# Novel Applications of Wavelet Transforms based Side-Channel Analysis

Youssef Souissi[1], M. Abdelaziz El Aabid[1], Nicolas Debande[1,2], Sylvain Guilley[1], Jean-Luc Danger[1]

(1) Telecom ParisTech, COMELEC, 75 634 PARIS Cedex 13, FRANCE.
(2) Morpho, 95 523 OSNY, FRANCE.
*This work is partially funded by the JST/ANR SPACES project.*
Email: {youssef.souissi, aziz.elaabid, nicolas.debande, sylvain.guilley, jean-luc.danger}@TELECOM-ParisTech.fr

**Abstract.** In this paper we propose, in a methodological manner, four different applications of Wavelet transforms in the Side-channel context. The proposed applications are involved in several Side-channel analysis aspects: acquisition of traces, patterns detection, noise filtering and secret-key recovery.

**Keywords:** Side-channel Preprocessing and Attacks, Time-frequency (multi-resolution) analysis, Wavelet transforms, Security evaluation, Secret key encryption.

## 1 Introduction

Recently, a new kind of threats called Side-Channel Analyses (SCA) have attracted much attention in embedded security areas. These analyses are serious concerns as they are able to retrieve the secret information from the cryptographic implementations without tampering with the system. Therefore, the need of securing and evaluating the robustness of embedded systems against such malicious attacks becomes obvious. From the security evaluation point of view, it is important to get rid off all possible external issues like the access to the device, which are not related to the cryptographic process itself, so as to make the evaluation in the best conditions. Moreover, it is better to know whether the deployed countermeasures can be broken by an attacker. Today's evaluator already has at his disposal several tools to properly analyse Side-channel traces. However, in the SCA literature, the analysis is usually performed either in the time or the frequency domain. Indeed, it is often reported that analysing a power or electromagnetic signal in the time domain is more efficient than analysing it for its frequency content, based on the Fourier transform. Nonetheless, a Fourier analysis turns out to be useful especially if the signals acquired are misaligned. The Fourier transform is only able to retrieve the global frequency content of a signal, thus the time information is lost. This is overcome by the short time Fourier transform (STFT) which computes the Fourier transform of a windowed part and shifts the window over the signal. The STFT gives the time-frequency content of a signal with a constant frequency and time resolution due to the fixed window length. This is unlikely the most appropriate resolution. For low frequencies, often a good frequency resolution is required over a good time resolution. Whereas, for high frequencies, the time resolution is more important. An alternative tool with some attractive properties is *the Wavelet transform*. Although Wavelet theory has been successfully applied in many applications in science and engineering, it has been rarely invoked in the SCA context. Actually, Wavelets have only been used in two occasions: first they are used to remove noise from Side-channel traces (which is already known) [1], and second used to estimate the probability density function of the leaked information [2]. In this paper, we propose new ways to allow the evaluator to get benefits from the multi-resolution

provided by Wavelet analysis. First, we show how Wavelets can be properly used to make real-time compression of Side-channel traces without loss of information. Indeed, it is known that, when acquiring Side-channel signals, cryptographic patterns need a high sampling rate to be detected by the existing means, especially when the implementation is protected. Therefore, a large memory depth and storage are required. Obviously, the goal of compressing SCA traces is to minimize the storage capacity needed to hold or to convey, without loss, the sensitive information. Second, we propose an improvement of the basic method of Pelletier et .al, proposed in [1] and used for Side-channel traces noise filtering. Indeed, we discuss the problem of noise filtering from the SCA context and propose to employ an information theoretic approach as a complementary tool and an enhancement for the basic method. Third, we highlight the ability of Wavelets to detect and extract the patterns of a cryptographic process when performing a simple Side-channel analysis (*i.e.*, a direct interpretation of the traces acquired). Eventually, we show that SCAs when performed with a multi-resolution analysis are much better, in term of security metrics, than considering only the time or the frequency resolution. Actually, the gain in number of traces needed to recover the secret key is about 50%, relatively to an ordinary attack. For this purpose, two attacks are considered: Correlation Power Analysis (CPA) and Principal Components Analysis (PCA) based template attacks, both performed on real acquired Side-channel traces. Besides, we note that such analysis is generic as it can be seen as a plug-in used to improve existing Side-channel attacks. The overall goal of this talk is about valuing, in a methodological manner, the use of Wavelet transforms in Side-channel Analysis — not only in the noise filtering preprocessing, but in the very core of the attack.

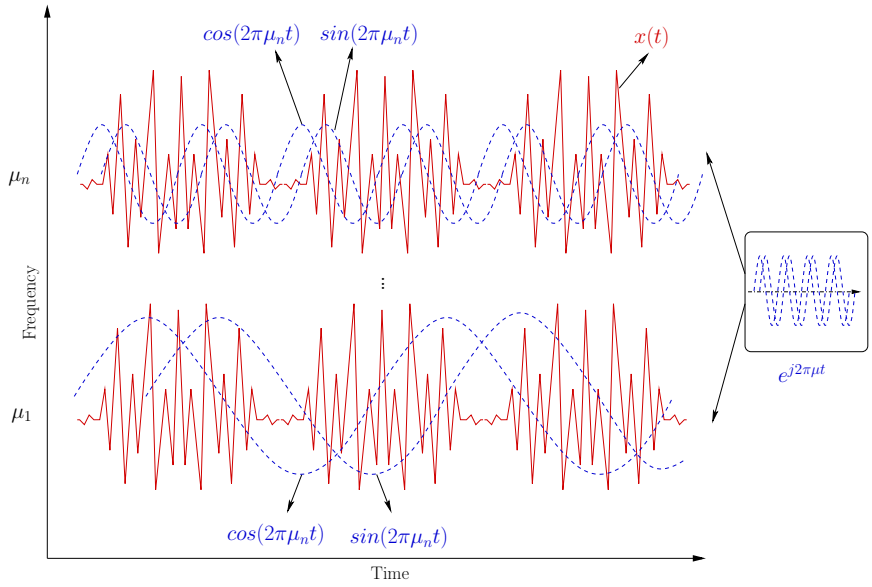


**Fig. 1.** Fourier transform illustration.

## 2 An understanding of the multiresolution analysis

In practice, signals acquired are usually analysed in the time domain. The information encompassed by the signals can be translated into different representations. Actually, thanks to these representations, more details about the signal, such as

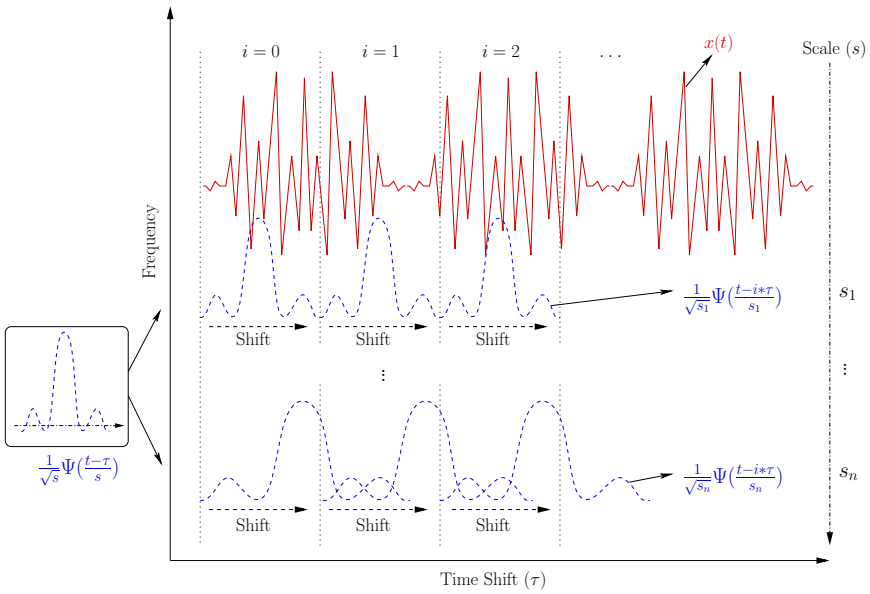


**Fig. 2.** Wavelet transform illustration.

the frequency contents, can be emphasized very nicely in order to highlight the hidden dynamics related to the cryptographic process. The most commonly used representation used to analyse a time signal for its frequency contents is the Fourier transform and its alternative the Short Time Fourier Transform (STFT). Understanding the Fourier transform and the STFT is necessary to understand the basics of the multiresolution analysis, which is often known as *the Wavelet transform*.

### 2.1 Fourier transform

The Fourier transform is a mathematical representation that decomposes a function exactly into many components, each of which has a precise frequency. From the mathematical point of view, any periodic function $f(t)$, with period $2p$, that is $f(t) = f(t + 2p) = ... = f(t + 2np)$ for $n = 1, 2, 3, ...$, can be expressed as a linear combination of all *cosine* and *sine* functions, which have the same period:

$$f(t) = \frac{1}{2}a_0 + \sum_{n=1}^{+\infty}(a_n cos_(\frac{n\pi t}{p}) + b_n sin_(\frac{n\pi t}{p})). \tag{1}$$

Where

$$a_n = \frac{1}{p}\int_{-p}^{p} f(t)cos(\frac{n\pi t}{p})dt; n = 0, 1, 2, 3, ...,$$
$$b_n = \frac{1}{p}\int_{-p}^{p} f(t)sin(\frac{n\pi t}{p})dt; n = 1, 2, ... \tag{2}$$
$$. \tag{3}$$

These series of sines and cosines are called *Fourier series*. Note that each individual term $cos_(\frac{n\pi t}{p})$ or $sin_(\frac{n\pi t}{p})$ is a periodic function, which period $T_n$ is determined by the relation that when $t$ is increased by $T_n$, the function returns to its previous value,

$$cos_(\frac{n\pi}{p}(t + T_n)) = cos_(\frac{n\pi t}{p} + \frac{n\pi T_n}{p}) = cos_(\frac{n\pi t}{p}). \tag{4}$$

Thus,

$$\frac{n\pi}{p}T_n = 2\pi \ , \ T_n = \frac{2p}{n}. \tag{5}$$

The frequency $\mu$ is defined as the number of oscillations in one second. Therefore, each of them is associated with a frequency $\mu_n$,

$$\mu_n = \frac{1}{T_n} = \frac{n}{2p}. \tag{6}$$

Often, the angular frequency, defined as $\omega_n = \frac{n\pi}{p} = 2\pi\mu_n$, is used to simplify the writing. The *Fourier series* can be generalized to complex numbers and further generalized to derive the Fourier transform. Actually, it can be shown that the *Fourier series* of a function repeating itself in the interval of $2p$ can be written as:

$$f(t) = \sum_{n=-\infty}^{+\infty} c_n e^{i\frac{n\pi}{p}t} \ , \ c_n = \frac{1}{2p}\int_{-p}^{p} f(t)e^{-i\frac{n\pi}{p}t}dt. \tag{7}$$

hence

$$f(t) = \sum_{n=-\infty}^{+\infty} [\frac{1}{2p}\int_{-p}^{p} f(t)e^{-i\frac{n\pi}{p}t}dt]e^{i\frac{n\pi}{p}t}. \tag{8}$$

Now, if we denote $\delta\omega = \omega_{n+1} - \omega_n = \frac{\pi}{p}$, hence the series can be expressed as:

$$f(t) = \sum_{n=-\infty}^{+\infty} [\frac{1}{2\pi}\int_{-p}^{p} f(t)e^{-i\omega_n t}dt]e^{i\omega_n t}\delta\omega = \sum_{n=-\infty}^{+\infty} \frac{1}{2\pi}\hat{f}_p(\omega_n)e^{i\omega_n t}\delta\omega. \tag{9}$$

where $\hat{f}_p(\omega_n) = \int_{-p}^{p} f(t)e^{-i\omega_n t}dt$.

The Fourier transform may be regarded as the formal limit of the Fourier Series as the period tends to infinity. Indeed, if we let $p \to +\infty$, then $\delta\omega \to 0$ and $\omega_n$ becomes a continuous variable. Hence, we have:

$$\hat{f}(\omega) = \lim_{p\to+\infty} \hat{f}_p(\omega_n) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}dt. \tag{10}$$

Besides, The original signal $f(t)$ can be reconstructed using the inverse Fourier transform as follows:

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)e^{i\omega t}dt. \tag{11}$$

In practice, signals acquired experimentally are not continuous in time, but sampled as discrete time intervals $\delta T$. Besides, T-boner length is finite with a total acquisition time $T$, divided into $N = \frac{T}{\delta T}$ intervals. In this case, the frequency analysis is conducted by using *the Discrete Fourier Transform* (DFT). Generally, the Fourier transform is useful at providing the frequency content that can not be easily detected in the time domain. However, the temporal structure of the analysed signal is not revealed; and therefore such transforms are not suitable for brief signals, signals that change suddenly (short bursts), or in fact any non-stationary signals. Actually, using Fourier transform does not reveal how the signal's frequency contents vary with time. This may limit the merit of Fourier transform specially when both time and frequency information are required.

### 2.2 Short Fourier Transform (STFT)

In order to overcome the limitations of the Fourier transform, one can introduce an analysis window of a certain length that shifts along the signal's time axis to perform a *time-localized* Fourier transform. This concept, which has been initially proposed

by D.Gabor in [3], is referred to as *Short Time Fourier Transform* (STFT). The STFT is capable to retrieve both frequency and time information from a signal. Technically, STFT employs a sliding window function $g(t)$ that is centered at time $\tau$; and can be expressed by the following equation:

$$f_{STFT}(\tau, \mu) = \int_{-\infty}^{+\infty} f(t) g^*(t-\tau) e^{-i2\pi\mu t} dt. \tag{12}$$

Where $*$ denotes the complex conjugated operator. When the window is moved by $\tau$, a time-localized Fourier transform is performed on the original signal $f(t)$ within the window. This way, the STFT decomposes a time domain signal into a two dimensions time-frequency representation. The mapping of the time domain onto a function of time and frequency provides some information about if a certain frequency is continuously present throughout the time observation or only at a specific time interval. However, using a sliding window with fixed length results in a new problem. Actually, the STFT analysis is critically dependent on the chosen window $g(t)$. Theoretically, it is not possible to know exactly what frequencies occur at what time, only a range of frequencies can be detected. In other words, it is not possible to get both a good time resolution and a good frequency resolution with STFT. In fact, on the one hand, a short window, which is suitable for high frequencies detection, provides a good time resolution, but several frequencies are not revealed. On the other hand, a long window, which is suitable for low frequencies detection, provides an inferior time resolution, but a better frequency resolution.

### 2.3 Wavelet transform

We have seen that the STFT is not sufficient to describe, with accuracy, both the time and the frequency content of a signal acquired. Recently, a powerful tool called *Wavelets analysis*, has been introduced to overcome the limitation of the STFT. In contrast to STFT, where the sliding window size is fixed, the Wavelet transform enables variable window sizes (or resolutions) in analysing the frequency content of a signal. The Wavelet analysis correlates the original signal $f(t)$ with a set of template functions obtained from the scaling (*i.e.* dilation and contraction) and the shift (*i.e.* translation along the time axis) of a wavelet function $\Psi$, referred to as *the mother wavelet*. In comparison with Fourier transform, Wavelet transform offers more flexibility as the analysing function can be chosen with more freedom, without the need of using sines and cosines forms. The similarity between the original function and the wavelet function is calculated separately for different time intervals, producing a two dimensional representation.

Basically, when dealing with Wavelet analysis, two types of transforms can be used: *the continuous Wavelet transform* and *the discrete Wavelet transform*.

**Continuous Wavelet transform (CWT)** The continuous Wavelet transform (CWT) is the sum over all time of scaled and shifted versions of the wavelet function $\Psi$. The CWT, when applied on the original signal $f(t)$, is expressed as:

$$CWT_f(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} f(t) \Psi^*(\frac{t-\tau}{s}) dt. \tag{13}$$

The calculation of the CWT over a signal results in many *coefficients*, which are functions of the translation parameter $\tau$ and the scale parameter $s$. In fact, $\tau$, is proportional to time information. It specifies the location of the wavelet in time; by varying $\tau$ the wavelet can be shifted over the signal. The scale $s$ is inversely proportional to the frequency information. The variation of $s$ modifies not only the central frequency of the wavelet, but also the window length. Large scales are related

to low frequencies, giving global information of the signal. Whereas, small scales correspond to high frequencies, revealing the finer details of the signal. Moreover, the signal energy remains constant at every scale as it is normalized by $\frac{1}{\sqrt{|s|}}$. Although the high accuracy provided by the CWT in analysing a signal, the computation of CWT is redundant and very time consuming. Usually the calculation of the CWT is performed by taking discrete values for the scaling parameter $s$ and the translation parameter $\tau$.

**Discrete Wavelet transform (DWT)** In practice, signals acquired experimentally are not continuous in time, but sampled as discrete time intervals. Previously, we have seen that the CWT performs a time-frequency resolution (or multiresolution) by scaling (contraction and dilation) and shifting a wavelet function. Recently, it has been shown that such analysis can actually be performed using multiresolution filter banks and wavelet functions, resulting in *the Discrete Wavelet Transform* (DWT). We note that the DWT is not the discretized version of the CWT, which just uses a discretized version of the scale and the translation parameters. The DWT Provides the information necessary to reliably analyse the content of a signal acquired; and more importantly it is much faster than CWT calculations. One level DWT is basically composed of what we call *wavelet filters*, which involve two basic concepts: the filter banks and the down- and up-sampling operations. The filter banks aim at changing the resolution by separating a signal into frequency bands. Actually, a discrete time signal is first filtered by the filters $L$ and $H$ which separate the frequency content of the analyzed signal in frequency bands of equal length. The filters $L$ and $H$ are thus respectively a low-pass and a high-pass filter. When the signal is put through the high-pass filter, then the high frequency information is kept and the low frequency information is lost. Whereas, if it is put through the low-pass filter, then the low frequency information is kept and the high frequency information is lost. Therefore, the signal is effectively decomposed into two sub-signals called *the details* (or detail wavelet coefficients) and *the approximations* (or approximation wavelet coefficients), respectively. The approximations correspond to low frequencies and the details to high frequencies. Note that the output sub-signals each contains half the frequency content, but an equal amount of samples as the original signal. In other words, the two sub-signals together contain the same frequency content as the original signal, however the amount of data is doubled. At this point, a down-sampling of factor two is applied to each sub-signal, which indeed doubles the scale (*i.e.* make the analysing window larger), doubles the frequency resolution; and therefore avoid redundancy. This way, the resolution and the scale have been changed in a manner to increase the frequency resolution and reduce the time resolution. In the second level of DWT, the approximation sub-signal is used as the original signal and put through a wavelet filter (*i.e.* filter bank plus down-sampling), until reaching the required level of decomposition. For a $p-$level decomposition, the range of frequency content included by the approximations and the details, denoted by $f_{Approx}$ and $f_{Det}$ respectively, can be determined as follows:

$$f_{Approx} = [0, \frac{f_s}{2^{p+1}}] \ , \ f_{Det} = [\frac{f_s}{2^{p+1}}, \frac{f_s}{2^p}]. \tag{14}$$

Where, $f_s$ is the sampling frequency of the original signal. Eventually, the original signal can be represented by the final approximation, related to the last level of decomposition, and the accumulated details corresponding to all levels. The process can be expanded to an arbitrary level, depending on the desired resolution. An illustration of a 3-levels DWT is shown in Fig. 3. Note that the relation between CWT and DWT is similar. In fact, the wavelets in the CWT act as a band-pass filter in the convolution of the wavelet function with the original signal; and so does the DWT sequence of low-pass filter, high-pass filter and down-sampling. Now, for

the reconstruction of the original signal, the same mechanism is used for the transformation. Actually, the reconstruction is possible thanks to the final approximation and the accumulated details. The obtained sub-signals are first upsampled and then passed through filter banks, which are related to the initial filter banks $L$ and $H$.

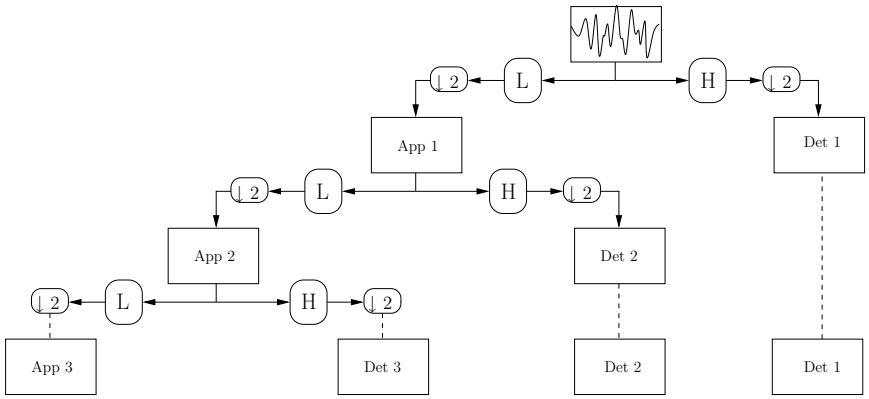


**Fig. 3.** An illustration of 3-levels DWT decomposition.

## 3 Applications to SCA

### 3.1 Wavelets for cryptographic pattern detection

In the Side-channel domain, the main challenge of the evaluator is to best analyse and exploit dependencies between the manipulated data and the electric (power, electromagnetic ...) consumption leaked from a CMOS circuit. In practice, the evaluator analyses a set of Side-channel consumption signals (or traces). Each trace includes a block of operations occurring during a cryptographic process (*i.e.* encryption or decryption). In real life applications, encrypted data bits are divided into small blocks of bits, called blocks of operations, depending on the specification of the cryptographic algorithm. For instance, Block ciphers, like the Advanced Encryption Standard (AES), are related to different modes of operations (*e.g.* ECB, CBC, CFB, OFB) that aim at dividing the encrypted data and managing the way of operation of the obtained blocks. Moreover, in the context of SCA, the alignment of traces is of great concern since deployed analysis are very sensitive to the magnitude of acquired traces. Consequently, in order to build a proper set of traces, the evaluator should be able to detect the start and the end of each block of operation. Unfortunately, in practice, it is almost impossible to perfectly collect aligned traces due to many factors. A frequent situation is that the trigger signal, which is precisely synchronized with the cryptographic process and used in functional testing or academic cases, is removed by the designer for security reasons. Additionally, the acquired traces are very often disturbed by the presence of noise. In such situation, we propose to use the CWT to reveal the global information involved by the cryptographic process. For this purpose, we recorded one signal involving the activity of three AES-128bit encryptions in CBC mode. Our measurement setup consists of one Xilinx Virtex 5 FPGA soldered on a XILINX LX30 platform, an 54855 Infiniium Agilent oscilloscope with a bandwidth of 6 GHz and a maximal sample rate of 40 GSa/s, antennas of the HZ–15 kit from Rohde & Schwarz. Fig. 4 is a capture generated from MATLAB (*wavemenu toolbox*), the powerful numerical computing environment. The top of this figure shows the original signal as it is

acquired by the oscilloscope. Clearly, the signal is disturbed by a high amount of noise and no information about the cryptographic process can be revealed. At the center of the figure, we can see the two dimensional representation of the CWT. Obviously, thanks to this representation, we can easily detect the limits of the three measured AES blocks. Actually, we have added the dashed lines to highlight these limits. We note that these limits are detected for high scales (low frequencies) of CWT. Moreover, we can reveal more details about the information content of the signal, such as the number of rounds composing each block of AES. This may be very helpful for the evaluator, specially when the analysed algorithm is unknown. Besides, as shown at the bottom of the figure, the MATLAB tool generates an approximate shape (*i.e.* extracting the global information) of the analysed signal based on the CWT coefficients.

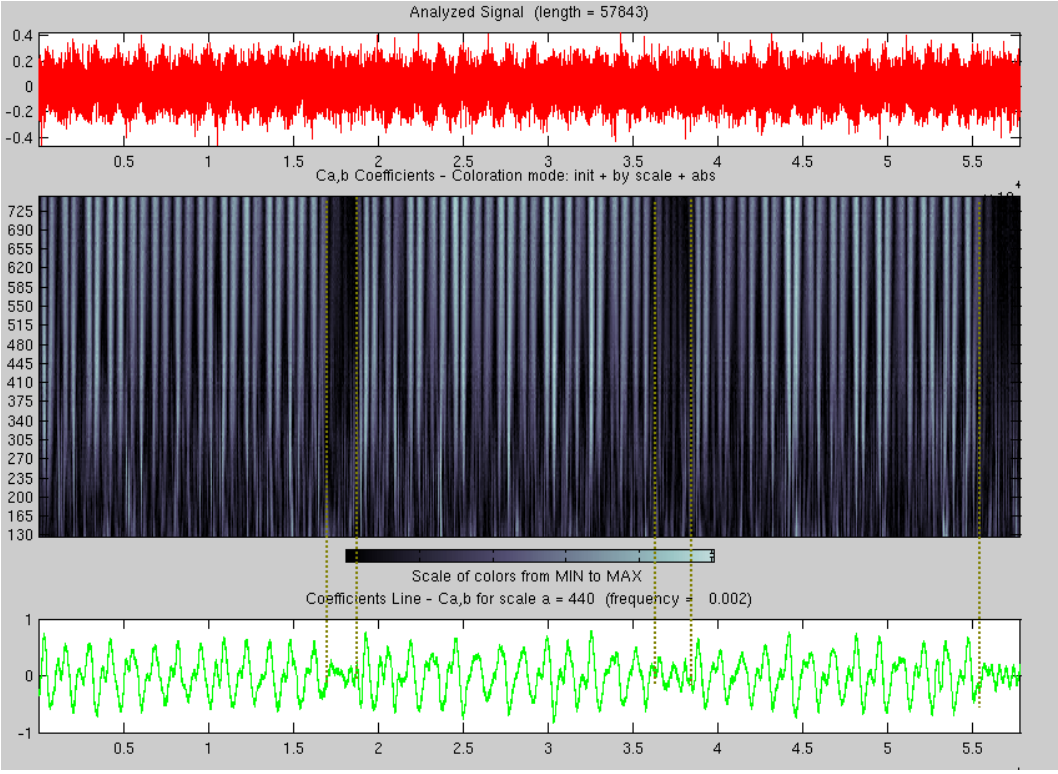


**Fig. 4.** An illustration of the CWT on three AES encryption blocks.

### 3.2 Wavelets combined Mutual information for Side-channel traces denoising

**Side-channel traces noise filtering using wavelets** In SCA litterature, the procedure of denoising Side-channel traces, using Wavelets transforms, was first proposed in [1]. The authors in [1], proposed the very general method to remove noise from signals as it is described in signal processing books, without detailed explanations. The method is based on the Discrete Wavelet Transform. Actually, we have seen that when a signal is decomposed using the DWT, we are left with a set of wavelet coefficients that correlates to the high frequency sub-bands, which consist of the details in the analysed signal. If the details coefficients are small enough, they

might be omitted without substantially affecting the main features of the signal. Moreover, these small details are often those associated with noise; thus, by setting these coefficients to zero, we are essentially removing the noise. This becomes the main idea behind *thresholding* all frequencies that are less than a particular threshold to zero and use these coefficients in an inverse Wavelet transformation to reconstruct the denoised version of the original signal. The Threshold value is estimated using the universal formula of Donoho [4]. The Donoho's threshold, $\lambda_{Donoho}$, can be computed as follows:

$$\lambda_{Donoho} = \sigma\sqrt{2log(len)}\ ,\ \sigma = \frac{median|details|}{0.6745} \tag{15}$$

$$\begin{cases} Th_{hard}(c) = c\ if\ |c| > \lambda_{Donoho} \\ \qquad\qquad = 0\ if\quad c \leq \lambda_{Donoho} \end{cases} \tag{16}$$

Where $\sigma$ and $len$ are respectively the noise variance and the length of the details coefficients. This threshold is applied only on the details coefficient for a specific wavelet scale level. In fact, a new threshold is computed for each scale and used in a *thresholding function*. There is two basic types of thresholding functions: A hard thresholding $Th_{hard}$, defined in Eqn. (16), that sets to zero all details coefficients that are below the threshold value $\lambda_{Donoho}$; and a soft thresholding $Th_{soft}$, defined in Eqn. (17), for which the details coefficients with magnitudes smaller than $\lambda_{Donoho}$ are set to zero, but the retained coefficients are also shrunk towards zero by the amount of the threshold value in order to decrease the effect of noise assumed to corrupt all the wavelet coefficients.

$$\begin{cases} Th_{soft}(c) = sign(|c| - \lambda_{Donoho})\ if\ |c| > \lambda_{Donoho} \\ \qquad\qquad\qquad\qquad = 0\ if\quad c \leq \lambda_{Donoho} \end{cases} \tag{17}$$

**Improvement of the noise filtering basic method** In the open literature of wavelet analysis, many scientific papers deal with the efficiency of Donoho's threshold to treat digitally acquired signals. However, we assume that such thresholding techniques are very generic and not sufficient to treat the problem of noise reduction from the SCA context. Actually, from the security evaluation perspective, the evaluator usually knows the value of the secret key. His goal is to make his analysis in the best conditions; and thus to know at which point a secure implementation can be resistant to Side-channel attacks. Therefore, the knowledge of the secret key can be a crucial advantage to the security evaluation analysis. In this context, we propose to use the powerful *information theoretic approach* as a complementary tool to Donoho's threshold. This information theoretic approach is basically used to measure the amount of the useful information, which is leaked from the cryptographic implementation. Technically speaking, this metric is mainly based on the mutual information theory. In the context of SCA, we evaluate the amount of information in the Side-channel leakages with the mutual information, measured in bits, between the global observation $O$ (*i.e.* the set of traces acquired) and the leakage $L$ when the secret-key is known, as:

$$MI(O;L) = \sum_{o}\sum_{l} P(o,l)log\frac{P(o,l)}{P(o)P(l)} \tag{18}$$

$$MI(O;L) = H(O) - H(O|L) \tag{19}$$

Where $H(O)$ is an estimation of the entropy of $O$, $p(o,l)$ is the joint probability density function of $O$ and $L$, $p(o)$ is the marginal probability density function of $O$ and $H(O|L)$ is the conditional entropy of $O$ knowing $L$. $MI(O;L)$ can be regarded as a positive (*i.e.* $MI(O;L) \geq O$) and symmetric (*i.e.* $MI(O;L) = MI(L;O)$) measure

of the strength of a 2-way interaction between two variables: the observation $O$ and the leakage $L$ that is related to the secret key. But more importantly, the higher the value of the mutual information is, the higher the dependency between $O$ and $L$ is. Statistically speaking, $MI(O; L) = 0$ if and only if $O$ and $L$ are independent random variables. Therefore, such tool allows the evaluator to detect the time instants that are directly related to the secret key, during the cryptographic process. This has a pure practical flavour for the analysis, as the evaluator will be able to improve the basic threshold of Donoho. In fact, the mutual information can be used as a complementary tool to Donoho's threshold in order to add more information about the real features of traces acquired; and therefore to improve the reliability and the accuracy of the thresholding. Our proposed mutual information based Donoho's threshold, can be stated in four steps as follows:

- **step 1** Applies the DWT over all traces acquired for a desired level of decomposition.
- **step 2** Computes the $MI$ over the set of DWT coefficients obtained.
- **step 3** Keeps in memory (table $T$) the temporal indexes of all coefficients which values are located below a certain threshold. (*i.g.* 90% of Max($MI$)).
- **step 4** Pre-processes the original traces as follows: For each trace, the DWT is first computed (with the same level of decomposition as used previously). Second, $\lambda_{Donoho}$ is calculated based on the obtained DWT coefficients. Finally, **for each** coefficient $c$ that value is above $\lambda_{Donoho}$ **and** that index belongs to the table $T$, **then** the value of $c$ is set to zero.

Clearly, the new algorithm aims at discarding more coefficients that are really related to noisy time samples; and therefore favorising only the actual coefficients that are related to the sensitive information. Basically, the Donoho's threshold is computed over the details; nonetheless it is noteworthy that in the open litterature of wavelet analysis other thresholds have been proposed to deal with both the approximation and the details [5]. We note that our proposed algorithm is generic and can be plugged to any proposed threshold. An illustration of $MI$ computation is shown in Fig. 5. The $MI$ is computed over the first-level DWT decomposition of an unprotected DES implementation (500 traces used and only the first two rounds were considered). We remark that for an unprotected implementation, the sensitive information is mainly located within the approximations which are related to the low frequencies. However, during our experiments, we have noticed that, for some unprotected implementations, sensitive information can not be negligible in the details' region. A logical result from thresholding is the compression of signals. Indeed, the evaluator is required to collect as much traces as he can in order to know whether the cryptographic implementation can be broken by an attacker. In practice, the acquisition of SCA traces, which is usually performed by an oscilloscope, requires a high storage capacity especially when countermeasures are implemented. In fact, the evaluator is required to evaluate the amount of information leaked from the cryptographic process. Hence, for each measurement, the whole cryptographic operations should be included, which requires a high storage capacity when taking into account all traces acquired. Moreover, a high sampling rate is usually used when acquiring traces; thus more memory resources are needed. In this context, the compression of signals that involve a big amount of information, has been recently the focus of extensive research. One of the most efficient solutions is to use the Discrete Wavelet Transform. The idea behind compression is to remove redundancies from the signal and keep only the useful information, in a real time manner (*i.e.* before storage). From the SCA perspective, the useful information can be defined as the power or electromagnetic consumption related directly to the activity of the cryptographic process, resulting in the recovery of the secret key. In the litterature, Wavelet transforms based signals compression is basically performed using the notion of *threshold*, detailed previously. We note that the threshold $\lambda_{MI_{Donoho}}$, that

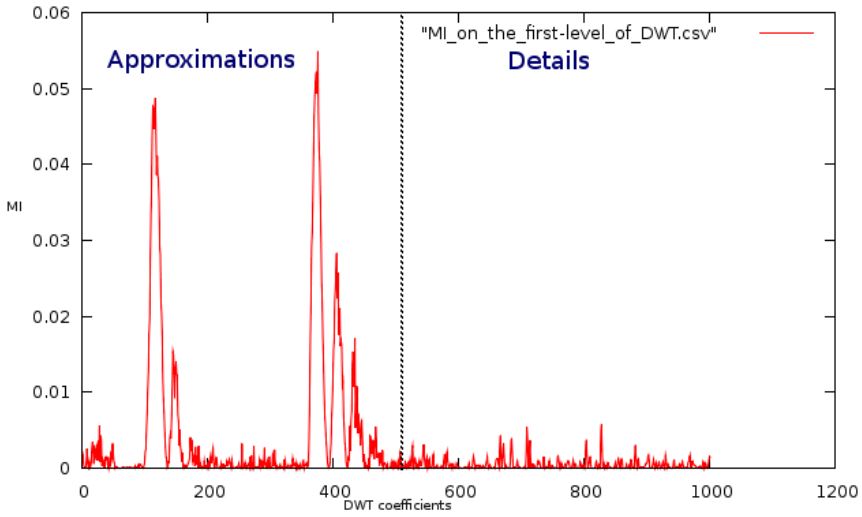


**Fig. 5.** Computation of the MI on the first-level DWT of unprotected DES traces.

we have proposed for the denoising context, can not be applied here. Indeed, the $\lambda_{MI_{Donoho}}$'s threshold needs the entire set of traces acquired to be computed, unlike the $\lambda_{Donoho}$'s threshold that is computed for each trace acquired. Several metrics such as *the zero ratio* and *the retained energy ratio* [5], have been proposed in the litterature, to assess the compression quality. However, when dealing with signals compression, the thresholding usually requires additional coding techniques like *the Huffman coding*, which slow down the compression process. From the SCA perspective, and specifically for unprotected implementations, we propose to keep only the approximation coefficients and reject the details. This way, the compression process is fast and the storage capacity gain is fixed. As experiments, we have measured the mutual information, in the same way as done for generating the Fig. 5, for different implementations (Data Encryption Standard (DES) and AES) and for different levels of DWT decomposition. As a result, we have realized that the sensitive information, within SCA traces, is basically represented by low frequencies. Therefore, the compression of traces, when taking into account only the approximations, is performed without loss of information. Indeed, in the litterature of multiresolution analysis, it has been often reported that a loss of information could happen only when reconstructing the signal. For this purpose, the so-called *bi-orthogonal wavelet family* is often recommended for the reconstruction.

### 3.3 Wavelets for secret key recovery

In SCA, the interest is to find some ways to accelerate the attack, taking into account that the scarce resource is the number of traces acquired. As stated before, the importance of wavelets relies on separating the different sources (*i.e.* an electrical activity with different frequencies content) composing the signal acquired. We recall that the signal content can be represented differently thanks to the concatenation of the approximations of the last level of decomposition with the details. From the SCA perspective, such representation is very helpful as the main goal of SCA is to separate and extract the sources related to the cryptographic process and responsible for the manipulation of the secret key. Our proposition is to perform the Side-channel analysis directly on the wavelet coefficients: the approximations

and the details. In other words, the leakage is not anymore represented by the time samples of the acquired signals but simply by the value of the wavelet coefficients. This way, in one sense, we may improve the efficiency of Side-channel distinguishers. In another sense, we avoid the problem of signal's reconstruction stated previously in the compression context.

**Experimental results** In our first experiment, we have considered a Correlation Power Attack (CPA) performed on real unprotected DES traces. More precisely, three analyses have been involved: CPA on the original traces (*orig*), CPA on the first level of wavelet decomposition, *i.e.* the approximations and the details, (*AppDet*) and a CPA on only the approximation coefficients (*App*). In order to assess the efficiency of the attack, we computed the most commonly used metrics, known respectively as *the first-order success rate* (SR) and *the guessing entropy* (GE), which are basically proposed by F.X Standeart in [6]. The first-order success rate, which is depicted in Fig. 6, shows that analysing the traces using wavelets is clearly more efficient than the basic analysis in the time domain. Actually, the SR is converging faster towards the maximal rate for *AppDet* and *App* than for *orig*. For instance, to reach a SR of 90%, we need around 150 traces for *orig*, and only 75 traces for *AppDet* (*i.e.* a gain of 50%). But more importantly, by examining and comparing the performance of *AppDet* and *App*, we prove again the importance of performing the analysis on only the approximation coefficients for unprotected implementations. Nonetheless, during our experiments, we have remarked that *AppDet* often performs better then *App* when the implementation is protected. Therefore, the evaluator is required to conduct his analysis on *AppDet*, specifically when the deployed countermeasures are unknown. In Fig. 7, the guessing entropy, unsurprisingly confirms the results found for the SR. In fact, the rank of the secret key when performing a ordinary CPA takes more time to converge to the lowest and best rank, compared to the analyses *AppDet* and *App*. Generally, from

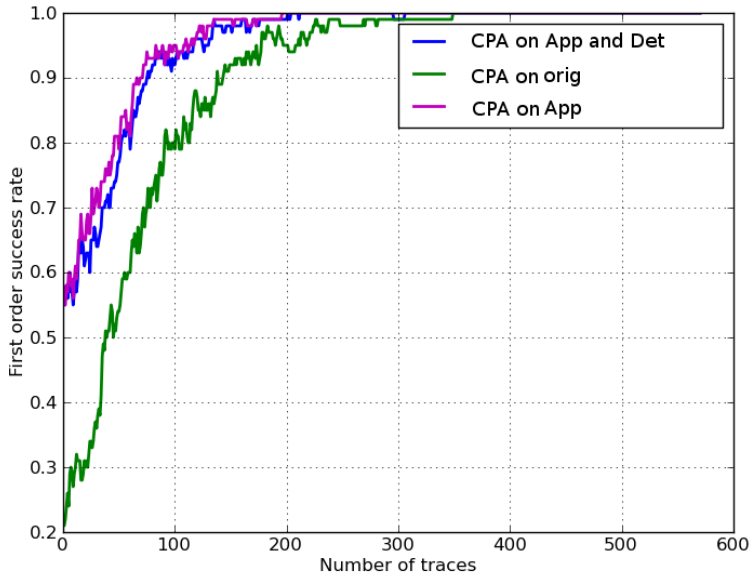


**Fig. 6.** CPA first-order success rate.

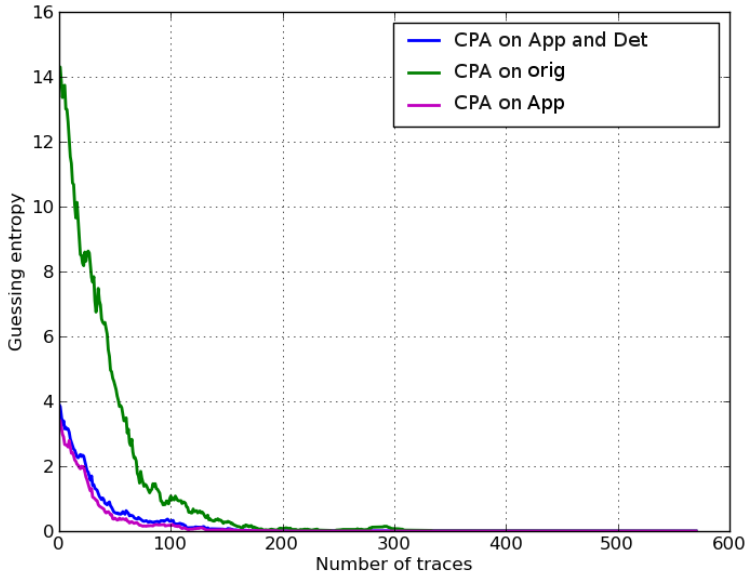


**Fig. 7.** CPA guessing entropy.

the evaluation perspective, the evaluator has at his disposal powerful tools, usually known as profiling analysis, such as *the template analysis* [7] and *the stochastic models* [8], which mainly aims at revealing the finer details related to the leakage; and therefore allowing the evaluator to reliably assess the robustness of the analysed cryptographic implementation. In our second experimentation, we are interested in *Principal Components Analyses (PCA)* based templates and examine the effect of combination with Wavelet transforms on SCA specially when increasing the level of wavelet decomposition. Indeed, in the open literature of wavelet analysis, the combination between *PCA* and Wavelet Transforms has been recently proposed in [9]; and has turned out to be efficient in many engineering fields [10] [11]. The main idea behind combining PCA with wavelets is to remove redundancy from wavelet coefficients; and therefore keep only de-correlated ones. According to Fig. 8 and Fig. 9, first we notice that the wavelet analysis outperforms, in terms of SR and GE, the standard analysis, *i.e.* when performed on the original trace; and this for all levels of decomposition. Moreover, the performance of the analysis is getting better when the level is increased. In principle, the more scale levels are used, the higher the performance is. Nonetheless, the energy can not be centralized endlessly. Hence, there is always a limit scale level for the Wavelet transform. Beyond this level, performance can not be increased. Besides, the increase in scale level means more computations and time complexity. Therefore, a proper scale level should be selected based on some points such as the amount of noise affecting the traces and the frequency sampling used for the acquisition. When performing a template analysis, the selection of a proper level of decomposition can be easily achieved thanks to the clone device used for the profiling.

### 3.4 Conclusion

In the perspective of reliably evaluate the robustness of secure devices against Side-channel attacks, four aspects of analysis are basically taken into consideration: the acquisition of Side-channel traces, the pre-processing of traces acquired, the detec-

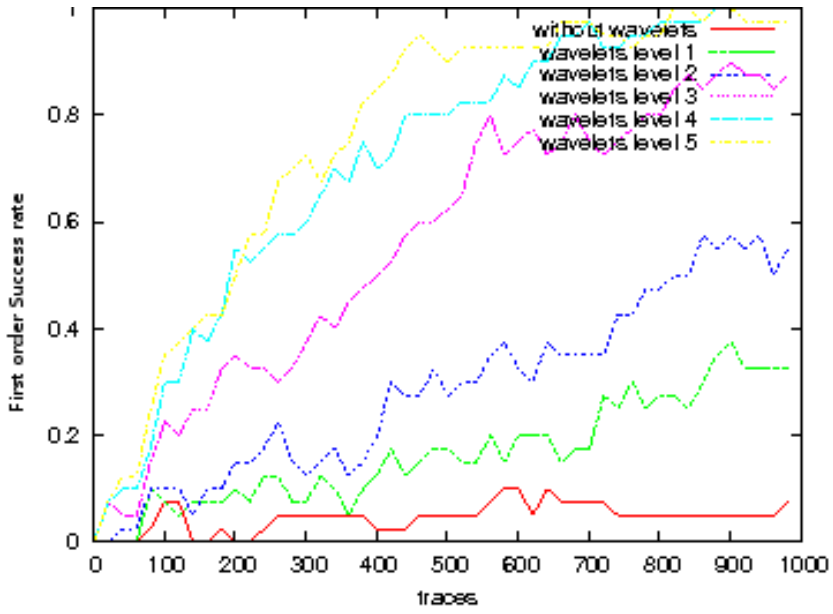


**Fig. 8.** Template attack first-order success rate.

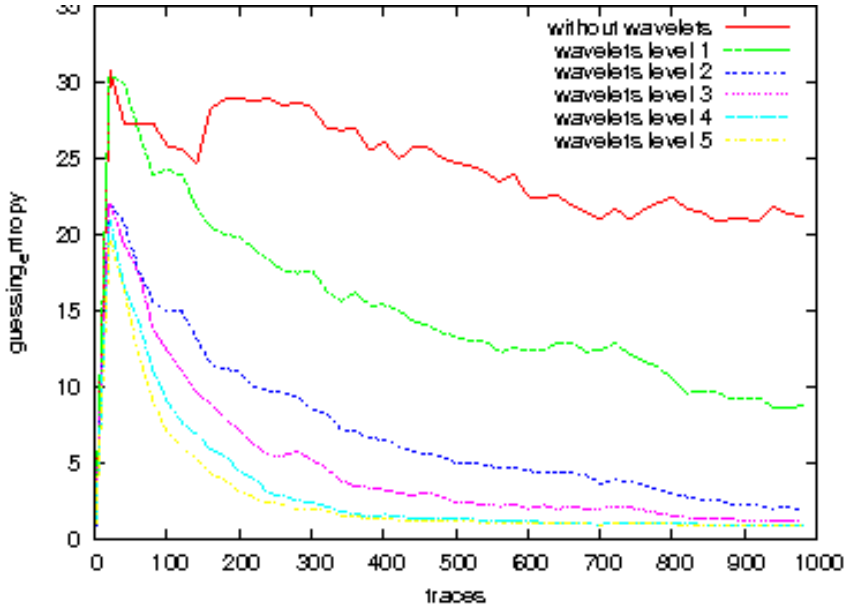


**Fig. 9.** Template attack guessing entropy.

tion and extraction of cryptographic patterns from the preprocessed traces, and finally the recovery of sensitive information, referred to as *the secret key*. In this paper, we provide the evaluator with novel applications of Wavelet transform in the context of Side-channel analysis. Although Wavelet theory has been successfully applied in many applications in science and engineering, it has been rarely invoked in the SCA context. Actually, Wavelets have only been used in two occasions: first they are used to remove noise from Side-channel traces (which is already known), and second used to estimate the probability density function of the leaked infor-

mation. In this paper, we show how Wavelet transform can be useful to reduce the storage capacity by compressing Side-channel traces. Besides, we highlight the way in detecting and extracting the patterns of the analysed cryptographic process. Additionally, we propose an improvement of signals denoising from the Side-channel context. Besides, we show that wavelet analysis is not only used for pre-processing purposes, but also in the very core of the attack. The overall goal of this paper is about valuing, in a methodological manner, the use of Wavelet transform in the SCA domain. Fig. 10 shows the involvement of the proposed applications in all SCA aspects.

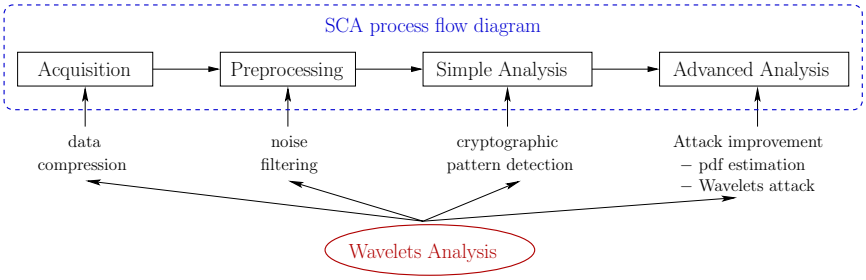


**Fig. 10.** Involvement of wavelet analysis in SCA security aspects.

## References


1. Pelletier, H., Charvet, X.: Improving the DPA attack using Wavelet transform (2005) Honolulu, Hawai, USA; NIST's Physical Security Testing Workshop. Website: `http://csrc.nist.gov/groups/STM/cmvp/documents/fips140-3/physec/papers/physecpaper14.pdf`.
2. Rhee, K.H., Nyang, D., eds.: Information Security and Cryptology - ICISC 2010 - 13th International Conference, Seoul, Korea, December 1-3, 2010, Revised Selected Papers. In Rhee, K.H., Nyang, D., eds.: ICISC. Volume 6829 of Lecture Notes in Computer Science., Springer (2011)
3. Gabor, D.: Theory of communication. J. Inst. Elect. Eng. **93** (1946) 429–457
4. Lin, S., Huang, X.: Advanced Research on Computer Education, Simulation and Modeling: International Conference, CESM 2011, Wuhan, China, June 18-19, 2011. Proceedings. Number ptie. 2 in Communications in Computer and Information Science Series. Springer (2011)
5. Shang, L., Jaeger, J., Krebs, R.: Efficiency analysis of data compression of power system transients using wavelet transform. In: Power Tech Conference Proceedings, 2003 IEEE Bologna. Volume 4. (2003) 6 pp. Vol.4
6. Standaert, F.X., Malkin, T., Yung, M.: A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In: EUROCRYPT. Volume 5479 of LNCS., Springer (2009) 443–461 Cologne, Germany.
7. Chari, S., Rao, J.R., Rohatgi, P.: Template Attacks. In: CHES. Volume 2523 of LNCS., Springer (2002) 13–28 San Francisco Bay (Redwood City), USA.
8. Schindler, W., Lemke, K., Paar, C.: A Stochastic Model for Differential Side Channel Cryptanalysis. In LNCS, ed.: CHES. Volume 3659 of LNCS., Springer (2005) 30–46 Edinburgh, Scotland, UK.
9. Bakshi, B.R.: Multiscale pca with application to multivariate statistical process monitoring. AIChE Journal **44** (1998) 1596–1610
10. Lee, D.S., Park, J.M., Vanrolleghem, P.A.: Adaptive multiscale principal component analysis for on-line monitoring of a sequencing batch reactor. J Biotechnol **116** (2005) 195–210
11. Aminghafari, M., Cheze, N., Poggi, J.M.: Multivariate denoising using wavelets and principal component analysis. Comput. Stat. Data Anal. **50** (2006) 2381–2398