Improving the DPA attack using Wavelet transform*

Xavier CHARVET, Herve PELLETIER[1]

[1] SAGEM DS – Avenue du Gros Chêne – 95610 Eragny-sur-Oise (France)
herve.pelletier@sagem.com

Abstract

Security of cryptographic systems might be considered under many points of view. Traditionally, security of cryptology is seen on an abstract way: we make sure that the algorithms do not present any mathematical weaknesses which would make a cryptanalysis possible. However, for a few years, in an extended security model, physical attacks which use potential vulnerabilities of the material implementations are taken into consideration. One of the most threatening physical attack is presently the DPA attack (Differential Power Analysis), invented by Kocher in 1998 and presented in [1]. In this model, the assailant finds the secret keys used by the algorithms by analysing electric signals of the current consumed by the smart card. From this time, many countermeasures have been proposed in order to make those attacks difficult or just impossible. One of the most popular countermeasures consists currently in varying the internal clock frequency, in order to make DPA inappropriate. In this article, we propose a new way to launch an attack upon the smart card despite the desynchronisation of signals. This approach consists in analysing the curves of electric current by wavelets, and then to resynchronize them thanks to a minimization algorithm which will be presented in the following. The last step consists in selecting only well resynchronised curves, using statistical sorting algorithms, and then to perform a classical DPA attack on this selection. The result is better than the one we would obtain without doing those operations and will be explained and justified with many graphs in this article.

1 Introduction

Many papers have been published on the side channel topic since the first publication by Paul Kocher. In most cases, such articles are focused on plausible attacks against cryptographic implementations. In fact, only few papers deal with the problem of the countermeasure effects, often the authors keep this problem secret. Power analysis is a particular and efficient type of side channel attack. By monitoring devices power consumption or electromagnetic emanation during operations and manipulation of data it is possible to collect information

*The work presented in this paper has been exclusively supported by SAGEM DS.

leakage about these data. If information leakage about the data can be directly observed this is called a SPA attack (Simple Power Analysis). In other cases, if it is necessary to use statistical method, this is called DPA attack (Differential Power analysis) [5]. In fact, the DPA needs to compute one mean on a large number of power consumption samples to establish one correlation between the data being manipulated (and depending on few key bits) and the information leakage. This kind of attacks supposes there is an observable difference in the power consumption when a bit is set or clear.

Countermeasures proposed for DPA may be classified in two groups: algorithmic countermeasures on one hand and physical countermeasures on the other hand.

Among the first category, we can mention the duplication method, invented by L.Goubin and J.Patarin [2]. Another method which has been proposed by M.-L.Akkar and C.Giraud [3] consists in masking all the intermediate variables by the same random variable. However, both methods are sensitive to an high-order DPA. Since then, L.Goubin and M.-L. Akkar proposed [4] a generic protection, but which would also not be completely secure.

Among the physical countermeasures, the most natural countermeasure consists in increasing noise in order to make the number of data acquisition very high. This method does not make a DPA impossible, but only more difficult and particularly slow. The second countermeasure which has been considered consists in varying the instant where the critical calculation is being performed from an acquisition to another. We can proceed by executing useless operations on a random number or by varying the internal clock frequency. In this way, the DPA computation (sum/subtract on the traces) will be on elementary instructions which have no connection to each other. We can also imagine physical countermeasures which consist in limiting leakages, but the practical realization might be tricky on smart cards with restricted size.

In the following, we focus our presentation on smart cards with a random internal CPU clock to desynchronize the power consumption. In this article we propose a real methodology to attack a recent component with one of its hardware countermeasure activated. We show how a wavelet analysis enables a smoothing of the power consumption curves in order to resynchronize themselves with a general process. In the final step we show its efficiency in the DPA process in relation to a classical DPA treatment.

2 Experimental plateform

2.1 Measurement setup

The basic equipment is a standard digital oscilloscope with a 500 MHz bandwith and a sample rate of 500 MSamples/s to measure the probe's output signal. This probe is an active differential probe to measure the differential voltage (and so the current consumption) through a shunt resistance on the power supply line of the smart card. This smart card reader power supply has been replaced with a voltage generator to reduce the signal noise. Moreover, to collect and store the power consumption curves, the oscilloscope is connected to a PC via a GPIB bus.

2.2 DES implementation

The standard algorithm DES is implemented in C without masking or blinding protections. In fact to speed up and facilitate the power consumption acquisition only the first round is entirely implemented. Then, this round is repeated 16 times instead of the other standard rounds. Consequently, our DPA partition function tries to guess the first bit of the Sbox's output after the last permutation. If the hypothesis is true a bias peak will appear on the DPA curves. DPA curves represent the difference between the average power consumption when the bit is set and when the bit is cleared. To easily manage the hardware countermeasure effects, a specific command has been implemented, in the card, to choose one security level among three. The three modes are:

- External clock,
- Internal clock,
- Internal clock plus hardware random clock frequency.

3 Wavelet basis

This mathematical theory, whose first outline goes back in the 50's to the works of Morlet in soil mechanics and Grossmann in physical wave mechanics, has known a spectacular progress in the last years. It makes possible a time-frequency analysis of a signal and enables in a more general way, an automatic analysis of the world of the transient phenomena. The most natural way to grasp a signal is to represent the components evolution during the time. At the beginning of the XIXth century, Fourier presents a new way to apprehend a periodical signal, by showing that such a signal is an infinite sum of sinusoids. This is always true for a signal of finite energy (in this case, the sum has to be understood as an integral). The Fourier transform is:

$$\forall f \in \mathcal{L}^2(\mathbb{R}),\ f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{i\omega t} \mathrm{d}\omega.$$

Each signal of finite energy is then a continuous sum of sinusoids weighted by coefficients $\hat{f}(\omega)$.

The principal drawback of a frequencial representation is that it completely masks timing information. Wavelet analysis is a solution for this problem. We only give in this article a short presentation of wavelets. Interested reader can refer to Mallat's book [6].

The mathematical definition of wavelet is rather simple and is given below:

Définition 1 (Wavelet) *A* wavelet *is a function ψ with an average value of zero,*

$$\int_{-\infty}^{\infty} \psi(t) dt = 0$$

normalized $\|\psi\| = 1$, and centered around $t = 0$.

In fact, a wavelet ψ is a function (waveform) of limited duration. A time-frequency atoms family might be made by dilating the wavelet ψ with a numerical factor s (scale), and then by doing a translation with u:

Définition 2 (Time-frequency atoms)

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s}),\ u \in \mathbb{R},\ s \in \mathbb{R}^+$$

form a time-frequency atoms *normalized family.*

We have now all we need to give the definition of the wavelet continuous transform:

Définition 3 (Wavelet Continuous Transform) *If* $f \in \mathcal{L}^2(R)$*, its* wavelet continuous transform *at time* u *and at scale* s *is*

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi^*(\frac{t-u}{s})dt$$

which is nothing else than the convolution product of f *and the function*

$$\overline{\psi}_s(t) = \frac{1}{\sqrt{s}}\psi^*(\frac{-t}{s}).$$

4 Smoothing and resynchronization

4.1 Denoising aspect

The principal idea, as we mentioned in the beginning, is at first to find a way to smooth acquisition curves in order to be able to fit them together. Wavelet analysis, by virtue of its correlation with multiresolution approximations, enables to approach precisely a signal at different scales that the user can choose. This is used for many applications in fields as various as image processing, where specific wavelets bases have been conceived (interested reader may refer to curvelets imagined by E.C. Candès [7]) or even in speech processing, mechanics, etc...

In this article the wavelet transform have been computed with the DWT (discret wavelet transform) algorithm (from the Matlab software). We have analysed wavelet transform in using several families of wavelet and it seems the best results are find with the "Symlet" family. Figures 1, 2 and 3 put together a few approximations of signals (from power traces) at different scales.

The Wavelet transform, at a coarse scale, displays the general pattern of the power consumption curves. In this case, in a first approximation, the noise has been removed.

First, we have tried to measure the effect of wavelet transform directly on the DPA computation. A classical DPA attack has been achieved with the lowest security mode to test the correct behaviour of our DPA function. Figure 4 represents this DPA attack (for 1000 acquisitions based on the DES described above) realised for 7 assumptions of subkeys, whose only one is correct. The curve which presents two typical peaks corresponds to the good assumption for the subkey. This graph gives an idea of the ratio signal/noise we get with our experimental setup. Then, we have applied wavelet tranform (Symlet wavelet family) on each power traces before the DPA computation. The result is presented on Figure 5. It appears the DPA peaks are more sharper and deeper. The ratio signal/noise is increased of 30 %. However this improvement is not sufficient (as will be explained in Section 5) to defeat hardware protection like random clock mechanism. In fact, we can use the denoising aspect of the power traces (after wavelet transorm) to resynchronize them before the DPA computation.

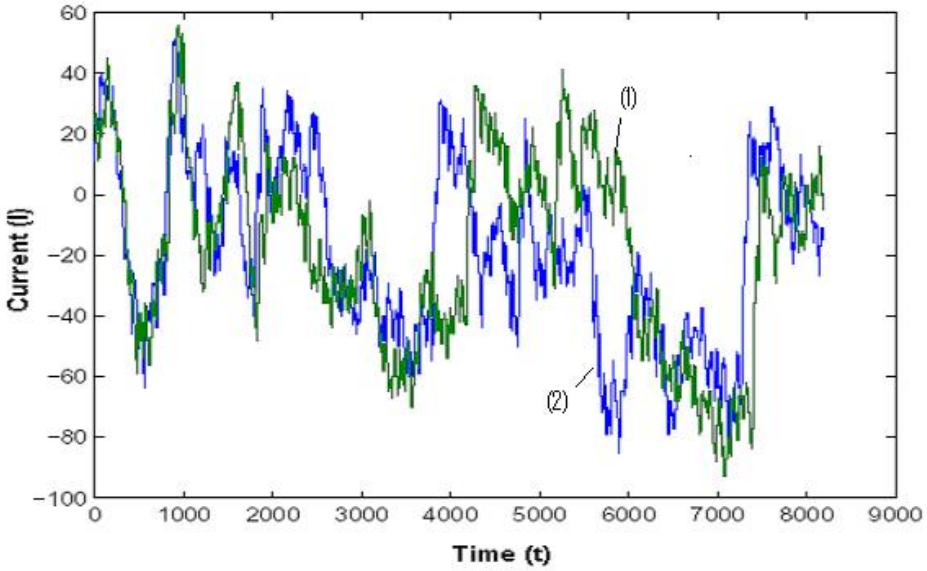

Figure 1: Two power consumption measurements taken during one round of the DES - internal random clock

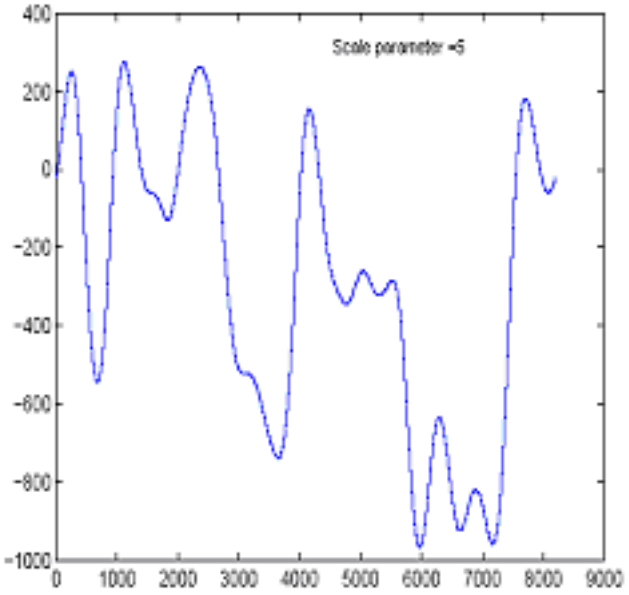

Figure 2: Approximation of the first signal at a coarse scale - internal random clock

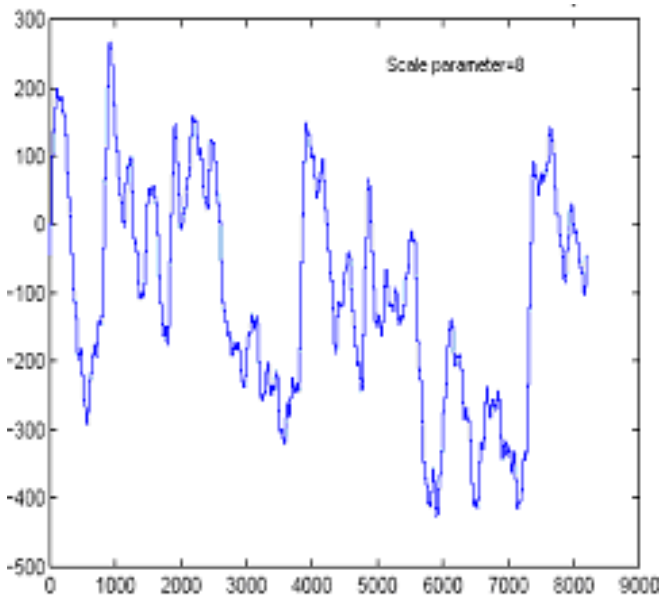

Figure 3: Approximation of the first signal at a fine scale - internal random clock

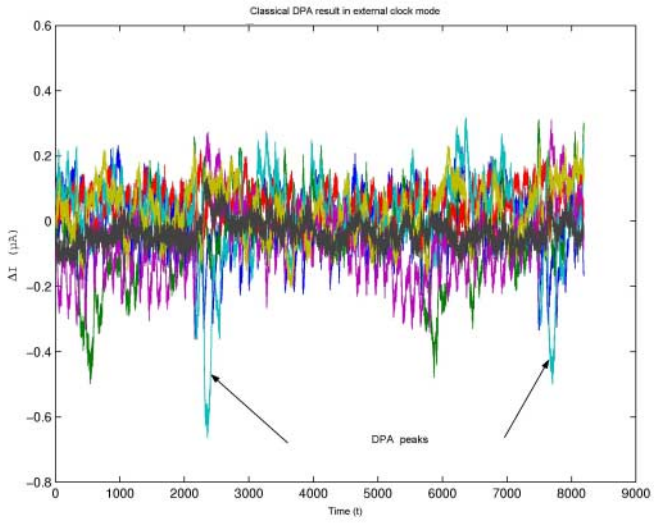

Figure 4: DPA attack on a smart card powered with an external clock .

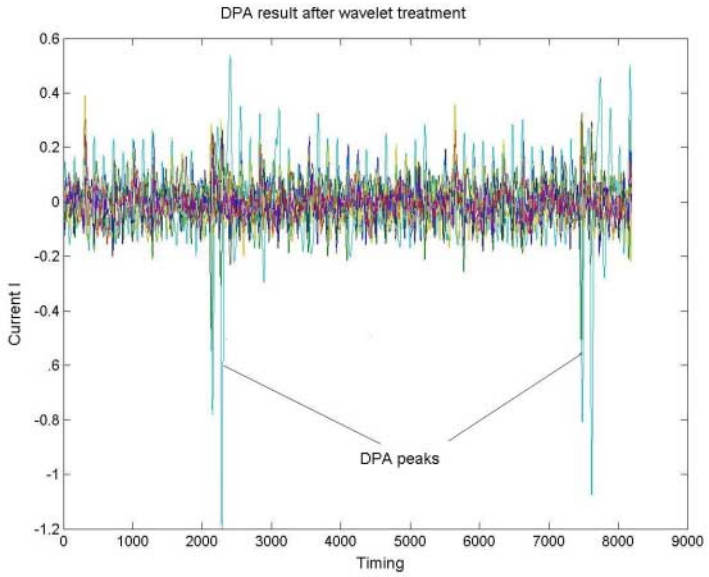

Figure 5: DPA attack on the wavelet tranform of the power traces.

4.2 Resynchronization problem

We assume in this section that we dispose of two curves which present a certain similarity in their structure. The first one will be taken as the "reference" and will be denoted by f_1, and the second one will be the curve to fit to the reference. The former one will be denoted by f_2. The problem may be outlined like this: find for each x axis of f_1 one axis $\tau(x)$ of f_2 such the covariant function $cov(f_1, \tilde{f}_2)$ be maximum with $\forall x\, \tilde{f}_2(x) = f_2(\tau(x))$.

Of course $\tau(x)$ must check some contraints:

1. $\tau(x)$ must be strictly decreasing. Indeed, if one event A occurs before a second event B, the order must remain unchanged after resynchronization.
2. In the same way $\tau(x)$ must not change brutally. In mathematical word $\frac{d\tau(x)}{dx}$ must be limited by a constant.
3. Reciprocally, it is not desirable that $\tau(x)$ stays constant on one interval, so $\frac{1}{\frac{d\tau(x)}{dx}}$ must be limited .

By this way, the resynchronization problem can be reduce to a classical minimization energy problem. Usually some additional constraints are incorporated in word of energy. For example, the constraints 2 and 3 can be modulated by a penalty coefficient in such a way that chosen functions do not show variations too slow or too abrupt. So the system energy can be modelized like this:

$$U(\tau) = \int (f_1(x) - f_2(\tau(x)))^2 + \alpha \max \tau'(x) + \beta \max \frac{1}{\tau'(x)} \qquad (1)$$

with α, β constants correctly chosen.

From this model, a minimization algorithm called "Simulated annealing" (SA) will be used. In fact, this algorithm tries to simulate the behaviour of a metal, in words of energy, when it is slowly cool down (if the temperature is abruptly decreased the metal will not crystallize in a configuration where the energy level is minimal).

4.3 The Simulated annealing algorithm

We let one set S of s_i sites, with $i \in \mathbb{N}$. Each site is associated with a descriptor x_s whith $x_s \in \mathbb{E}$. In this case, the set of configurations (in $\mathbb{E}^S$) is equal to Ω. Moreover, we can define a function U that for each configuration x, computes an associated energy $U(x)$.

If we want to find a minimum to this function $U(x)$, without estimating the energy for all the possible configurations, U must verify some properties. U must be the sum of "local" energy. By this word "local" we refer to energy computed on a subset of connected sites [9]. This subset of all neighbours which are pairwise adjacent is called a *clique* c and so C is the set of all c. If we call U_c this "local" energy on a subset c, the "Simulated Annealing"(SA) algorithm can be applied if we can write:

$$\forall x \in \Omega,\, U(x) = \sum_{c \in \mathcal{C}} U_c(x).$$

To use this SA algorithm it is necessary to quickly present the Gibb's sampler. It will be used as the "neighbour selection" algorithm. This algorithm defines the following distribution:

$$P(X = x) = \frac{1}{Z} \exp(-U(x))$$

with $U(x) = \sum_{c \in \mathcal{C}} U_c(x)$ and $Z = \sum_{x \in \Omega} \exp(U(x))$.

From an initial configuration x_0, this algorithm gives a configuration x such that $P(X = x) = \frac{1}{Z} \exp(-U(x))$. This algorithm is an iterative method to build each configuration. At step n:

- Select one site s.
- For the site s, according to the configuration of the neighbours V_s associated to the configuration x_s, it is necessary to compute the local conditional probability:

$$P(X_s = x_s | V_s) = \frac{-U_s(x_s|V_s)}{\sum_{\xi \in E}(\exp(-U_s(\xi|V_s)))}.$$

- The site s is updated with a random draw from the probability law $P(X_s = x_s|V_s)$.

This algorithm is iterated until the configuration is stable.

The SA algorithm selects a minimal energy among the set of possible configurations. It is an iterative algorithm based on the following steps. At the beginning, the initial state is characterized by an initial configuration x_0 and a "large" temperature T_0. For the step n:

- Simulation of one configuration x_n with the energy law $\frac{U(x)}{T^n}$ from the previous configuration x_{n-1}. Here the Gibb's sampler can be used for this simulation.
- Compute the difference $\Delta U = U(x_n) - U(x_{n-1})$.
- If $\Delta U < 0$ then the transition is accepted else this transition is accepted with a probability of $p = \exp(\frac{-\Delta U}{T})$.
- Decrease the temperature T (not so fast).

This process is repeated until the energy is minimal and remains stable. In this way, the algorithm must converge through a configuration of minimal energy. Such an algorithm allows to find a global minimum for the energy and not only a local minimum. Moreover this process is relatively independent of the initial configuration.

4.4 Experimental resynchronization example

Figure 6 shows, after wavelet transform (with the same scale), two power current traces taken from the same round of DES using a random internal clock. We

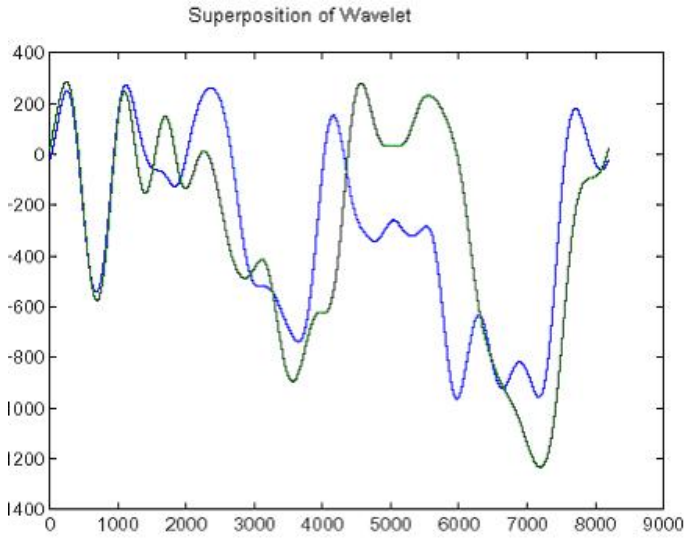

Figure 6: Superposition of two Wavelet Continuous Transform signal obtained in a random clock mode at a coarse scale.

can note that the shift between the two curves is not constant with the time. It is necessary to synchronize the curves by shifting them to the right or the left. Once the simulated annealing has been applied, Figure 7 is obtained.

In spite of initial differences between the two curves, the SA algorithm has clearly resynchronized these traces.

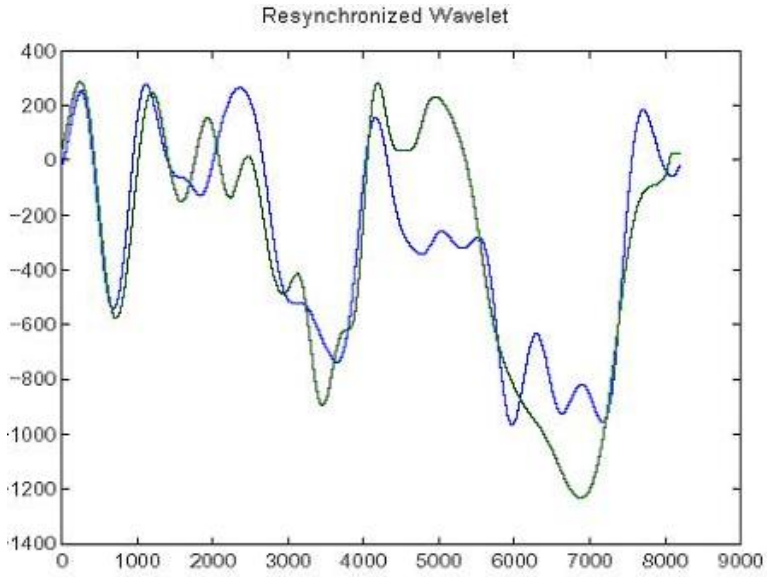

Figure 7: Superposition of the two Wavelet Continuous Transform after resynchronization by the "Simulated Annealing" .

5 Results

To test and check the efficiency of our method it was necessary to confront it with experimental results. In the following, we present some experimental DPA results according to the security level of the card and to the post processing treatment applied.

Let's take a closer look at a classical DPA attack against a card with the highest security mode activated (random clock oscillator). In this case the number of power consumption acquisitions has been augmented up to 100000 to try to reduce the noise. Figure 8 shows the result of this attack and it seems the DPA does not work, even if there is a large bias peak at the beginning it is not associated to the good key hypothesis. This result confirms the good efficiency of this countermeasure based on an internal random clock.

After this unsuccessful attack, in a first time, to get better results, we select only a part of all power traces acquired. By applying a simple correlation function to each power consumption curve it is possible to exclude the worst curves (the curves which are very different from others) of the DPA selection function. By this way we can reduce the noise injected in the DPA computation. It seems from Figure 9 this solution can increase the ratio signal/noise. However, sometimes, it is always insufficient to discriminate the good key hypothesis (the good hypothesis of key is not associated to the larger peak).

In order to increase the ration signal/noise it was interesting to estimate the efficiency of our method in this case. By applying wavelet transform on the same power traces used for the previous computation and consequently the simulated annealing to resynchronize them (before DPA computation). In this case, Figure 10 shows the result of this approach with a large bias peak associated to the good key hypothesis. This peak is always well characterized and largest than in the previous graph. This result shows clearly the advantage of our process. Nevertheless this method needs a largest amount of computations (several hours).

In fact, wavelet transform of the power curves shows the presence of specific pattern, probably the internal random clock does not completely mask the characteristic of the DES operation. In using wavelet transform (with a specific scale) it is possible to display this particular pattern on all curves. So we have developed a specific hand made algorithm (more fast as simulated annealing algorithm) to only resynchronize these patterns (and consequently the power curves) before applying the DPA computation.

Again, the ration signal/noise of the DPA bias is clearly increased by this approach in relation to the previous result.

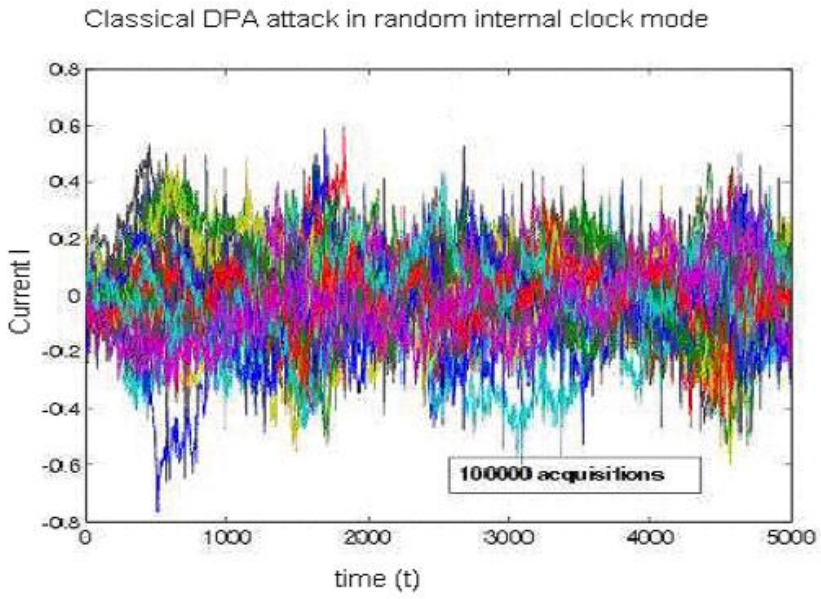

Figure 8: DPA attack on a smart card powered with the highest security level .

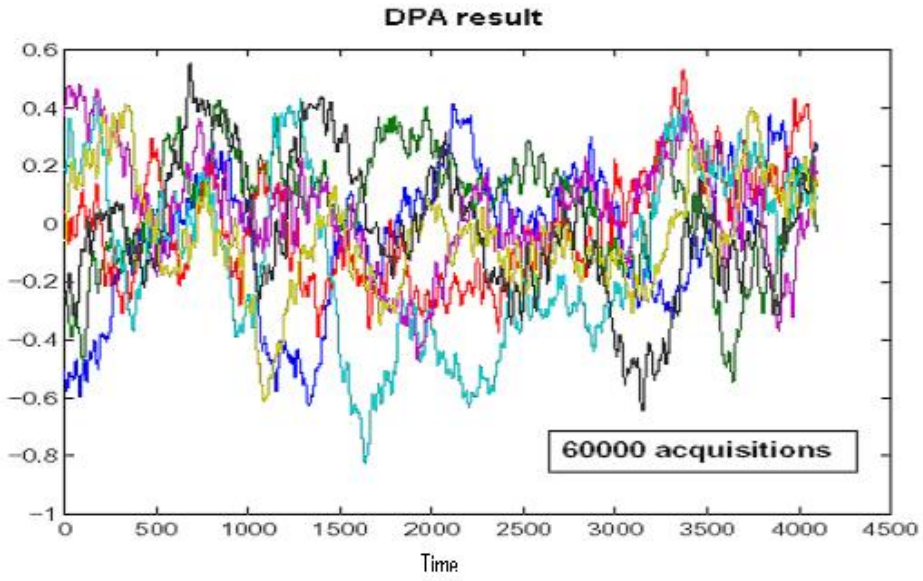

Figure 9: DPA attack on a smart card powered with the highest security level and after selection of the "good" traces (after a specific sort).

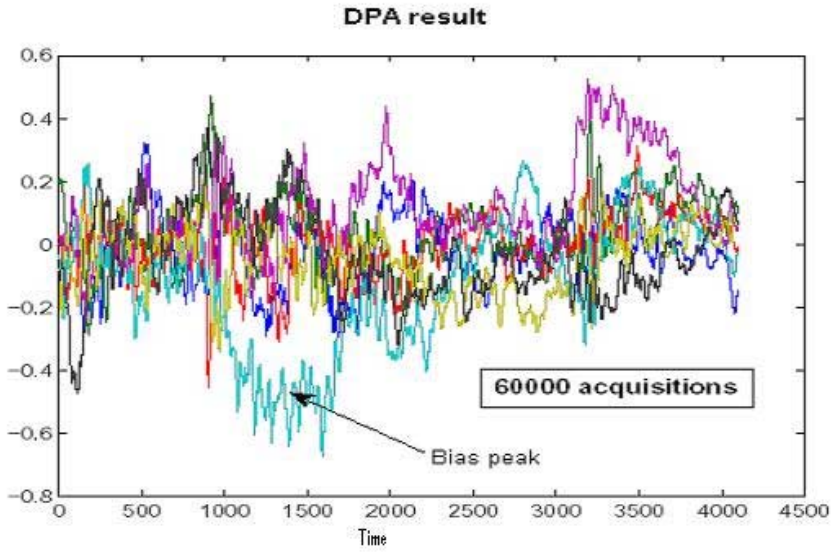

Figure 10: The same DPA attack as previous, after denoising(wavelet) and resynchronization(Simulated annealing) treatment.

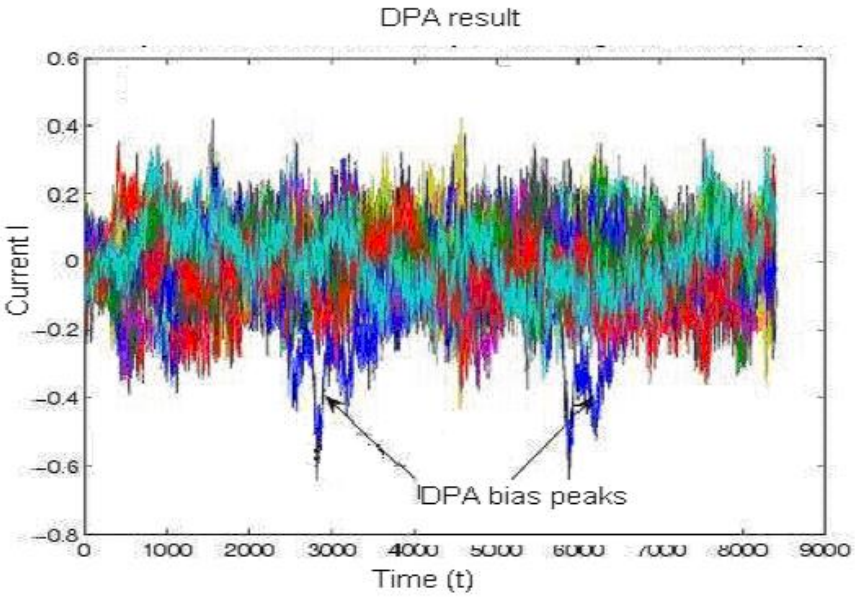

Figure 11: The same DPA attack as previous, after denoising(wavelet) and a specific resynchronization treatment

6 Conclusion

In this paper we have explored a general technique to increase the ratio signal/noise of the DPA attack by using an advanced alignment algorithm. This technique is based on the connection between a wavelet continuous transform and a general minimization energy function. We have shown, in a first time, the efficiency of such technique on a widely known countermeasure applied to a recent card. Even if better results can be obtained by using a specific resynchronization algorithm after wavelet transform, this algorithm is too specific to the characteristics of the smart card's implementation. Indeed, in this case, this resynchronization method is not applicable for others algorithms or other cards contrary to our general method. In fact, maybe, a general resynchronization algorithm applied to a multi scale wavelet transform can improve the efficiency of the DPA attack.

References

[1] P. Kocher, J. Jaffe, B. Jun. *Introduction to Differential Power Analysis and Related Attacks*. Technical Report, Cryprography Research Inc., 1998.

[2] L. Goubin. *DES and Differential Power Analysis - The "Duplication" Method*. In proceedings of CHES'99, LNCS 1717, pp.158-172, Springer-Verlag, 1999.

[3] M.-L. Akkar, C. Giraud. *An Implementation of DES and AES Secure against some Attacks*. In proceedings of CHES'2001, LNCS 2162, pp.309-318, Springer-Verlag, 2001.

[4] M.-L. Akkar, L. Goubin. *A Generic Protection against High-Order Differential Power Analysis*. In proceedings of FSE 2003, LNCS 2887, pp.192-205, International association for Cryptologic research, 2003.

[5] T.S. Messerges. *Power analysis attacks and countermeasures for cryptographic algorithms*. Doctoral Thesis, January 2000.

[6] S. Mallat. *A wavelet tour of signal processing*. Academic Press, 1998.

[7] E. J. Candès, D. L. Donoho (2000). *Curvelets, Multiresolution Representation, and Scaling Laws*. Wavelet Applications in Signal and Image Processing VIII, A. Aldroubi, A. F. Laine, M. A. Unser eds., Proc. SPIE 4119.

[8] E. J. Candès, D. L. Donoho (2000). *Curvelets and Reconstruction of Images from Noisy Radon Data*. Wavelet Applications in Signal and Image Processing VIII, A. Aldroubi, A. F. Laine, M. A. Unser eds., Proc. SPIE 4119.

[9] M. Sigelle. *Champs de Markov* In Analyse des Images, Chapitre 6, ENST, 2002.