

Deconvolving Protected Signals

Mohaned Kafi

GEMALTO Security Labs
6, rue de la verrerie
92 190 Meudon
FRANCE

mohaned.kafi@gemalto.com

Sylvain Guilley

TELECOM ParisTech
46, rue Barrault
75 634 Paris Cedex 13
FRANCE

sylvain.guilley@telecom-paristech.fr

Sandra Marcello

THALES
160, bld de Valmy, BP 52
92 704 Colombes Cedex 04
FRANCE

samarcello@hotmail.com

David Naccache

École Normale Supérieure
45, rue d'Ulm
75 230 Paris Cedex 05
FRANCE

david.naccache@ens.fr

Abstract—The variable clock (VC) side-channel countermeasure consists in clocking a chip with an internal oscillator whose parameters (frequency, duty cycle, shape *etc.*) vary randomly in time. In this paper, we use parametric deconvolution to process VC-power consumption curves. We also analyze experimental results in order to show its efficiency.

I. INTRODUCTION

The variable clock (VC) side-channel countermeasure consists in clocking a chip with an internal oscillator whose parameters (frequency, duty cycle, shape *etc.*) vary randomly in time. Several authors attempted to work around VCs. An attacker can either try to re-synchronize individual power consumption curves (PCCs) [5], [2], [9] or – more rarely – apply DPA [6], [12] on protected signals and endeavor to fix the resulting differential curves [3].

In this paper, we use *parametric deconvolution* to process VC-protected PCCs. This is done by approximating the PCCs as the convolution of a (given) kernel function by an (unknown) irregular, sparse, comb of δ -functions. The PD game consists in estimating as accurately as possible the comb's form (teeth locations and amplitudes).

Fig. 1 shows the effect of a randomly varying clock ($8\text{MHz} < \text{VC} < 14\text{MHz}$) and the effect of a stable 3.57MHz clock on a smartcard's PCC. Note that despite the VC's irregularity, the typical (hardware-dependent) cycle shape doesn't fundamentally change. Periodicity disappears at the current spectrum level and a much larger spectrum of an 8 to 14MHz bandwidth appeared (central frequency of 10MHz), this confirms well the clock's fluctuation.

The minimal (Shannon) sampling frequency is 30MHz (spectrum's width of 15MHz). To overcome artifacts we sampled signals at 200 MHz to get an optimal temporal resolution.

The power traces synchronization technique used in this paper is *parametric deconvolution* (PD). PD will allow us to estimate the parameters used in our model. We will first define our model, and explain the deconvolution process. Then we will analyze our method with the help of an example and assess its experimental effectiveness.

II. MODEL AND PARAMETRIC DECONVOLUTION

A. Modeling The Power Consumption Signals

The power consumption model is illustrated Fig. 2. Fig. 3

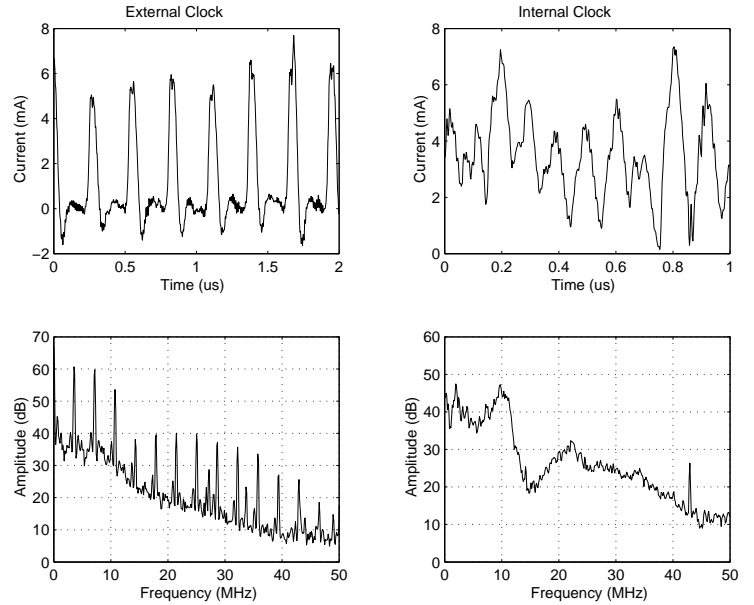


Fig. 1. Smart-card PCC (signal and spectrum); Normal clock versus VC

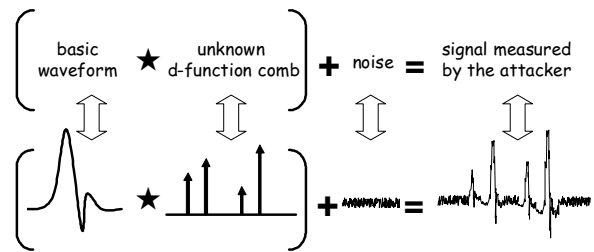


Fig. 2. Modeling of Current Consumption

shows the typical power consumption waveform during a clock cycle.

We will model the power consumption as a convolution product of an unknown signal by a known kernel. This kernel is the basic power consumption waveform shown in Fig. 3. To better reflect reality we add a white noise component to our model.

We assume that PCC signals $x_i(j)$ follow the model

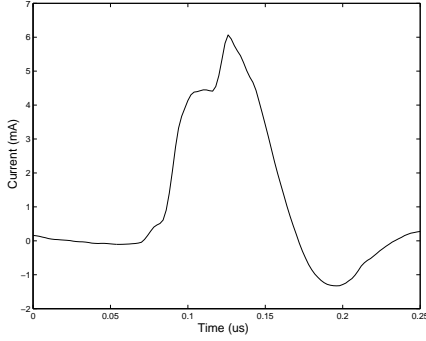


Fig. 3. Typical Cycle Power Consumption Waveform at 3.57 MHz.

$$\mathcal{M}(\omega, \tau_{ik}, \alpha_{i,k})$$

$$x_i = \omega \star \phi_i + \xi_i \quad (1)$$

$$\phi_i(j) = \alpha_0 + \sum_{k=0}^n \alpha_{i,k} \delta(j - \tau_{i,k})$$

where

- ω is the typical (Fig. 3) waveform observed during a clock cycle.
- ξ_i are centered gaussian (variance σ^2) noise components.
- $\tau_{i,k}$ is the time shift of each cycle with respect to the origin zero.
- $\alpha_{i,k}$ are amplitudes of successive waveforms $\omega(j - \tau_{i,k})$.
- δ is a Dirac spike

This can also be written as:

$$x_i(j) = \alpha_0 + \sum_{k=0}^n \alpha_{ik} \omega(j - \tau_{i,k}) + \xi_{ij}$$

The amplitudes $\alpha_{i,k}$ depend on the energy dissipated during each clock cycle. The duration separating two successive peaks is the duration separating the global maxima of two successive cycles.

The signals $y_i(j)$ represent what would have been obtained if the clock would have been regular. We use the model $\mathcal{M}(\omega, k\tau_0, \alpha_{ik})$ for the signal's reconstruction $y_i(j)$:

$$y_i(j) = \alpha_0 + \sum_{k=0}^n \alpha_{ik} \omega(j - k\tau_0)$$

Our goal is to restore the same DPA working conditions as if the VC would not have existed.

III. PARAMETRIC DECONVOLUTION

A. Linear Time Invariant system

Let \mathcal{S} denote a Linear Time-Invariant (LTI) system (Fig. 4) of impulse response $h(t)$. If one inputs into \mathcal{S} a signal $x(t)$ then a signal $y(t)$ appears at \mathcal{S} 's output. $y(t)$ is nothing but the convolution product of $h(t)$ by $x(t)$.

Given that in all attacks, the signals $x(t)$, $y(t)$ and $h(t)$ are sampled, we will substitute all continuous signals by the

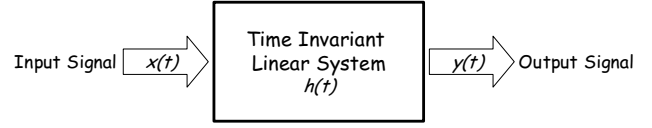


Fig. 4. Linear time-invariant system's representation.

quantized discrete sets $x(j)$, $y(j)$ and $h(j)$. The quantization's effect on the convolution equation is: $y(j) = \sum_{k \in \mathbb{Z}} x(k) h(j - k)$.

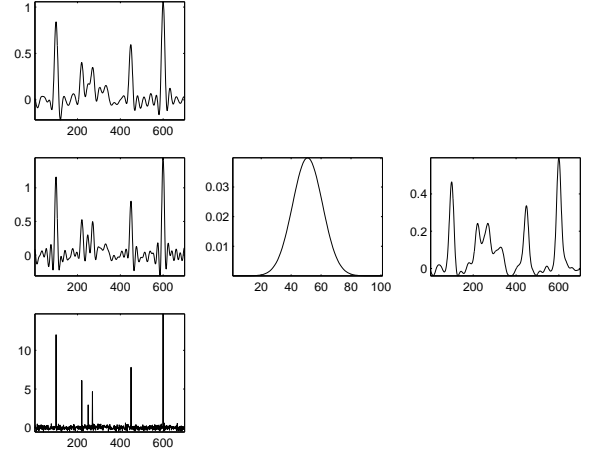


Fig. 5. Deconvolution is an ill-posed problem.

In most experimental situations one tries to recover the entry signal $x(t)$ from a measured output signal $y(t)$ and an approximation of the impulse response $h(t)$. In such instances we are confronted with an *inverse problem* and, more specifically, a *deconvolution* problem.

Deconvolution is an ill-posed problem *i.e.* a given deconvolution problem instance usually admits several solutions. To illustrate this, observe that the leftmost column Fig. 5 contains three different signals : $x_1(j)$, $x_2(j)$ and $x_3(j)$. The convolution of any of these with a gaussian kernel function represented in the central column yields the same signal $y(j)$ represented on the rightmost column.

This simple example illustrates the difficulty of solving deconvolution problems and the fact that one needs to take into account additional criteria to obtain a "good" solution. Such a process of adding such constraints is called *regularization*. More on deconvolution can be found in [7],[8]. To estimate the parameters (τ_k, α_k) of our model $\mathcal{M}(\omega, \tau_k, \alpha_k)$, we proceed as follows: we first apply Hunt's deconvolution, estimate the parameters τ_k and then the parameters α_k .

B. Hunt's Deconvolution

We use our model $\mathcal{M}(\omega, \tau_{ik}, \alpha_{ik})$ (§II-A) for $x_i(j)$. In the sequel we will omit the index i for x_i , ϕ_i and ξ_{ij} . With these notations Eq. (1) becomes

$$x(j) = \sum_{l=0}^{2p} \omega(l) \phi(j-l) + \xi_j \quad \forall j \in \{0 \dots n\} \quad (2)$$

Using matrix notation Equ. (2) becomes: $X = \Omega \Phi + \Xi$.

On several examples, we observed that the maximum likelihood criterion yields the same solution obtained by least-squares estimation. We will hence use least-squares.

$$J(\Phi) = \|X - \Omega\Phi\|^2 = (X - \Omega\Phi)^T(X - \Omega\Phi)$$

The solution set in the least-squares sense is non-empty, closed and convex, so there exists a unique explicit solution with minimal norm that we will denote $\hat{\Phi}$.

C. Regularization

In our practical case, we observe that $\Omega^T\Omega$ is not well conditioned (often the case for inverse problems) so we need to regularize $J(\Phi)$. We use Philips and Twomey's regularization criterion [11] (discrete case of Tikhonov's). Let $\lambda \in \mathbb{R}^{+*}$, $k \in \mathbb{N}$ and D_k be an order k differential operator.

$$\begin{aligned} J_\lambda(\Phi) &= \|X - \Omega\Phi\|^2 + \lambda \|D_k\Phi\|^2 \\ &= (X - \Omega\Phi)^T(X - \Omega\Phi) + \lambda (D_k\Phi)^T(D_k\Phi) \end{aligned}$$

As the signals are quite irregular, we will choose $D_0 = I = D$ for our application. The solution will be $\hat{\Phi}_\lambda$ (sometimes written $\hat{\Phi}$). We first find a solution depending on λ and then choose the optimal value for λ .

We will minimize the criterion $J_\lambda(\Phi)$, by zeroing the gradient, $\nabla J_\lambda = 2\Omega^T(X - \Omega\Phi) + 2\lambda D^T D\Phi = 0$ i.e

$$\hat{\Phi}_\lambda = (\Omega^T\Omega + \lambda D^T D)^{-1} \Omega^T X$$

In practice the major difficulty is the cost of inverting the matrix $\Omega^T\Omega + \lambda D^T D$. To overcome this difficulty we resort to Hunt's method, or circulant approximation, which is based on the use of the Fast Fourier Transform (FFT).

1) *Hunt's Algorithm:* $D\Phi$ is a linear filtering operation.

$$D\Phi = d \star \phi \quad \text{where} \quad d = \delta_0 = [1]$$

The first step is to extend the vectors x , ϕ and ω with zeroes to get vectors of same size $m \geq n + 2p$: $x_e(j)$, $\phi_e(j)$, $\omega_e(j)$. Then $X_e = \Omega_e \Phi_e$ where Ω_e is a circulant matrix, that we can diagonalize by an FFT. We have:

$$\Omega_e = F \Lambda_h F^{-1} \quad \Lambda_h = \text{diag}(\lambda_0^h \dots \lambda_m^h)$$

with $F_{k,l} = \exp^{j2\pi \frac{kl}{m}}$, $F_{k,l}^{-1} = \frac{1}{m} \exp^{-j2\pi \frac{kl}{m}}$

$$(\lambda_0^h, \dots, \lambda_m^h) = \text{FFT}(\omega(0), \dots, \omega(2p), 0, \dots, 0)$$

Similarly, we can extend vector $d = [1]$ with zeroes: $d_e(j)$. We diagonalize by a FFT the matrix D_e (where $D_e\Phi_e = d_e \star \phi_e$) such that:

$$\begin{aligned} D_e &= F \Lambda_d F^{-1}, \\ \Lambda_d &= \text{diag}(\lambda_0^d \dots \lambda_m^d), (\lambda_0^d, \dots, \lambda_m^d) \\ &= \text{FFT}(d(0), 0, \dots, 0) \end{aligned}$$

So we can write the following relation:

$$\hat{\Phi}_e = (\Omega_e^T \Omega_e + \lambda D_e^T D_e)^{-1} \Omega_e^T X_e$$

Writing $\hat{\Phi}(\nu)$, $\Omega(\nu)$, $D(\nu)$ and $X(\nu)$ for the FFT of ϕ_e , ω_e , d_e and x_e we get :

$$\hat{\Phi}(\nu) = \frac{1}{\Omega(\nu)} \times \frac{|\Omega(\nu)|^2}{|\Omega(\nu)|^2 + \lambda |D(\nu)|^2} \times X(\nu) \quad (3)$$

Empirically, the computed solutions (Eq. 3) show a big sensibility to λ 's value, so we need to find an optimal value for λ .

2) *Parameter Optimization Using The L-Curve Method:* To find the optimal parameter λ_{opt} , we use the L-curve method [4] which is based on the distance criterion.

The L-curve is usually used with a logarithmic scale i.e. the parametric curve is (ρ, η) where: $\rho = \log \|X - \Omega\Phi\|^2$, $\eta = \log \|D\Phi\|^2$. On Fig. 6 (computed using real experimental data) we can observe the characteristic form in L.

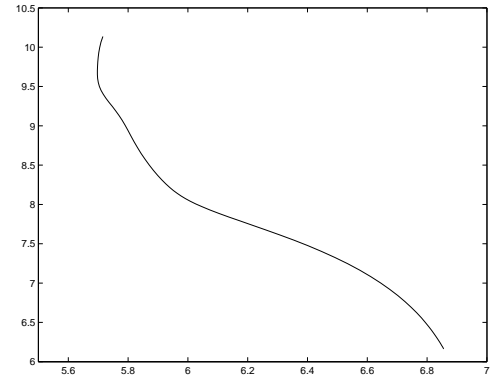


Fig. 6. Card Power Consumption : The L-Curve.

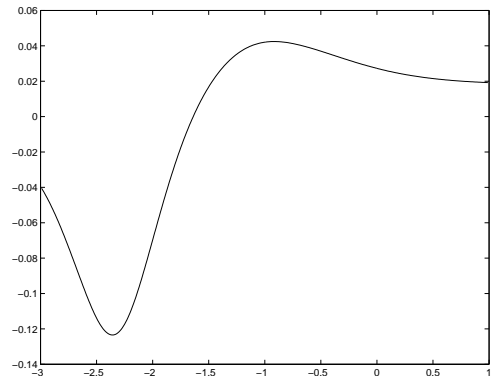


Fig. 7. The L-Curve's Curvature for a card power consumption (real data, logarithmic scale.)

To find λ_{opt} , we compute the L-curve's curvature $\kappa = 2 \frac{\dot{\rho}\ddot{\eta} - \ddot{\rho}\dot{\eta}}{((\dot{\rho})^2 + (\dot{\eta})^2)^{\frac{3}{2}}}$. When $|\kappa|$ is maximal, we obtain the value λ_{opt} . On Fig. 7, we can see that the curvature's absolute value is maximal for λ_{opt} close to $10^{-2.3} \simeq 0.005$.

3) *Shift Estimation* τ_k : In §II-A we define some notations, using them we define:

- $\hat{\tau}_k$: shift's approximation τ_k .
- $\hat{\delta}(j - \hat{\tau}_k)$: Dirac's peaks approximation $\delta(j - \tau_k)$.
- β_k : spike amplitudes $\hat{\delta}(j - \hat{\tau}_k)$.
- ε_j : errors due to $\omega(j)$'s approximation and to noise ξ_j .

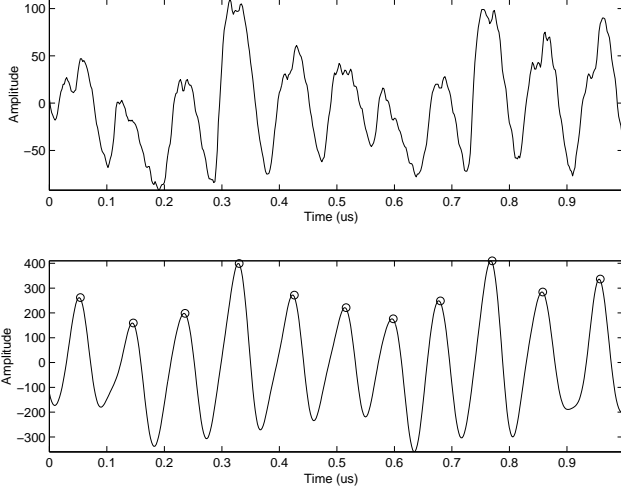


Fig. 8. Card's PCC signal (top) and signal after Hunt's deconvolution (bottom)

Even with an optimal regularization, the estimated signal $\hat{\phi}(j)$ is not a Dirac's peaks distribution with interval $\tau_k - \tau_{k-1}$ following $\mathcal{M}(\omega, \tau_k, \alpha_k)$ (§II-A).

After the Hunt's deconvolution, we obtain the following signal:

$$\hat{\phi}(j) = \beta_0 + \sum_{k=0}^n \beta_k \hat{\delta}(j - \hat{\tau}_k) + \varepsilon_j$$

We also should have, $\|\beta_k \hat{\delta}\| \approx \|\alpha_k \delta\| = |\alpha_k|$. So, using the computed $\hat{\phi}(j)$'s maxima, we will find the shifts $\hat{\tau}_k$. But there exist some peaks $\hat{\delta}(j - \hat{\tau}_{k^*})$ which don't correspond to "real peaks". These are the model's limits and there is a real need to fix these errors.

To that end, we propose a heuristic method. The PCC's signal doesn't contain power cycles with a period lower then the VC's minimal period. Moreover, the false impulsions' errors due to the $\hat{\phi}(j)$ have, in general, amplitudes lower then adjacent impulsions amplitudes. Hence, we propose to eliminate all the impulse's peaks which generate a shift $\hat{\tau}_k$ verifying the following condition:

$$(\hat{\tau}_{k+1} - \hat{\tau}_k) + (\hat{\tau}_k - \hat{\tau}_{k-1}) < 2T_{\min}$$

and

$$\beta_k < \min(\beta_{k-1}, \beta_{k+1})$$

where $T_{\min} = \frac{1}{F_{\max}}$ and $F_{\max} = 14\text{MHz}$ is the VC's maximal frequency.

4) *Amplitude's Estimation* α_k : We recall Equ. (1) : $x(j) = \alpha_0 + \sum_{k=0}^n \alpha_k \omega(j - \tau_k) + \xi_j$. We will assume that $\|\omega\| = 1$. Let $\alpha = (\alpha_0 \dots \alpha_n)^T$, $\Psi = (\psi_{\alpha_0} \dots \psi_{\alpha_n})^T$, with $\psi_{\alpha_0} = 1$, and $\psi_{\alpha_k} = \omega(j - \tau_k)$. So we can write

$$\langle x(j) - \xi_j, \Psi \rangle = \langle \Psi, \Psi^T \rangle \alpha \quad (4)$$

where

$$\langle \Psi, \Psi^T \rangle = \begin{pmatrix} \langle \psi_{\alpha_0}, \psi_{\alpha_0} \rangle & \dots & \langle \psi_{\alpha_0}, \psi_{\alpha_n} \rangle \\ \vdots & \ddots & \vdots \\ \langle \psi_{\alpha_n}, \psi_{\alpha_0} \rangle & \dots & \langle \psi_{\alpha_n}, \psi_{\alpha_n} \rangle \end{pmatrix}$$

We follow Lei and Speed [7]. Let $\hat{\alpha}$ be the solution, to find the solution of (4) using the maximum likelihood principle, we can use Gauss-Newton's algorithm [10]. In practice it suffices to iterate the algorithm twice (at most), to obtain good approximations of $\hat{\alpha}$. For this algorithm, we need to solve a linear system for each iteration. For a symmetric positively defined matrix, the conjugate gradient algorithm [1] is very efficient, the matrix $\langle \Psi, \Psi^T \rangle$ is of this kind so we use this algorithm. The impulse answer $\omega(j)$ is of compact basis and the maximal time between two cycles is $T_{\max} = \frac{1}{8\text{MHz}}$, the VC's maximal time. We infer that matrix $\langle \Psi, \Psi^T \rangle$ is sparse:

$$\forall l > 0 \quad \exists k_1 < l \quad \forall k \leq k_1 \wedge k \neq 0 \quad \langle \psi_{\alpha_l}, \psi_{\alpha_k} \rangle = 0$$

and, $\exists k_2 > l \quad \forall k \geq k_2 \quad \langle \psi_{\alpha_l}, \psi_{\alpha_k} \rangle = 0$

IV. PARAMETRIC DECONVOLUTION: AN ILLUSTRATION

A. Synthetic Data

Let $x(t)$ be the continuous signal defined by

$$\begin{aligned} x(t) = & 0.5 + \omega(t - 1.3) + 1.25 \omega(t - 1.6) \\ & + 1.25 \omega(t - 1.9) + \omega(t - 3.2) \\ & \dots + 1.25 \omega(t - 3.7) + 1.1 \omega(t - 4.2) \\ & + 1.25 \omega(t - 5.7) + \xi_t, \quad \forall t \in [0, 2\pi] \end{aligned}$$

where ξ_t is a centered gaussian white noise of variance σ^2 and

$$\omega(t) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{t^2}{2\sigma^2}} & \forall t \in [-4\sigma, 4\sigma] \\ 0 & \forall t \notin [-4\sigma, 4\sigma] \end{cases}$$

where $\sigma_\omega = \frac{1}{8}$.

We sample the signal $x(t)$ at $t_j = \frac{2\pi j}{1024}$ where $j \in \{0 \dots 1023\}$. The resulting signal $(x(j))_{j \in \{0 \dots 1023\}}$ follows the model $\mathcal{M}(\omega, \tau_k, \alpha_k)$ ($k < 8$) where $w(j)$ is the sampled version of $w(t)$ at instants t_j and:

$$(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6, \tau_7) = (1.3, 1.6, 1.9, 3.2, 3.7, 4.2, 5.7)$$

and

$$(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7) = (0.5, 1, 1.25, 1.25, 1, 1.25, 1.1, 1.25)$$

Fig. 9 illustrates the importance of λ 's choice.

We can see that $\lambda = 0.5$ is not a good value (still a lot of the peak's noise). For $\lambda = 10, 100$ we observe the good peaks

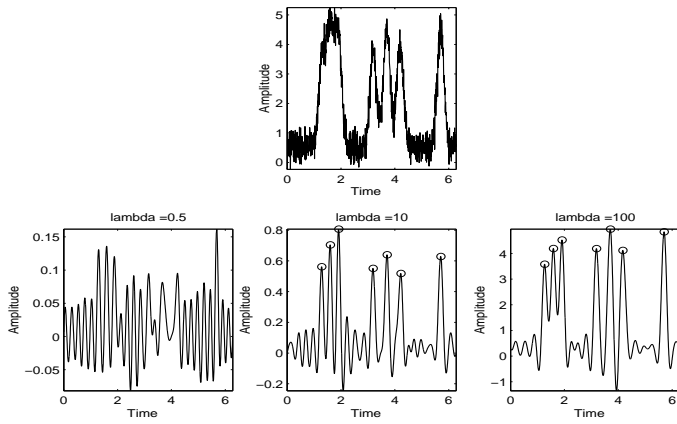


Fig. 9. Signal $x(j)$ (top) and Hunt's deconvolution results for $\lambda = 0.5, 10$ and 100 (bottom).

and narrower for $\lambda = 10$. The interpretation is that the bigger λ is the more important is the regularization.

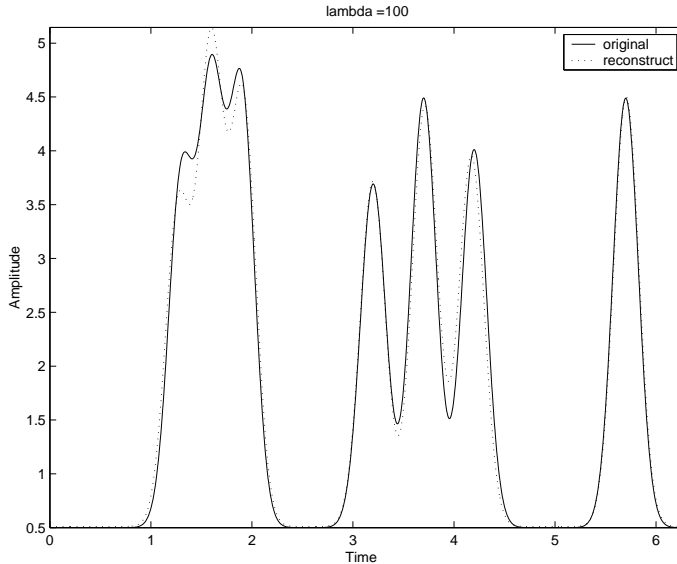


Fig. 10. Signal reconstruction $(x(j) - \xi_j)$ following the model $\mathcal{M}(\omega, \tau_k, \alpha_k)$ for $\lambda = 100$.

In Fig 10, we observe the original signal and its reconstruction for $(x(j) - \xi_j)$ with the model $\mathcal{M}(\omega, \tau_k, \alpha_k)$ for $\lambda = 100$. The reconstruction's errors are more significant for the first three cycles $\alpha_1\omega(j - \tau_1)$, $\alpha_2\omega(j - \tau_2)$ and $\alpha_3\omega(j - \tau_3)$. These cycles have an important temporal coverage. This is a lack of information which is needed for the cycles' reconstruction.

1) *Reconstruction Error Measurements:* We made some measurements of the estimated parameters (τ_k, α_k) for the signal $x(j)$ cf. to the table below.

With a constant $\sigma = 0.3$ noise, we observe the influence of the regularization parameter λ . We observe better bias cancellation for $\lambda = 10$ then for $\lambda = 100$. It appears that to select a good λ value we have to strike a trade-off between bias and standard-deviation for the estimators $\hat{\tau}_k$. Similar observations

apply to $\hat{\alpha}_k$, as well $\hat{\tau}_k$. With $\lambda = 100$ and changing the noise level, we observe that the estimators are robust in terms of bias. So we conclude that the estimators's $(\hat{\tau}_k, \hat{\alpha}_k)$ standard deviation is more or less stable from a covering standpoint. Only the $(\hat{\tau}_k, \hat{\alpha}_k)$ bias is strongly influenced by important coverings, e.g for the three first cycles.

		$\lambda = 10$ $\sigma = 0, 3$		$\lambda = 100$ $\sigma = 0, 3$		$\lambda = 100$ $\sigma = 0, 6$	
		bias $\times 10^4$	stand. dev. $\times 10^3$	bias $\times 10^4$	stand. dev. $\times 10^3$	bias $\times 10^4$	stand. dev. $\times 10^3$
τ_1	1.3	124	11	249	6	230	13
τ_2	1.6	17	8	38	6	21	12
τ_3	1.9	-105	9	-200	5	-205	10
τ_4	3.2	32	20	-89	9	-52	16
τ_5	3.7	7	15	0	7	2	13
τ_6	4.2	-38	16	65	7	92	15
τ_7	5.7	-12	13	6	6	-1	11
α_0	0.5	-13	13	-131	12	-101	27
α_1	1	-187	30	-347	24	-350	52
α_2	1.25	624	37	1177	24	1144	50
α_3	1.25	-120	30	-113	25	-212	52
α_4	1	-65	21	32	15	-16	37
α_5	1.25	-25	21	5	16	-70	32
α_6	1.1	-46	21	27	16	-59	28
α_7	1.25	-15	20	40	14	53	34

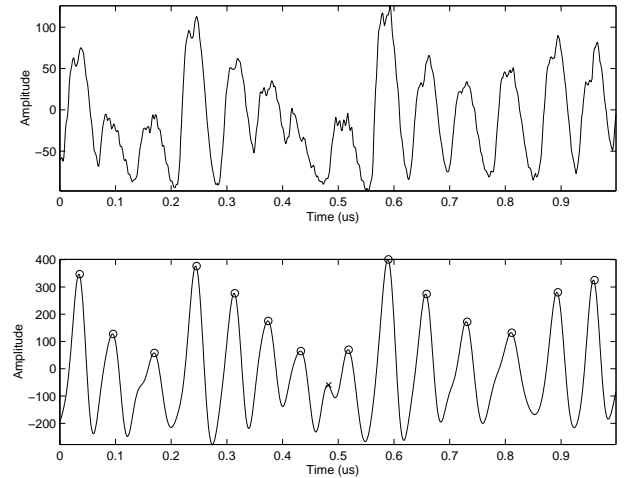


Fig. 11. Example: Hunt deconvolution (bottom) of a card PCC signal (top).

V. EXPERIMENTAL VALUES

The $x_i(j)$ values used in this section come from an experimental smart-card PCC acquisition, regularized using an optimal parameter λ_{opt} .

Fig. 11 shows the peaks tops (circled), with a statistical bias corresponding to the maximum of each clock cycle. But a false peak exists (cross marked). In other words our model doesn't reflect reality in full but an algorithm detecting false shifts τ_{k_0} corrects such errors. The false peak has been detected thanks to information known *a priori* on the PCC signal (§III-C3).

One can observe in Fig. 12 the signal $z(j)$ (plain line) with a noise level equal to zero and the estimators constructed with the PCC's signal $x(j)$ (dotted line).

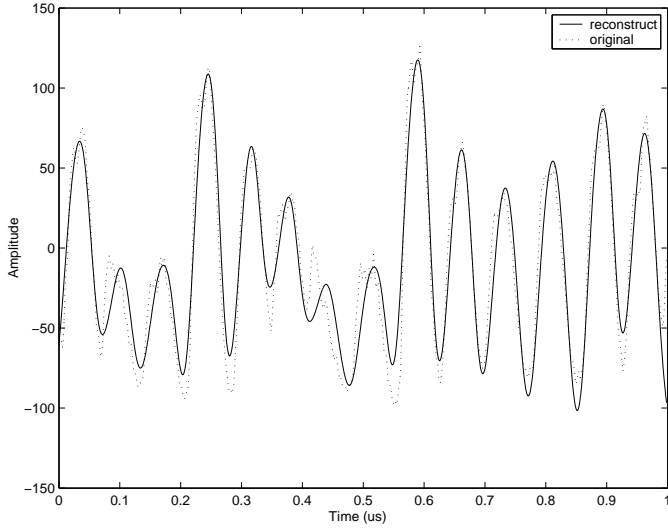


Fig. 12. Example : card PCC signal reconstitution.

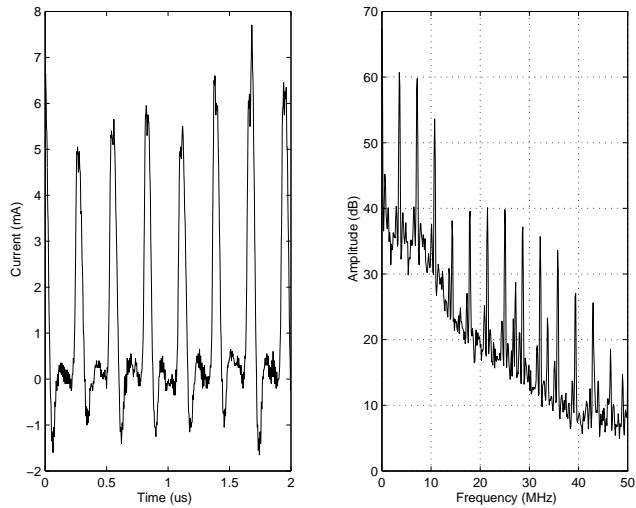


Fig. 13. Time and frequency-domain PCC features at 3.57MHz.

A. DPA on Deconvolved signals

We observe experimentally that a DPA on a VC-protected chip requires at least ten times as PCC many acquisitions as is necessary to conduct a DPA on the same chip wherein the VC was turned off.

1) *Experimental Conditions*: Fig. 13 shows $x(t)$, the typical PCC of a CMOS chip clocked at 3.57MHz. A quick look reveals that $x(t)$ breaks into fixed-period cycles having a similar shape and mainly varying in amplitude. The spectrum of $x(t)$ features equidistant spectral lines separated by 3.57MHz that, indeed, betray a fundamental 3.57MHz frequency.

As $x(t)$'s spectral band is rather large (circa 10MHz, i.e. a 20MHz) Nyquist frequency, all signal acquisitions reported here were over-sampled (100MHz) to avoid frequency overlaps, aliasing and other artifacts.

To reflect the efficiency of our signal processing process

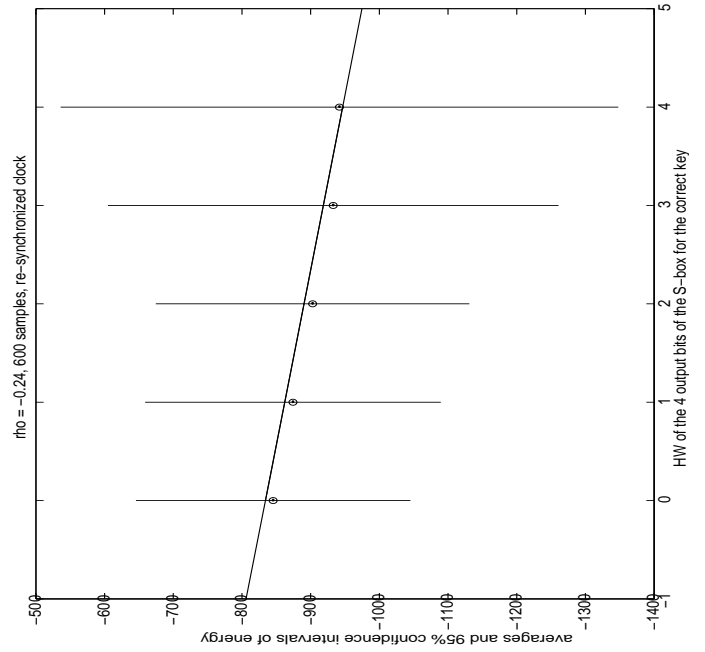


Fig. 14. Averages (circled points) and confidence intervals of $(S_i(j^*))_{i \in 1 \dots N}$ as a function of $\text{HW}(\bar{D}_i, K_{\text{true}})$ in the case of a re-synchronized VC

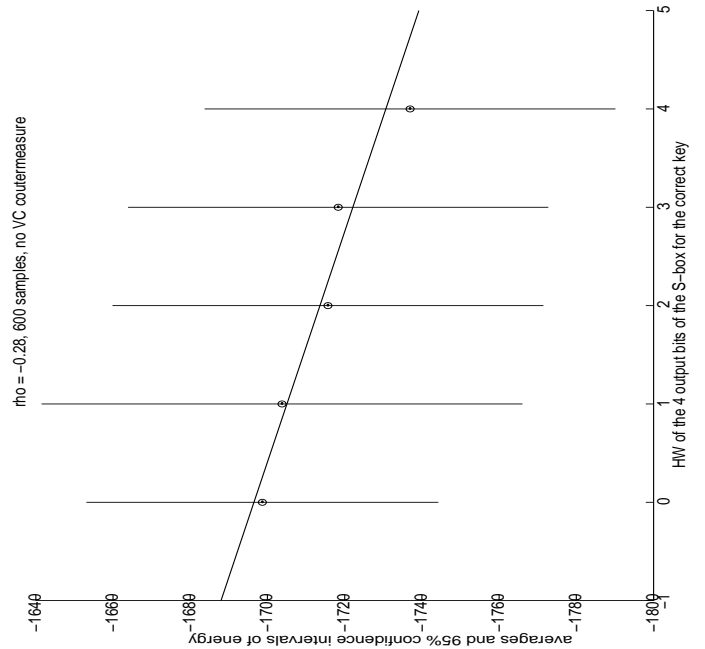


Fig. 15. Averages (circled points) and confidence intervals of $(S_i(j^*))_{i \in 1 \dots N}$ as a function of $\text{HW}(\bar{D}_i, K_{\text{true}})$ in the absence of a VC

we compare the DPA bias (differential curve) obtained using with a switched-off VC and with deconvolved VC-protected PCCs. Performance is very good as results appear to be nearly identical in both cases.

2) *Linearity Test*: Finally, we wish to measure rigorously the VC's re-synchronization quality. Let $D_{i,K}$ denote the DPA's

selection function for the i -th plaintext and the key hypothesis K . We further denote by HW the Hamming weight of a binary word.

A multibit DPA tests the linearity between $\text{HW}(D_{i,k})$ and the energy curve $S_i(j)$ (in physical terms, the integral of $x_i(j)$).

We determine the instant j^* where the linearity between $\text{HW}(D_{i,k})$ and $S_i(j)$ is the best. The multibit DPA attack signal is maximal for $j = j^*$ and $K = K_{\text{true}}$ where K_{true} denotes the actual key present in the device.

We can choose Pearson's linear correlation coefficient as a statistic $T_{\tilde{K}}(j)$ to spot the correct the key,

$$T_{\tilde{K}}(j) = \rho_{\text{HW}(\bar{D}_{i,K_{\text{true}}}), S_i(j)} \quad \text{with} \quad \rho_{x,y} = \frac{\text{cov}(x,y)}{\sqrt{\text{var}(x)\text{var}(y)}}.$$

We observe the average and the 95% confidence intervals for the current $(S_i(j^*))_{i \in 1 \dots N}$ in the absence of a VC (Fig. 15) and with a re-synchronized PCCs (Fig. 14).

Fig. 15 and Fig. 14 show that linearity is well respected in both cases. However, the confidence intervals, for deconvolved PCC's, seem to increase when the complemented Hamming weight $\bar{D}_{i,K_{\text{true}}}$, corresponding to $S_i(j^*)$ decreases.

We conclude that the reconstruction of a VC-protected signal is degraded as the energy/cycle ratio decreases. In other words, the α_k estimators loses in precision as SNR decreases, which is – albeit – what one would reasonably expect.

CONCLUSION AND PERSPECTIVES

The experiments undertaken in this work reveal that parametric de-convolution efficiently defeats the VC countermeasure as implemented on the tested chip. Indeed, we managed to obtain, while applying this technique, DPA spikes identical to these obtained in the countermeasure's absence.

Overcoming limitations of the current consumption's model and further improvements of regularization techniques deployed are interesting directions for further research.

REFERENCES

- [1] R. Barrett, M. Berry, T.F. Chan, J. Demmel, J. Donato, J. Dongara, V. Eijkhout, R. Pozo, C. Romine and H. van der Vorst, "Templates for the solution of linear systems: building blocks for iterative methods", SIAM, book, 1994.
- [2] X. Charvet and H. Pelletier "Improving the DPA attack using Wavelet transform, <http://csrc.nist.gov/groups/STM/cmvp/documents/fips140-3/physec/papers/physecpaper14.pdf>
- [3] C. Clavier, J.-S. Coron and N. Dabbous, "Differential Power Analysis in the presence of hardware countermeasures", in Proc. of CHES 2000, LNCS 1965, 252-263, 2000.
- [4] P. C. Hansen and D. P. O'Leary, "The use of the L -curve in the regularisation of discrete ill-posed problems", SIAM Journal on Sci. Comp., 14, 1487-1503, 1993.
- [5] N. Homma, S. Nagashima, Y. Imai, T. Aoki and A. Satoh: "High-resolution side-channel attack using phase-based waveform matching", Proceedings of CHES 2006, LNCS 4249, 187-200, 2006.
- [6] P. Kocher J. Jaffe and B. Jun, "Differential power analysis", CRYPTO '99 Proceedings of Advances in Cryptology, Springer-Verlag, 388-397, 1999.
- [7] L. Li and T. P. Speed, "Parametric deconvolution of positive spike trains", Ann Statist. vol 28, No. 5 1279-1301, 2000.
- [8] L. Li and T. P. Speed, "Deconvolution of sparse positive spikes: is it ill-posed?", Berkeley Technical Report No. 586, 2000, <http://stat-www.berkeley.edu/tech-reports/586.ps.Z>.

- [9] D. Moyart and R. Bevan, "A method for resynchronizing a random clock on smart cards", Eurosmart 2001, <http://www.nmda.or.jp/nmda/ic-card/proceedings/30-1440-DMoyart.pdf>
- [10] J. Nocedal and S. Wright, "Numerical optimization", New York, Springer 1999.
- [11] A. Tarantola, "Inverse Problem Theory", SIAM, book, 2004.
- [12] Sei Nagashima, Naofumi Homma, Yuichi Imai, Takafumi Aoki and Akashi Satoh, "DPA Using Phase-Based Waveform Matching against Random-Delay Countermeasure", ISCAS 2007 (IEEE), pages 1807-1810.