\Delta}\dot{\rho} > \mathbf{0})\\
&= f_{\boldsymbol{\Delta}\dot{\rho}}(\mathbf{0} < \boldsymbol{\Delta}\dot{\rho} < \infty)\\
&= f_{\boldsymbol{\Delta}\dot{\rho}}(-\mu_{\boldsymbol{\Delta}\dot{\rho}} < \boldsymbol{\Delta}\dot{\rho} - \mu_{\boldsymbol{\Delta}\dot{\rho}} < \infty)\\
&= f_{\boldsymbol{\Delta}\dot{\rho}}(-\infty < \boldsymbol{\Delta}\dot{\rho} - \mu_{\boldsymbol{\Delta}\dot{\rho}} < \mu_{\boldsymbol{\Delta}\dot{\rho}})\\
&= \Phi_{\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\dot{\rho}}}(\mu_{\boldsymbol{\Delta}\dot{\rho}})
\end{aligned}$$

where $f_{\boldsymbol{\Delta}\dot{\rho}}$ denotes the pdf of the distribution of $\boldsymbol{\Delta}\dot{\rho}$.

■

To analyse the first order success rate further, from Corollary 1 we note that $\mu_{\boldsymbol{\Delta}\dot{\rho}} = a\mu_{\boldsymbol{\Delta}\mathbf{P}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\dot{\rho}} = \frac{\sigma_N^2}{q}\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}$ where $\mu_{\boldsymbol{\Delta}\mathbf{P}}$ be the vector $\{Cov(\boldsymbol{\Delta}\mathbf{P}_k, \mathbf{P}_{k^*})\}_{k\in\mathcal{K}\setminus\{k^*\}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}$ be the $(|\mathcal{K}|-1)\times(|\mathcal{K}|-1)$ matrix with elements $Cov(\boldsymbol{\Delta}\mathbf{P}_{k_1}, \boldsymbol{\Delta}\mathbf{P}_{k_2})$ for all $(k_1,k_2) \in (\mathcal{K}\setminus\{k^*\})^2$. Let us also define the signal-to-noise ratio [23] of traces as

$$SNR = \frac{Var(E[L|P_{k^*}])}{Var(L - E[L|P_{k^*}])} = \frac{a^2\sigma_{k^*}^2}{\sigma_N^2} \tag{11}$$

Then we state the following result.

Corollary 2: The first order success rate ($1OSR^{\mathscr{D}}$) of CPA for the input distribution $\mathscr{D}$ is given by

$$1OSR^{\mathscr{D}} = \Phi_{\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}}(\sqrt{q}\sqrt{SNR}\sigma_{k^*}^{-1}\mu_{\boldsymbol{\Delta}\mathbf{P}}) \tag{12}$$

where Φ, $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}$ and $\mu_{\boldsymbol{\Delta}\mathbf{P}}$ is defined as before.

Proof: Let us first denote the multi-dimensional intigration $\int_{ll_0}^{ul_0}\cdots\int_{ll_{|\mathcal{K}|-1}}^{ul_{|\mathcal{K}|-1}} f(y_0,\ldots,y_{|\mathcal{K}|-1})dy_{|\mathcal{K}|-1}\cdots dy_0$ as $\int_{\mathbf{ll}}^{\mathbf{ul}} f(y_0,\cdots,y_{|\mathcal{K}|-1})d\mathbf{y}$ where $\mathbf{ll} = \{ll_0,\ldots,ll_{|\mathcal{K}|-1}\}$, $\mathbf{ul} = \{ul_0,\ldots,ul_{|\mathcal{K}|-1}\}$ and $\mathbf{y} = \{y_0,\ldots,y_{|\mathcal{K}|-1}\}$. From Proposition 2, we get

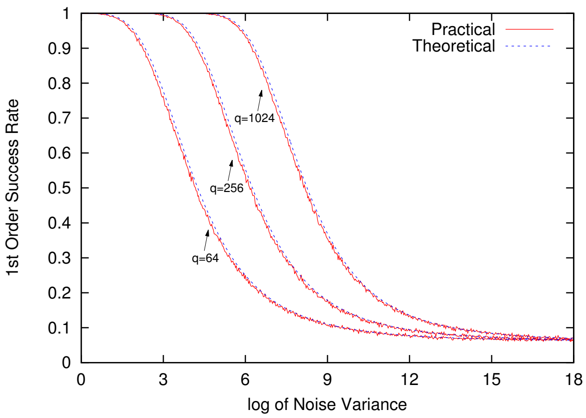


Fig. 2: Plots of the practical and theoretical $1OSR$ for CPA on the output of Present S-box using HW model. The $10SR$ is estimated using number of traces 64, 256 and 1024 respectively.

$$\begin{aligned}
1OSR^{\mathscr{D}} &= \int_{-\infty}^{\mu_{\boldsymbol{\Delta}\dot{\rho}}} \frac{1}{\sqrt{(2\pi)^{k-1}|\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\dot{\rho}}|}} e^{-\frac{1}{2}\mathbf{y}'\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\dot{\rho}}^{-1}\mathbf{y}} d\mathbf{y}\\
&= \int_{-\infty}^{\frac{\sqrt{q}}{\sigma_N}\mu_{\boldsymbol{\Delta}\dot{\rho}}} \frac{1}{\sqrt{(2\pi)^{k-1}|\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}|}} e^{-\frac{1}{2}\tilde{\mathbf{y}}'\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}^{-1}\tilde{\mathbf{y}}} d\tilde{\mathbf{y}},\\
&\qquad\qquad\qquad\qquad \text{where } \tilde{\mathbf{y}} = \frac{\sqrt{q}}{\sigma_N}\mathbf{y}\\
&= \Phi_{\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}}(\frac{\sqrt{q}}{\sigma_N}\mu_{\boldsymbol{\Delta}\dot{\rho}})\\
&= \Phi_{\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}}(\frac{\sqrt{q}a}{\sigma_N}\mu_{\boldsymbol{\Delta}\mathbf{P}})\\
&= \Phi_{\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}}(\sqrt{q}\sqrt{\frac{a^2\sigma_{k^*}^2}{\sigma_N^2}}\sigma_{k^*}^{-1}\mu_{\boldsymbol{\Delta}\mathbf{P}})\\
&= \Phi_{\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}}(\sqrt{q}\sqrt{SNR}\sigma_{k^*}^{-1}\mu_{\boldsymbol{\Delta}\mathbf{P}})
\end{aligned}$$

■

When the input random variable X is uniformly random, for sufficiently large value of q, the input distribution $\mathscr{D}$ converges. Therefore, $1OSR^{\mathscr{D}}$ also converges to the global first order success rate. Thus, Corollary 2 expresses the first order success rate of a CPA attack in terms of number of power traces q, side channel characteristic SNR and some algorithm dependent parameters like $\mu_{\boldsymbol{\Delta}\mathbf{P}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{P}}$ and σ_{k^*}. Moreover, it shows how the first order success rate of CPA depends on the SNR of the power traces.

To experimentally validate Eq. 12, we computed practical $1OSR$ by simulation. For the simulation, we generated power traces by adding Gaussian noise to the Hamming weight of the output of Present S-box. The success rate is computed by repeating CPA on the simulated power traces 10000 times. On the other hand, we estimated theoretical $1OSR$ using Eq. 12. Both the results are plotted in Fig. 2 with the increasing variance of Gaussian noise.

Similar relation between more general o^{th} order success rate with SNR can be found. Thus for a given algorithm and a fixed set of traces, maximization of the success rate requires the maximization of SNR. In this work, we combine the leakages $L_0, \cdots, L_{T-1}$ using a linear function g in such a way that it maximizes the SNR of the resultant leakage $g(L_0, \cdots, L_{T-1})$. However, such combining is not possible in non-profiling setup without any estimation of the information contained in each sample point. Thus in the following sections, we try to estimate the information at each sample point using some parameters which can be computed without knowing the correct key.

IV. Multivariate Leakage Model: Extending the leakage model over multiple time samples

A. Profiling the Power Traces of AES

In this section, our objective is to investigate how leakage due to a known computation varies over a range of sample points. The nature of leakages at several sample points have been investigated with respect to the correct predicted leakage P i.e. $P = P_{k^*} = \Psi(S)$ using the following metrics.

1) *Squared Pearson's Correlation between Data Dependent Leakage and Predicted Leakage (SCDP)*: It is defined as follows:

$$SCDP_t = Corr^2(E[L_t|P], P)$$

$E[L_t|p]$ quantifies the deterministic leakage at sample point t due to the predicted leakage p for $p \in \mathcal{P}$. Since, Pearson's correlation detects the linear relation between two variables [11], $SCDP_t = Corr^2(E[L_t|P], P)$ reveals the linear dependency between the deterministic leakage at t and the predicted leakage P. It should be noted that if the leakage of a sample point t follows Eq. 2, then the value of $SCDP$ at t is almost one. On the other hand, if L_t and P are almost independent, $E[L_t|p]$ will be almost constant for all $p \in \mathcal{P}$, resulting to $SCDP_t$ almost zero.

2) *Variation of Data Dependent Leakage (VDL)*:

$$VDL_t = Var(E[L_t|P])$$

It reveals the variations in leakage caused by the predicted leakage P at sample point t. Sometimes, it is used to quantify the signal in the leakage. On the other hand, noise is quantified by $Var(L_t - E[L_t|P])$.

3) *Squared Mean Leakage (SML)*:

$$SML_t = E^2[L_t]$$

It quantifies the magnitude or the strength of the leakage at a sample point.

Fig. 3 shows the plot for the above three metrics which are estimated over $20,000$ traces of AES encryptions. The AES is implemented using parallel iterative hardware architecture on the setup described in Appendix A. The correct predicted leakage P is taken as the Hamming distance between the ciphertext and the input to the last round. The metrics are plotted only for 400 sample points around the last round register update.

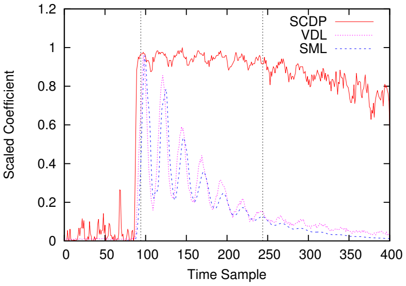


Fig. 3: Plots of the chosen metrics in the last round of unprotected implementation of AES.

The figure shows that as the cycle begins, with the mean leakage (SML), SCDP also rises rapidly, remains almost constant for about 150 sample points and then it decreases slowly. The slight fluctuations in the curve are due to the presence of small amount of noise after averaging a limited number of power traces. This leads us to the following observation:

Observation 1: The deterministic leakages at a large number of sample points show high linear dependencies with the correct predicted leakage P.

In other words, a large number of sample points contains information about the correct predicted leakage P. It should be noted that various profiling attacks optimally extract the information from multiple sample points by estimating the multivariate leakage distribution of the sample points using a profiling step.

From the figure, we can also see that VDL almost superimposes on SML i.e. VDL is highly correlated to SML. This leads us to the following observation:

Observation 2: The variation in deterministic leakage of a sample point is correlated to the square of the mean leakage of the sample point.

In other words, the second observation states that the magnitude of the variation in leakage at a sample point due to some computation is proportional to the mean value (strength) of the leakage at that sample point. It should be noted that similar kinds of observation can be found in Chapter 4.3.2 of [23] for the leakages of a micro-controller. The authors have also suggested several trace compression techniques based on the observation and have shown their usefulness to attack software implementation of AES. However, to the best of our knowledge, no attempt has been made to incorporate these observations into the leakage model.

In the next sub-section, we extend the conventional leakage model for multiple sample points using the above two observations which will be latter used to derive an impulse response of the matched filter.

B. Modeling the Leakage over Multiple Time Samples

Observation 1 and 2 immediately extend the conventional leakage model given by Eq. (2), into the following multivariate leakage model:

$$L_t = a_t \cdot \Psi(S) + N_t = a_t \cdot P + N_t \qquad (13)$$

for $t_0 \leq t < t_0 + \tau$ where $a_t \in \mathbb{R}$ and the random vector $\mathbf{N} = \{N_{t_0}, \ldots, N_{t_0+\tau-1}\}$ follows a multivariate Gaussian distribution with zero mean vector and covariance matrix $\mathbf{\Sigma_N}$. It should be noted that the linear relation in Eq. (13) is a consequence of Observation 1 while Observation 2 enforces the mean vector of $\mathbf{N}$ to be a zero vector.

In a parallel iterative hardware architecture, a single round consists of several parallel S-boxes and the attacker targets only a part of it (usually a single S-box). Thus, in addition to the predicted leakage P due to the computation of the target $S = F_{k^*}(X)$, leakage due to the computation of the other parallel bits adds to it. This is known as algorithmic noise and we denote it by U. It should be noted that for a fully serialized architecture, U takes the value zero. Leakages due to the key bits and the control bits is denoted by c. Since key scheduling and the controlling operations are fixed for a specific round in all the encryptions, c is constant for all the inputs.

Thus, we can adopt Eq. 13 to incorporate these new variables as follows:

$$\begin{aligned} L_t &= a_t \cdot (P + U + c) + N_t & (14)\\ &= a_t \cdot (I + c) + N_t, \qquad t_0 \leq t < t_0 + \tau & (15) \end{aligned}$$

where $I = P + U$. We are interested in the leakages of the above window namely $\{t_0, t_0 + 1, \ldots, t_0 + \tau - 1\}$ that can be roughly determined by the clock cycle in which the target operation is being performed. We denote this time span by $\{0, \ldots, \tau - 1\}$ and in the rest of the paper, power trace is referred by the sample points of this time span only.

V. Derivation of the matched filter impulse response

In this section, our objective is to explore Approach 1 of Fig. 1 in order to maximize the success rate of CPA. Since we are restricting the combining function g to be a linear function, our objective is to find a vector $\mathbf{h} = \{h_0, \ldots, h_{\tau-1}\}$ which maximizes the success rate of CPA on the combined leakage $L_o = \sum_{t=0}^{\tau-1} h_t L_t$. According to Corollary 2, this objective is achieved by finding $\mathbf{h}$ which maximizes the SNR of the output leakage. Such a pre-processing is commonly referred to as time-domain filtering and the vector $\mathbf{h}$ is referred to as impulse response of the filter. Time-domain filters have been previously applied in side-channel attacks in [24], [10], [25]. In [25], the success rate of CPA has been maximized by searching the optimal linear FIR filter coefficients $\mathbf{h}$. However, there the authors have assumed a semi-profiling approach. In this section, we find the optimum impulse response of $\mathbf{h}$ using the multivariate leakage model in a non-profiling setup.

In practice, a linear filter is applied on each of the traces separately. Thus, we need to define the SNR of a sample point with respect to a single trace. The output l_o of a linear filter with impulse response $\mathbf{h} = \{h_0, \ldots, h_{\tau-1}\}$ applied on a trace $\mathbf{l} = \{l_0, \ldots, l_{\tau-1}\}$ is given by

$$l_o = \sum_{t=0}^{\tau-1} h_t L_t = \mathbf{h}'\mathbf{l} \qquad (16)$$

Since the trace $\mathbf{l}$ satisfies the multivariate leakage model given in Eq. 15, t^{th} element of $\mathbf{l}$ satisfies $l_t = a_t \cdot (i + c) + n_t$ where i and n_t are the instants of random variables I and N_t. Hence, the output leakage l_o can be written as

$$l_o = \sum_{t=0}^{\tau-1} (h_t a_t \cdot (i + c) + h_t n_t) = (i + c)\mathbf{h}'\mathbf{a} + \mathbf{h}'\mathbf{n} \qquad (17)$$

Without loss of generality, we assume that the traces are centered to zero with respect to its mean over all the traces. Thus, the centred trace $\tilde{\mathbf{l}}$ can be represented by

$$\begin{aligned} \tilde{\mathbf{l}} &= \mathbf{l} - E[\mathbf{l}]\\ &= (i - E[I])\mathbf{a} + \mathbf{n}\\ &= \tilde{i}\mathbf{a} + \mathbf{n} & (18) \end{aligned}$$

where $\tilde{i} = i - E[I]$. Similarly the centered output leakage can be represented by

$$\begin{aligned} \tilde{l}_o &= \mathbf{h}'(\mathbf{l} - E[\mathbf{L}]) & (19)\\ &= \tilde{i}\mathbf{h}'\mathbf{a} + \mathbf{h}'\mathbf{n} & (20) \end{aligned}$$

where $\tilde{i}$ is defined as before.

Now following the standard definition of SNR [32], we define the SNR of a sample point t with respect to a single trace $\tilde{\mathbf{l}}$ as

$$\begin{aligned} SNR_t^{\tilde{\mathbf{l}}} &= \frac{\text{Power of the signal in the trace } \tilde{\mathbf{l}} \text{ at sample point } t}{\text{Average noise power}}\\ &= \frac{(E[\tilde{l}_t | I = i])^2}{E[|n_t|^2]}\\ &= \frac{\tilde{i}^2 \cdot a_t^2}{\sigma_{N_t}^2} & (21) \end{aligned}$$

It should be noted that SNR of the sample point t over all the traces can be obtained as

$$SNR_t = E_{\tilde{\mathbf{l}}}[SNR_t^{\tilde{\mathbf{l}}}] = \frac{a_t^2 \sigma_I^2}{\sigma_{N_t}^2}$$

which is equivalent to the definition given by Eq. 11.

From Eq. 20, we compute the SNR of the output leakage $\tilde{l}_o$ as follows:

$$\begin{aligned} SNR^{\tilde{l}_o} &= \frac{|\tilde{i}\mathbf{h}'\mathbf{a}|^2}{E[|\mathbf{h}'\mathbf{n}|^2]}\\ &= \tilde{i}^2 \times \frac{|\mathbf{h}'\mathbf{a}|^2}{E[(\mathbf{h}'\mathbf{n})(\mathbf{h}'\mathbf{n})']}\\ &= \tilde{i}^2 \times \frac{|\mathbf{h}'\mathbf{a}|^2}{\mathbf{h}'\mathbf{\Sigma_N}\mathbf{h}} & (22) \end{aligned}$$

Recall that $\mathbf{\Sigma_N}$ be the $\tau \times \tau$ covariance matrix of the multivariate Gaussian noise $\mathbf{N} = \{N_0, \cdots, N_{\tau-1}\}$.

The filter which maximizes the $SNR^{\tilde{l}_o}$ is commonly referred to as matched filter in DSP and its impulse response $\mathbf{h}$ involves auto-correlation function or the covariance matrix $\mathbf{\Sigma_N}$ [32]. However, computation of $\mathbf{\Sigma_N}$ requires the knowledge of the secret key which cannot be obtained in non-profiling DPA. Thus instead of optimizing the SNR of the output signal, we will optimize a different metric which does not involve the covariance matrix of noise. Hence, we define *Signal Ratio* (SR) of the output as the ratio of the power of the output due to the deterministic leakage and the average output power:

$$SR^{\tilde{l_o}} = \frac{|\tilde{i} \cdot \mathbf{h}'\mathbf{a}|^2}{E[|\tilde{l_o}|^2]}$$

We simplify the above definition as:

$$\begin{aligned} SR^{\tilde{l_o}} &= \frac{\tilde{i}^2|\mathbf{h}'\mathbf{a}|^2}{E[|\mathbf{h}'(\mathbf{l}-E[\mathbf{L}])|^2]}, && \text{using Eq. 19} \\ &= \tilde{i}^2 \times \frac{|\mathbf{h}'\mathbf{a}|^2}{\mathbf{h}'E[(\mathbf{l}-E[\mathbf{L}])'(\mathbf{l}-E[\mathbf{L}])]\mathbf{h}} \\ &= \tilde{i}^2 \times \frac{|\mathbf{h}'\mathbf{a}|^2}{\mathbf{h}'\mathbf{\Sigma_L}\mathbf{h}} && (23) \\ &= \tilde{i}^2 \times \frac{|\mathbf{h}'\mathbf{a}|^2}{\mathbf{h}'\mathbf{\Sigma_D}\mathbf{h} + \mathbf{h}'\mathbf{\Sigma_N}\mathbf{h}} \\ &= \tilde{i}^2 \times \frac{|\mathbf{h}'\mathbf{a}|^2}{\sigma_I^2|\mathbf{h}'\mathbf{a}|^2 + \mathbf{h}'\mathbf{\Sigma_N}\mathbf{h}} && (24) \end{aligned}$$

where $\mathbf{\Sigma_L}$ and $\mathbf{\Sigma_D}$ be the covariance matrices of the total leakage and the deterministic leakage respectively. The last step follows because $Cov(a_{t_1}(I+c), a_{t_2}(I+c)) = a_{t_1}a_{t_2}\sigma_I^2$. Our objective is to find $\mathbf{h}$ such that $SNR^{\tilde{l_o}}$ of the output is maximum. Interestingly, both the $SNR^{\tilde{l_o}}$ and the $SR^{\tilde{l_o}}$ reaches their maximum simultaneously. It is stated in the following lemma.

Lemma 1: The SNR of the output leakage l_o reaches its maximum if and only if SR of that also reaches its maximum.

Proof: From Eq. 24,

$$SR^{\tilde{l_o}} = \frac{1}{\frac{\sigma_I^2}{\tilde{i}^2} + \frac{\mathbf{h}'\mathbf{\Sigma_N}\mathbf{h}}{\tilde{i}^2|\mathbf{h}'\mathbf{a}|^2}} = \frac{1}{c_1 + \frac{1}{SNR^{\tilde{l_o}}}}$$

where $c_1 = \frac{\sigma_I^2}{\tilde{i}^2}$. We can rewrite the above equation as,

$$\frac{1}{SR^{\tilde{l_o}}} = c_1 + \frac{1}{SNR^{\tilde{l_o}}}$$

Since c_1 is constant for a given trace, the $SR^{\tilde{l_o}}$ of the output leakage reaches its maximum if and only if the $SNR^{\tilde{l_o}}$ reaches its maximum. ■

In Lemma 2, we state an expression of $\mathbf{h}$ which maximizes the SR of the output leakage.

Lemma 2: The impulse response $\mathbf{h}$ of a linear filter which maximizes the SR of the output leakage l_o can be derived as $\mathbf{\Sigma_L^{-1}}\mathbf{a}$.

The proof can be followed from the derivation of the matched filter given in [3] by replacing SNR with SR using Eq. 23. We, now, state and prove our final result in Theorem 1. Before that let us denote by $\mu_\mathbf{L}$ the mean leakage vector $E[\mathbf{L}] = \{E[L_0], \cdots, E[L_{\tau-1}]\}$.

Theorem 1: The impulse response $\mathbf{h}$ of the optimum linear filter for the leakage $\mathbf{L}$ which follows Eq. 15 can be given by $\mathbf{\Sigma_L^{-1}}\mu_\mathbf{L}$.

Proof: Taking the expectation on both sides of Eq. 15 we get, $\mathbf{a} = \mu_\mathbf{L}/(E[I]+c)$. Putting this value of $\mathbf{a}$ in the expression of the impulse response obtained from Lemma 2, we get the impulse response of the linear filter which maximizes the SR of the output l_o to be $\mathbf{h} = \mathbf{\Sigma_L^{-1}}\mathbf{a}/(E[I]+c)$. Since, by Lemma 1, maximization of SR also leads to the maximization of SNR, by neglecting the constant divisor of $\mathbf{h}$, we conclude that the impulse response of the optimum linear filter for the output response is $\mathbf{\Sigma_L^{-1}}\mu_\mathbf{L}$. ■

Thus, the impulse response of an optimum linear filter can be computed using the expression $\mathbf{\Sigma_L^{-1}}\mu_\mathbf{L}$. It should be noted that neither $\mathbf{\Sigma_L}$ nor $\mu_\mathbf{L}$ requires the knowledge of the correct key to estimate. Hence, the filter can be useful in non-profiling DPA also.

Elimination of the Matrix Inversion: Computation of $\mathbf{\Sigma_L^{-1}}\mu_\mathbf{L}$ involves the computation of the inverse of a $\tau \times \tau$ matrix which has a computational complexity $\mathcal{O}(\tau^3)$. Moreover, the inverse operation is highly susceptible to the error in the estimation of the covariance matrix. We avoid this operation by setting the off-diagonal elements of the covariance matrix $\mathbf{\Sigma_L}$ to zero which results in the approximated impulse response

$$\tilde{\mathbf{h}} = \{\frac{\mu_{L_0}}{\sigma_{L_0}^2}, \cdots, \frac{\mu_{L_{\tau-1}}}{\sigma_{L_{\tau-1}}^2}\} \qquad (25)$$

In other words, the above approximation neglects the correlation between different sample points of the power traces.

VI. Computing in a new Basis

When the leakages of different sample points are significantly correlated, the approximation of Eq. 25 might result into sub-optimal pre-processing. To avoid this, the leakage $L = \{L_0, \cdots, L_{\tau-1}\}$ can be transformed into a new basis system $\tilde{L} = \{\tilde{L}_0, \cdots, \tilde{L}_{\tau-1}\}$ by some linear transformation such that the leakage components along two different axes $\tilde{L}_{t_1}$ and $\tilde{L}_{t_2}$ become uncorrelated. Here, we discuss two such basis conversions.

Principal Component Analysis: Principal Component Analysis (PCA) [12] is a mean to convert a data set into the basis of eigenvectors of its covariance matrix. In this new basis, components along different axes (Principal Components or PCs) are uncorrelated to each other. Moreover, PCs are sorted by their variance i.e. the first PC has maximum variance, the second PC has second maximum variance, and so on. Thus in low noise scenario, where most of variations in traces is due to the target S, PCA projects the data dependent variations (signal) into the first PC while variations in all other PCs are mainly caused by noise. Thus, performing DPA on the first PC greatly increases performance of a DPA attacks [4], [29], [6]. However in high noise scenario, data dependent variations are rather scattered in all the PCs [6], [18]. Since, PCA is a linear transformation [12], Eq. 15 is valid in the domain of eigenvector also. Consequently, we can apply Eq. 25 on this domain i.e. on the PCs.

Discrete Fourier Transform: Other alternative is to use Discrete Fourier Transform (DFT) to convert the leakage samples to a new orthonormal basis (frequency domain). In frequency domain, the absolute value of the complex coefficients obtained from the DFT is commonly used to attack [14], [25]. By taking only the absolute value, phase component is ignored which is useful to attack misaligned traces. However, we do not use it since the absolute operation is not a linear operation. Rather, we keep both the real part (cosine coefficient) and the imaginary part (sine coefficient) as separate sample points. Since, both the real and the imaginary part are obtained using linear transformations and the linear transformation does not destroy the statistical property of the power signal, the resulting DFT traces also follows Eq. 15. It should be noted that this approach aims at gaining efficiency in the presence of misaligned traces by optimally combining leakages spread over multiple sample points due to the misalignment. Moreover, even if there exist significant correlations among sample points in time domain, we can assume the covariance matrix of the sample points in frequency domain is sparsed. Hence, we can use Eq. 25 to compute approximated matched filter on this domain.

Determination of Window: The model is valid only in the clock cycle in which the target operation is being performed (called *target clock cycle*). For an iterative hardware architecture, the window can be set to the whole period of the target clock cycle which is relatively easier to find. However in our experiments, we have roughly chosen the window from the beginning of the target clock cycle up to a sample point for which the mean leakage is slightly greater than zero.

VII. Experimental Results

For experimental evaluation, we have collected 40 sets of $3,000$ traces of AES encryptions. The cipher is implemented using parallel iterative hardware architecture on SASEBO-GII using the setup described in Appendix A. The S-boxes are implemented using Xilinx device primitive: distributed ROM. The setup is properly calibrated to reduce the quantization noise.

We performed CPA (1) on all the sample points independently and (2) on the output of approximated matched filter (AMF) applied on (a) time domain traces, (b) frequency domain traces and (c) on the PCs by adding Gaussian noise at each sample point. In addition to these, we also performed CPA on the output of matched filter (MF) on time domain traces. Fig. 4 shows global success rate [2] of CPA after applying all the above pre-processing. Global success rate is defined by the probability of getting the correct key for all the 16 bytes simultaneously. The figure shows that both MF and AMF optimizes the performance of CPA in each of the three domains.

We have further evaluated the pre-processing techniques by adding a constant noise to each of the sample points of the traces. Such noise resemblances very low frequency noise such as flicker noise [21]. In the presence of constant noise, leakages of the different sample points gets positively correlated. Thus filtering using AMF which neglects the correlation between two different sample points becomes sub-optimal. This can be seen in Fig. 5. The figure shows the GSR of CPA on the output of AMF is badly affected by the constant noise. However, AMF in frequency domain and on PCs performs almost optimally since in the new basis the sample points get sparsely correlated.

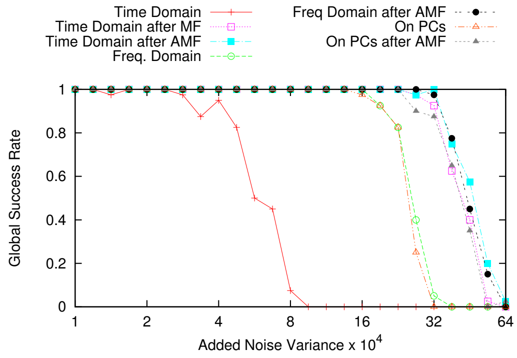


Fig. 4: Plots of the global success rate (GSR) of CPA after applying various pre-processing techniques on real traces of AES encryption. The GSR is computed over 40 sets of $3,000$ power traces.

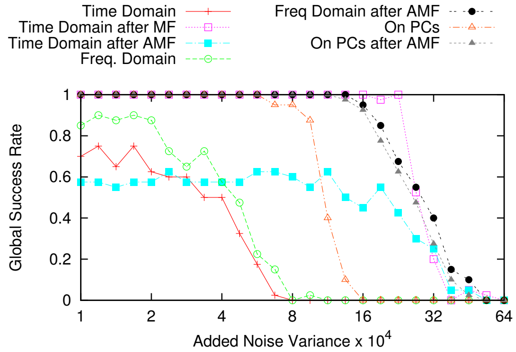


Fig. 5: Plots of the global success rate (GSR) of CPA after applying various pre-processing techniques on AES encryption traces. A constant noise is added to each sample point of the traces and then GSR is computed by adding independent Gaussian noise to each sample points of increasing variance.

VIII. Optimality of Matched Filter

In this section, our objective is to verify the optimality of matched filter as a pre-processing technique. To compare with we choose the Stochastic attack with a profiling step as an optimal attack since it can "learn" quickly using smaller number of traces [16]. Profiling Stochastic attack consists of three phases. In the first phase the deterministic leakage is estimated in b-dimensional vector space, and in the second phase the multivariate density of noise in estimated. Third phase is the key recovery phase where maximum likelihood principal [28] or the minimum principal [28] is used to find an unknown key using a new set of traces.

For our experiments, we chose the vector space using (1) the bit model where 9-dimensional space is chosen by taking each bit of the target S as the first eight dimensions and last one corresponds to constant leakage [28], [16], and (2) Hamming

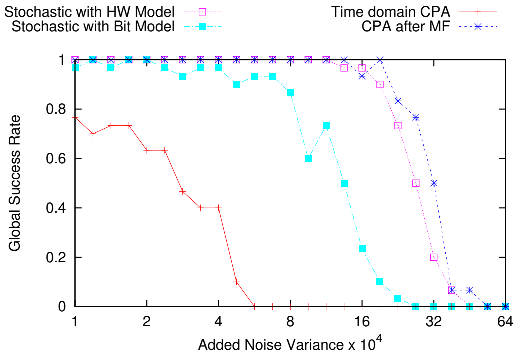


Fig. 6: Plots of the global success rate (GSR) of profiling Stochastic attack and CPA on the output of matched filter.

weight model where 2-dimensional space is chosen by taking the Hamming weight of the target S and the constant leakage. We used maximum likelihood principal in the third phase since it performs better than minimum principal method [28]. To evaluate the optimality of matched filter, we used first $60,000$ traces of $120,000$ traces to build the Stochastic models and the impulse response of matched filter and the rest of the traces for key recovery. To compute the success rate in the key recovery phase, we divided the $60,000$ traces into 30 groups of $2,000$ traces. The GSRs of all the attacks with increasing variance of added Gaussian noise are shown in Fig. 6. From the figure we see that CPA on the output of matched filter performs best which is slightly better than Stochastic attack using HW model. On the other hand, Stochastic attack using bit model which is more sensitive to the error in model estimation using a lesser number of traces performs worse than the one with HD model but better than classical time domain CPA.

IX. Conclusion

In this paper, we have derived the impulse response of the linear FIR filter (matched filter) which optimizes the SNR of the power traces for non-profiling DPA attacks. The derivation is based on the multivariate leakage model which is introduced for Virtex-5 FPGA device. We have experimentally evaluated several matched filter based pre-processing techniques. The experimental results reveal significant improvements of CPA using the proposed pre-processing techniques over the existing techniques in various noisy scenarios. We have further evaluated the optimality of the best proposed method by comparing it with profiling Stochastic attack.

References

[1] Multivariate normal distribution, en.wikipedia.org.

[2] Dpa contest/v2/, http://www.dpacontest.org/v2/index.php, 2012.

[3] Matched filter, en.wikipedia.org, 2013.

[4] C. Archambeau, E. Peeters, F.-X. Standaert, and J.-J. Quisquater. Template Attacks in Principal Subspaces. In Goubin and Matsui [17], pages 1–14.

[5] L. Batina, B. Gierlichs, and K. Lemke-Rust. Differential Cluster Analysis. In C. Clavier and K. Gaj, editors, *CHES*, volume 5747 of *Lecture Notes in Computer Science*, pages 112–127. Springer, 2009.

[6] L. Batina, J. Hogenboom, and J. G. J. van Woudenberg. Getting More from PCA: First Results of Using Principal Component Analysis for Extensive Power Analysis. In O. Dunkelman, editor, *CT-RSA*, volume 7178 of *Lecture Notes in Computer Science*, pages 383–397. Springer, 2012.

[7] E. Brier, C. Clavier, and F. Olivier. Correlation Power Analysis with a Leakage Model. In M. Joye and J.-J. Quisquater, editors, *CHES*, volume 3156 of *Lecture Notes in Computer Science*, pages 16–29. Springer, 2004.

[8] S. Chari, J. R. Rao, and P. Rohatgi. Template Attacks. In B. S. K. Jr., Çetin Kaya Koç, and C. Paar, editors, *CHES*, volume 2523 of *Lecture Notes in Computer Science*, pages 13–28. Springer, 2002.

[9] C. Clavier, J.-S. Coron, and N. Dabbous. Differential Power Analysis in the Presence of Hardware Countermeasures. In . K. Ko and C. Paar, editors, *CHES*, volume 1965 of *Lecture Notes in Computer Science*, pages 252–263. Springer, 2000.

[10] C. Clavier, J.-S. Coron, and N. Dabbous. Differential Power Analysis in the Presence of Hardware Countermeasures. In Çetin Kaya Koç and C. Paar, editors, *CHES*, volume 1965 of *Lecture Notes in Computer Science*, pages 252–263. Springer, 2000.

[11] S. Dowdy, S. Wearden, and D. Chilko. *Statistics for research*. John Wiley & Sons, third edition, 2004.

[12] R. O. Duda, P. E. P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley, pub-WILEY:adr, second edition, 2001.

[13] Y. Fei, Q. Luo, and A. A. Ding. A Statistical Model for DPA with Novel Algorithmic Confusion Analysis. In E. Prouff and P. Schaumont, editors, *CHES*, volume 7428 of *Lecture Notes in Computer Science*, pages 233–250. Springer, 2012.

[14] C. H. Gebotys, S. Ho, and C. C. Tiu. EM Analysis of Rijndael and ECC on a Wireless Java-Based PDA. In Rao and Sunar [26], pages 250–264.

[15] B. Gierlichs, L. Batina, P. Tuyls, and B. Preneel. Mutual Information Analysis. In E. Oswald and P. Rohatgi, editors, *CHES*, volume 5154 of *Lecture Notes in Computer Science*, pages 426–442. Springer, 2008.

[16] B. Gierlichs, K. Lemke-Rust, and C. Paar. Templates vs. Stochastic Methods. In Goubin and Matsui [17], pages 15–29.

[17] L. Goubin and M. Matsui, editors. *Cryptographic Hardware and Embedded Systems - CHES 2006, 8th International Workshop, Yokohama, Japan, October 10-13, 2006, Proceedings*, volume 4249 of *Lecture Notes in Computer Science*. Springer, 2006.

[18] S. Hajra and D. Mukhopadhyay. Pushing the Limit of Non-Profiling DPA using Multivariate Leakage Model. Cryptology ePrint Archive, Report 2013/849, 2013.

[19] T. Katashita, A. Satoh, T. Sugawara, N. Homma, and T. Aoki. Development of side-channel attack standard evaluation environment. In *Circuit Theory and Design, 2009. ECCTD 2009. European Conference on*, pages 403–408, 2009.

[20] P. C. Kocher, J. Jaffe, and B. Jun. Differential Power Analysis. In M. J. Wiener, editor, *CRYPTO*, volume 1666 of *Lecture Notes in Computer Science*, pages 388–397. Springer, 1999.

[21] K. H. Lundberg. Noise sources in bulk cmos.

[22] S. Mangard. Hardware Countermeasures against DPA ? A Statistical Analysis of Their Effectiveness. In T. Okamoto, editor, *CT-RSA*, volume 2964 of *Lecture Notes in Computer Science*, pages 222–235. Springer, 2004.

[23] S. Mangard, E. Oswald, and T. Popp. *Power analysis attacks - revealing the secrets of smart cards*. Springer, 2007.

[24] T. S. Messerges, E. A. Dabbish, and R. H. Sloan. Investigations of Power Analysis Attacks on Smartcards. In *In USENIX Workshop on Smartcard Technology*, pages 151–162, 1999.

[25] D. Oswald and C. Paar. Improving Side-Channel Analysis with Optimal Linear Transforms. In S. Mangard, editor, *CARDIS*, volume 7771 of *Lecture Notes in Computer Science*, pages 219–233. Springer, 2012.

[26] J. R. Rao and B. Sunar, editors. *Cryptographic Hardware and Embedded Systems - CHES 2005, 7th International Workshop, Edinburgh, UK, August 29 - September 1, 2005, Proceedings*, volume 3659 of *Lecture Notes in Computer Science*. Springer, 2005.

[27] M. Rivain. On the Exact Success Rate of Side Channel Analysis in the Gaussian Model. In R. M. Avanzi, L. Keliher, and F. Sica, editors, *Selected Areas in Cryptography*, volume 5381 of *Lecture Notes in Computer Science*, pages 165–183. Springer, 2008.

[28] W. Schindler, K. Lemke, and C. Paar. A Stochastic Model for Differential Side Channel Cryptanalysis. In Rao and Sunar [26], pages 30–46.

[29] Y. Souissi, M. Nassar, S. Guilley, J.-L. Danger, and F. Flament. First Principal Components Analysis: A New Side Channel Distinguisher. In K. H. Rhee and D. Nyang, editors, *ICISC*, volume 6829 of *Lecture Notes in Computer Science*, pages 407–419. Springer, 2010.

[30] F.-X. Standaert, T. Malkin, and M. Yung. A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In A. Joux, editor, *EUROCRYPT*, volume 5479 of *Lecture Notes in Computer Science*, pages 443–461. Springer, 2009.

[31] A. Thillard, E. Prouff, and T. Roche. Success through Confidence: Evaluating the Effectiveness of a Side-Channel Attack. In G. Bertoni and J.-S. Coron, editors, *CHES*, volume 8086 of *Lecture Notes in Computer Science*, pages 21–36. Springer, 2013.

[32] W. A. Woyczynski. *A First Course in Statistics for Signal Analysis*. Birkhuser Boston, 2011.

Appendix A
Experimental Setup and Pre-processing

For all the experiments, we have used standard side-channel evaluation board SASEBO-GII [19]. It consists of two FPGA device: Spartan-3A XC3S400A and Virtex-5 xc5vlx50. Spartan-3A acts as the control FPGA where as Virtex-5 contains the target cryptographic implementation. The cryptographic FPGA is driven by a clock frequency of 2 MHz. During the encryption process, voltage drops across VCC and GND of Virtex-5 are captured by Tektronix MSO 4034B Oscilloscope at the rate of 2.5 GS/s i.e. $1,250$ samples per clock period.

The traces acquired using the above setup are already horizontally aligned. However, they are not vertically aligned. The vertical alignment of the traces are performed by subtracting the DC bias from each sample point of the trace. The DC bias of each trace is computed by averaging the leakages of a window taken from a region when no computation is going on. This step is also necessary since the derived impulse response of the matched filter is sensitive to the absolute value of mean leakages.

For mounting the attacks, we selected a window of 300 sample points around the last round register update. After transforming into a different domain, variance of some of the sample points may become very close to zero in the new domain. As a result, while applying approximate matched filter in this new domain, the weights (which are mean/variance of the sample points) of those sample points may become very high even if their mean values are very less. In other words, due to very low variance, some low SNR sample points may get very high weight. We solved this problem by setting the weight of a sample point having variance less than a fraction of $1/200$ of the maximum variance to zero.

Appendix B
Experimental Validation of the Multivariate Leakage Model

To validate Eq. 15, we first classify all the traces according to the values of I. Then we estimate the deterministic leakage $\mathbf{d}^i = \{E[L_t|I = i]\}_{t_0 \le t < t_0+\tau}$ for all $i \in \mathcal{I}$ by computing the mean leakage curve of each class. Lastly, we verify the linear equation $E[L_t|I = i] - E[L_t|I = 0] = a_t \cdot i$ for all $i \in \mathcal{I} \setminus \{0\}$ and $t_0 \le t < t_0 + \tau$ using linear regression. However, we do not know the values of a_t, $t_0 \le t < t_0 + \tau$. Thus, we start with correlating $\mathbf{d}^{i_1}$ and $\mathbf{d}^{i_2}$ for all $i_1, i_2 \in \mathcal{I}$ and then use the high correlation among them to estimate $\mathbf{a} = \{a_t\}_{t_0 \le t < t_0+\tau}$.

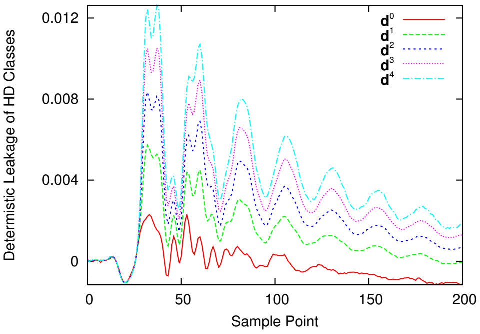


Fig. 7: Mean Leakage for the five Hamming distance classes across the 200 sample points.

We implemented an iterative structure of 32 parallel 10×4 S-boxes using distributed ROM in the setup described in Appendix A. All of the S-boxes were connected to the same input to increase the SNR of the power traces by the synchronous computations of the S-boxes. It should be noted that though the duplication of a single S-box increases the SNR of all the sample points, their relative SNR remains same. We collected $1,600$ power traces each having 200 sample points with random inputs. The values of the target variable S is taken to be the output of the S-box. We have also considered the Hamming distance model i.e. $\Psi(s)$ is taken to be the Hamming distance between s and the least significant 4 bit of the S-box input for all $s \in \mathcal{S}$. Since all the parallel S-boxes have the same input and the output, the algorithmic noise U is zero i.e. $I = P = \Psi(S)$.

The classification involves partitioning all the $1,600$ traces into five HD classes for $I = 0$ to 4. Fig. 7 shows the deterministic leakage curve $\mathbf{d}^i = \{E[L_t|I = i]\}_{t_0 \le t < t_0+\tau}$ for $0 \le i \le 4$ i.e. for each of the five classes. It is seen in the figure that the deterministic leakage for different HD classes i.e. different values of I are following almost same pattern. However, the non-zero leakage for HD class 0 is caused by the switching activities of the control bits and the DC power consumption which is also present in the leakages of all other classes. To remove this factor, we computed absolute deterministic leakage curves as $\bar{\mathbf{d}}^i = \mathbf{d}^i - \mathbf{d}^0 = \{E[L_t|I = i] - E[L_t|I = 0]\}_{t=t_0}^{t_0+\tau-1} = \{a_t \cdot i\}_{t=t_0}^{t_0+\tau-1}$ (from Eq. 15) for $i = 1, \cdots, 4$. Table I shows the correlation between $\bar{\mathbf{d}}^{i_1}$ and $\bar{\mathbf{d}}^{i_2}$ for all $i_1, i_2 \in \mathcal{I} \setminus \{0\}$. The values of these correlations are close to one which ensure that all of these

Correlation	$\bar{\mathbf{d}}^1$	$\bar{\mathbf{d}}^2$	$\bar{\mathbf{d}}^3$	$\bar{\mathbf{d}}^4$
$\bar{\mathbf{d}}^1$	1	0.9991	0.9981	0.9978
$\bar{\mathbf{d}}^2$	0.9991	1	0.9995	0.9992
$\bar{\mathbf{d}}^3$	0.9981	0.9995	1	0.9997

TABLE I: Pearson's correlation between absolute deterministic leakage curves of different pairs of HD Classes

vectors follow linear relations with a common vector namely $\mathbf{a} = \{a_{t_0}, \cdots, a_{t_0+\tau-1}\}$. We estimate $\mathbf{a}$ by $\frac{\sum_{i=1}^{4} \bar{\mathbf{d}}^i}{\sum_{i=1}^{4} i}$.

Next, we plot the vectors $\bar{\mathbf{d}}^i$ for all $i \in \mathcal{I} \setminus \{0\}$ against the estimated $\mathbf{a}$. The plot is shown in Fig. 8. The figure shows the linear relationships of $\bar{\mathbf{d}}^i$'s with the estimated $\mathbf{a}$. So, we have further used linear regression to find the closest linear models of the relation between each of $\bar{\mathbf{d}}_1$, $\bar{\mathbf{d}}_2$, $\bar{\mathbf{d}}_3$ and $\bar{\mathbf{d}}_4$ and the estimated $\mathbf{a} = \{a_t\}_{t=0}^{\tau-1}$. The relations obtained using linear regression are sufficiently close to the expected relation which are shown in Table II. This provides an evidence of the validity of Eq. 15.

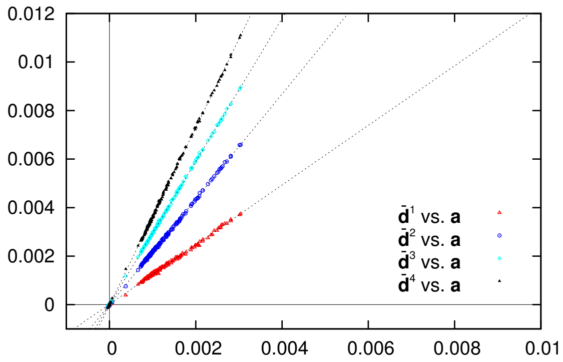


Fig. 8: Scatter Plots of $\bar{\mathbf{d}}^1$, $\bar{\mathbf{d}}^2$, $\bar{\mathbf{d}}^3$ and $\bar{\mathbf{d}}^4$ against $\mathbf{a}$.

Variable	Obtained Relation	Expected Relation
$E[L_t \| I = 1] - E[L_t \| I = 0]$	$a_t \times 1.23 - 1.60 \times 10^{-5}$	$a_t \times 1$
$E[L_t \| I = 2] - E[L_t \| I = 0]$	$a_t \times 2.17 - 7.26 \times 10^{-8}$	$a_t \times 2$
$E[L_t \| I = 3] - E[L_t \| I = 0]$	$a_t \times 2.95 - 1.41 \times 10^{-6}$	$a_t \times 3$
$E[L_t \| I = 4] - E[L_t \| I = 0]$	$a_t \times 3.65 - 1.75 \times 10^{-5}$	$a_t \times 4$

TABLE II: Relations of $\bar{\mathbf{d}}_1$, $\bar{\mathbf{d}}_2$, $\bar{\mathbf{d}}_3$ and $\bar{\mathbf{d}}_4$ with $\mathbf{a} = \{a_t\}_{t=0}^{\tau-1}$.