Novel Approaches for Improving the Power Consumption Models in Correlation Analysis

Thanh-Ha Le, Quoc-Thinh Nguyen-Vuong, Cécile Canovas, Jessy Clédière

CEA-LETI
17 avenue des Martyrs, 38 054 Grenoble Cedex 9, France,
{thanhha.le, quoc-thinh.nguyen-vuong, cecile.canovas,
jessy.clediere}@cea.fr

Abstract. Differential Power Analysis (DPA) is a powerful technique for revealing secret data of cryptographic algorithms such as DES, AES and RSA implemented on a specific platform. In recent years, Correlation Power Analysis (CPA) allowed to better formalize the differential approaches of DPA with the use of a power model. We propose here two methods in order to optimize the power model for the targeted bits of the analysed algorithm. We will consider that all the targeted bits do not give the same contribution to the power consumption. Our first method consists in finding out the optimal ratio among the bits of a specific device. The second method is based on a statistical analysis of attack results while applying different possible ratios among the bits. The experimental electromagnetic radiation signals intercepted from an ASIC during DES operations show that our proposed methods allow to improve significantly the attack performance.

1 Introduction

Side channel analysis has been a dangerous smart card attack technique since its discovery. The basic of this method is to extract physical information during the operation of a cryptographic device. Rather than regarding the regular I/O interface, the physical phenomena like timing of operation, power consumption or electromagnetic emanation, so-called side channel signals, are used for deducing secret informations. The effectiveness of side channel attacks has been tested in many types of devices (ASIC, FPGA) implemented with different cryptographic algorithms such as DES, AES, RC4, ECC and RSA. Side channel attacks become a serious threat for the cryptographic modules since they are easy to implement. From a different point of view, it is an important instrument for reinforcing the level of security.

The first side channel analysis, also known as timing attack, was introduced by Kocher in 1996 [16]. Some years later, the well known Differential Power Analysis (DPA) was introduced [17]. From then to now, many explanations, variations and improvements on DPA have been presented [8, 21, 20, 22, 11, 23, 28, 4, 6, 5, 18]. Several countermeasures have been proposed and are successfully used to secure those algorithms from first and high order side channel attacks [14,

10, 1, 2]. Also, the origin of the leaked signal has evolved. From the original power consumption signal used in [17], electromagnetic radiation signals [13, 26, 27], acquired by dedicated sensors, took the place of the most powerful side channel signals. Electromagnetic radiation signals can be used in close or far field. In the first case, the major advantage comes from the selectivity of the radiation signal, in the second case, the signal can be intercepted without the need of direct device access. Whatever the power consumption signals or electromagnetic signals are used, the principle of side channel analysis remains the same and is closely linked to a power model.

Several power models have been proposed [6, 15, 30] and allow to better formalize the differential analysis with the introduction of the correlation approach [6, 8, 11, 20]. This formalized analysis is denoted as the Correlation Power Analysis (CPA) introduced in [6]. Let's note that although CPA brings a powerful formalism, it has been shown in [19] that CPA is a multi-bit DPA with a Hamming distance approach and a normalization by the standard deviation of the side channel signal.

Besides that, the power consumption model can not always be fully used for an analysis. This is specially true for a hardware DES where the full RL register is synthetized and is targeted. Considering the entire register is not possible due to the 48 bit key-hypothesis required to compute the bit values. An attacker will need to focus the attack on a more restricted bit size, typically on four output bits of a single S-box. In this case, the power consumption of this restricted register is evaluated.

The multi-bit DPA concepts, as proposed by [23, 4] and formalized in [19] by the Partitioning Power Analysis (PPA) represent an intuitive way to modelize the power consumption model restricted to the analysed bits. At some points, it may be easier and faster to tune and optimize the coefficients of PPA rather than playing with the restricted model itself.

In this paper, we will see that considering an imbalanced effect of the targeted bits can greatly outperform the classic analysis linked to a power model where all the bits have the same contribution. More precisely, we exploit the non-equivalence among the targeted bits and then proposed two novel approaches to enhance the multi-bit coefficients and thus the underlying power model of side channel analysis.

2 Correlation Power Analysis and Consumption model

The CPA introduced by Brier et al. [6] is based on the correlation denoted $Corr$ between the power consumption signal $W(C_i)$ and its model $M(f(C_i, K_s)_{\mathcal{B}})$ where the C_i $(i \in 1 \dots n)$ are the cipher texts (or plain texts), K_s is the supposed key, f is the selection function and $\mathcal{B} = \{b_1 b_2 \dots b_d\}$ the d-bit targeted set.

$$\texttt{CPA}(K_s) = Corr(\{W(C_i), i \in 1 \dots N\}, \{M(f(C_i, K_s)_{\mathcal{B}}), i \in 1 \dots N\}) \qquad (1)$$

The choice of the model depends on the platform and is important for improving the attack effectiveness.

The model given by Messerges in [22] is based on the Hamming weight:

$$M_{\mathtt{hw}}(D_t) = a.H(D_t) + b_t \tag{2}$$

where a is a real constant value, b_t is the noise over time, D_t represents the value processed in the microchip at instant t and H stands for the Hamming weight. As stated by Messerges, this simple model can be used to understand the first and second order DPA attacks.

Some other authors presented models based on the Hamming distance, i.e. taking into account the previous state of the internal registers of the CMOS device. Such a model has been proposed by Brier et al.[6] and is very concisely given by:

$$M_{\mathtt{hd}}(D_t) = a.H(D_t \oplus D_{t-1}) + b_t \tag{3}$$

where D_{t-1} represents the previous value of D_t processed in the microchip.

Peeters et al. [24] gave a switching distance model that distinguishes the $0 \rightarrow 1$ and $1 \rightarrow 0$ bit transitions:

$$M_{\mathtt{sd}}(D_t) = a\left((1-\frac{\delta}{2})H(D_t \oplus D_{t-1}) + \frac{\delta}{2}(H(D_t) - H(D_{t-1}))\right) + b_t \tag{4}$$

where δ is the normalized difference of the transition leakages as defined in [24].

Although these concepts are derived from different approaches, they are based on the common assumption that all tested bits are equivalent. We are only interested in the Hamming weight or the Hamming distance of $\mathcal{B}$ and we do not care about which bits among d bits $b_1, b_2, \ldots, b_d$ are flipped.

3 Bit-asymmetry hypothesis

Contrary to the above *bit-equivalence* hypothesis, the side channel leakage of a real device depends on the handled specific flipping bits. This dependence can be merely due to some imbalance conception effects such as output capacitance, wire length of an output gate, and data dependant effects that modify the fan out of the considered gate. This dependence has been noted in [28].

Let's have a look on models that take into account imbalanced bit effects.

Bevan gives in [4] and [5] a model that can be easily adapted:

$$M_{\mathtt{ba}}(D_t) = \sum_{i=1}^{d} (1-D_t^i)(D_{t-1}^i)c_{10} + (D_t^i)(1-D_{t-1}^i)c_{01} + Crest \tag{5}$$

where c_{01} represents the current consumption of a bit to flip from 0 to 1, c_{10} represents the current consumption of a bit to flip from 1 to 0, D_t^i is the i-th bit value of D_t, and $Crest$ denotes the current consumption independent of the data being processed. We can introduce a difference between bits by indexing c_{01} and c_{10} with the bit number i.

The better is the model, the greater is the correlation, the more efficient is the attack. The difficulty is to define a both precise and general model for all components.

Now Le et al. [19] show that with a model based on Hamming distance or Hamming weight like those presented in Sect. 2 the CPA is equivalent to PPA with a normalization factor. The PPA splits the signal curves $W(C_i)$ in groups G_j depending on $f(C_i, K_s)_{\mathcal{B}}$ and applies the same weight for each curve of the same group.

$$\texttt{PPA}(K_s) = \sum_{j=0}^{d} a_j \frac{\sum_{G_j} W(C_i)}{card(G_j)} \tag{6}$$

where a_j is a chosen weight, $card(G_j)$ is the number of elements of the group G_j. The choice of the groups and weights is based on the consumption model. In order to take into account the bit asymmetry, we define 2^d groups: $G_{v_1 v_2 \dots v_d} = \{W(C_i) i \in 1 \dots N | f(C_i, K_s)_{\mathcal{B}} = v_1 v_2 \dots v_d\}$. For example $G_{0\dots01}$ is the group of curves where only the last targeted bit changes. In that way, it is possible to directly tune the weights of PPA in order to take into account imbalanced bit effects. For example if the last bit leaks more than the others, we modify the weighting of group $G_{0\dots01}$.

However, if we consider four bits ($d = 4$), this method requires the tuning of 16 weights and also 16 PPA computings. So we need another computing that is equivalent to CPA or PPA, which requires less computings.

Le et al. [19] also show the equivalence of Bevan's multi-bit DPA [4] with PPA and CPA. Similarly, this computing does not take into account different behaviours of examined bits since it computes the sum of four mono-bit DPA curves with an identical weight for all bits.

$$\texttt{DPA}_{\texttt{Bevan}}(\mathcal{B}) = \sum_{i=1}^{d} \Delta_D(b_i) \tag{7}$$

where $\Delta_D(b_i)$ is the mono-bit DPA value of bit b_i.

An easy way to exploit the different behaviours of each bit is to use the multi-bit DPA proposed by Bevan with modified weights α_i as shown in the following equation:

$$\Sigma_D(\mathcal{B}) = \sum_{i=1}^{d} \alpha_i \Delta_D(b_i) \tag{8}$$

Thus the values of α_i can be chosen proportional to the significance of bit i in multi-bit combination. The α_i coefficients will be considered independent to the data. In the rest of the paper, we focus on improving this multi-weight power analysis.

4 Our proposition

4.1 Selective multi-weight power analysis

As stated earlier, the multi-bit differential attack efficiency can be improved if we have knowledge about the reaction of each examined bit and then deduce a

suitable set of weights α_i. Basically, the optimal set of weights α_i can be different from one cryptographic device to another. However, for identical devices, if the noise level in side channel signals is slight, the proportion among the contribution of examined bits (i.e. the proportion among the α_i) remains almost the same. Therefore, the optimal weights, which are considered as reference ones, should be determined using a large number of averaged side channel signals.

Briefly, the optimal set of α_i can be determined by using a reference device which is identical to the attacked one and an already known key. We compute the $\Sigma_D(\mathcal{B})$ for different sets $\{\alpha_i\}$ using (8). As the key is known, we will choose the set of weights which permits to detect the key the most efficiently. These values will be then employed to hack other devices of the same type. The attack approach is depicted in the following figure.

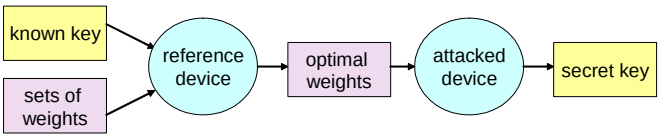

Fig. 1. Attack approach using selective weights

Note that the use of a reference device, which is identical to the device under test, has been proposed in the past in other side channel attacks such as the Inferential Power Analysis (IPA) [12] and the Template Attacks [9]. These attacks comprise two stages: a characterization stage, in which statistical operations are performed on a large number of power traces to learn details of the implementation, and a key extraction stage, in which the key is obtained from a very few power traces. Our method is also based on a profiling stage, but this allows to discover the optimal set of weights of the power model.

Theoretically, the values of the α_i can vary from -∞ to +∞. When α_i tends to ∞, the multi-weight power analysis becomes the mono-bit DPA related to bit b_i. On the other hand, when α_i tends to 0, the multi-weight power analysis does not consider the contribution of b_i. In order to reduce the number of tests, we can fix one weight to 1 and vary others. In reality, the variation of α_i from -10 to 10 is largely sufficient to discover the best values of α_i.

4.2 Statistical multi-weight power analysis

In this subsection, we describe a novel method to improve the multi-bit power analysis performance based on statistical analysis. Contrary to the previous proposed method, the statistical analysis based attack does not need a reference device as well as the suitable set of weights. Instead, we try all different sets of weigths for multi-bit power analysis computing and keep the best key.

Recall that the key that gives the highest differential peak is considered as the secret key. Obviously, there exist some sets of weights that imply a wrong key

detection results. Consequently, we observe all the results and compute for all key hypothesis the frequency of being designated as secret key. The hypothesis, which gives the highest occurrence frequency, is the wanted secret key. The attack approach is illustrated in Fig. 2.

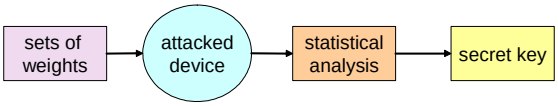

Fig. 2. Attack approach using statistical analysis

In order to see how the above statistical analysis can effectively help to detect the secret key, let's investigate the differential curve value using mathematical analysis. The switching distance model proposed in [24] was used to analyse both multi-weight power analysis attacks based on Hamming weight and Hamming distance. According to this model, a CMOS gate consumes energy differently when charging or discharging the load capacitance. Hence, we can assume the power consumption of a transition of bit b_i as follows: $P_{0\to0}^{b_i} = P_{1\to1}^{b_i} = 0$, $P_{1\to0}^{b_i} = p_i$ and $P_{0\to1}^{b_i} = p_i + \delta_i$.

Let's consider two key hypotheses: the correct one and a wrong one. We denote $\Sigma_D^c(\mathcal{B})$ and $\Sigma_D^w(\mathcal{B})$ the differential curves corresponding to the correct and the wrong key hypothesis respectively. We also assume that the texts C_j are independently and identically distributed.

The detailed calculation of the differential curve values is developed in appendix. The results are given in Table 1:

	$\Sigma_D^c(\mathcal{B})$	$\Sigma_D^w(\mathcal{B})$
Hamming weight	$\sum_{i=1}^{d} \alpha_i \frac{\delta_i}{2}$	$\sum_{i=1}^{d} \alpha_i (1 - \frac{n_i}{N_1} - \frac{n_i}{N_0}) \frac{\delta_i}{2}$
Hamming distance	$\sum_{i=1}^{d} \alpha_i (p_i + \frac{\delta_i}{2})$	$\sum_{i=1}^{d} \alpha_i (1 - \frac{n_i}{N_1} - \frac{n_i}{N_0})(p_i + \frac{\delta_i}{2})$

Table 1. Differential curve values at the instant τ

The relation $N = N_0 + N_1$ represents the partition of power consumption signals for each bit b_i during mono-bit DPA computing. If a wrong key is used, there are n_i curves which are wrongly placed.

It is important to note that $\left|1 - \frac{n_i}{N_1} - \frac{n_i}{N_0}\right| \leq 1$ as $n_i \leq N_0$ and $n_i \leq N_1$. The equalization happens only when $n_i = N_0 = N_1$ or when $n_i = 0$. By consequence,

$\left|\alpha_i(1-\frac{n_i}{N_1}-\frac{n_i}{N_0})\frac{\delta_i}{2}\right| \leq \left|\alpha_i\frac{\delta_i}{2}\right|$. Therefore, if we define, as Rivain suggests in [29], $\Sigma_{D_{abs}}(\mathcal{B}) = \sum_{i=1}^{d}\alpha_i|\Delta_D(b_i)|$ with $\alpha_i \geq 0$, we always have:

$$\Sigma^c_{D_{abs}}(\mathcal{B}) \geq \Sigma^w_{D_{abs}}(\mathcal{B}) \tag{9}$$

It means that with any set of positive weights $\{\alpha_i\}$, the correct key must always theoretically correspond to the highest $\Sigma_{D_{abs}}(\mathcal{B})$. In reality, this confirmation does not always hold due to the noise of measurement, the misalignment of signals and the imperfect distribution of texts. However statistically the correct key has a good probability to be guessed. Hence, our statistical analysis is based on $\Sigma_{D_{abs}}(\mathcal{B})$ with different positive values α_i. After testing all the sets of coefficients, the correct key is the one which is assigned the most frequently.

Note however that $(\Sigma^c_D(\mathcal{B}) - \Sigma^w_D(\mathcal{B}))$ in the case of Hamming distance is generally larger than in the case of Hamming weight. Therefore, the attack based on Hamming distance, if it is possible, is much more successful than the attack using Hamming weight. Unfortunately, the reference state of $\mathcal{B}$ is not always known and thereby we can not calculate the Hamming distance of $\mathcal{B}$ in many cryptographic devices.

5 Performance evaluation

5.1 Experiment description

In our experiment, we measure the electromagnetic emanations of a synthesized ASIC during a DES operation. Corresponding to each random text used in input, we acquire an electromagnetic signal (a side channel signal). An example of electromagnetic signal is depicted in Fig. 3 where we can observe 16 peaks corresponding to 16 rounds of DES.

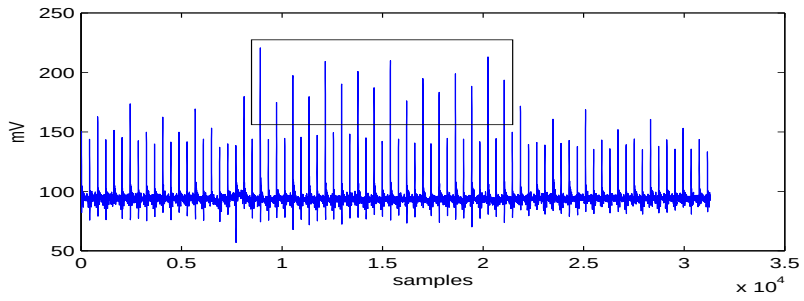

Fig. 3. Electromagnetic signal

In order to reduce the noise in electromagnetic signals, we repeated the encryption of each random text 10 times and calculated the averaged electromag-

netic signal of this text. The differential curves corresponding to different key hypothesis were computed using 3000 averaged electromagnetic signals of 3000 random texts. In our experiment, we examined four bits of an S-box output.

5.2 Performance metric

As the secret information of our electromagnetic signals are contained in sharp synchronized peaks, we define an attack-efficient index I as the ratio between the differential peak corresponding to the correct key (the expected peak) and the highest differential peak resulted from incorrect keys called ghost peaks [7]. These peaks are observed at the same moment τ when data are handled. If this index is greater than 1, the expected peak is higher than any ghost peak and the key detection is reliable. In contrast, if this index is smaller than 1, there exists a ghost peak which is higher than the expected peak. In the other words, the key detection method is not effective. It should be noted that this index is only applied in the attack using a reference device whose secret key is known.

In case of statistical multi-weight power analysis, we use another detection index based on the key probability of being guessed (see paragraph 4.2). Let's note F_{key} the occurrence frequency of each found key when we compute the multi-bit power analysis with differents sets of weigths. This index can be computed without any knowledge about the secret key, which happens in a real attack situation.

5.3 Hamming weight model

In this paragraph, we evaluate the performance of our proposed methods compared to Bevan's multi-bit DPA based on Hamming weight model. The experiment shows that with 3000 averaged electromagnetic signals, the four mono-bit DPA derived from b_1, b_2, b_3, b_4, the Bevan's multi-bit DPA and CPA do not allow to detect the secret key. Figure 4 represents 64 multi-bit differential curves corresponding to 64 key hypothesis. These curves are enlarged at the instant that data are handled. We observe that the differential curve of the correct key, which is plotted with star points, is covered by other curves. In this case, the attack-efficient index is equal to $I = 0.6$ and hence the secret key can not be detected. Briefly, the existing mono-bit and multi-bit power analysis do not allow to retrieve the secret key in this situation.

Firstly, we examine the performance of our selective multi-weight power analysis. We use an identical device with a known key and the 3000 texts to determine the set of weights α_i which optimizes the attack performance (i.e., maximize the index I). We fix $\alpha_1 = 1$ and vary α_2, α_3, α_4 from -2 to 2 with a step of 0.2. After testing all sets of weights, we realize that the index I is maximized when $\{\alpha_2 = 1, \alpha_2 = -1.8, \alpha_3 = -0.2, \alpha_4 = -1.6\}$. These are the optimal weights for multi-bit differential attack for this specific device. We then apply these weights instead of $\alpha_i = 1$, $i = 1..4$ in previous attack. The result of using these selective weights is illustrated in Fig. 5 where 64 differential curves

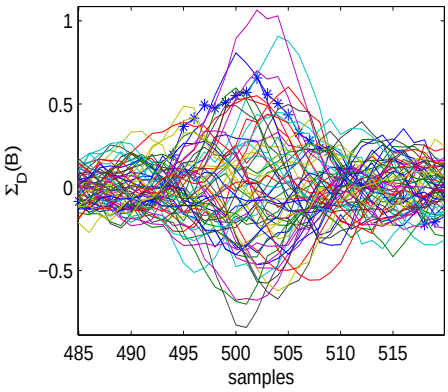

Fig. 4. Bevan's 4-bit DPA using Hamming weight model

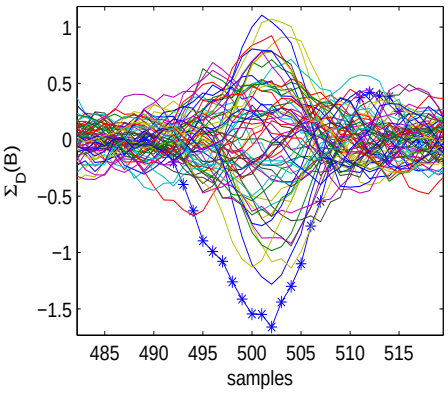

Fig. 5. Selective multi-weight power analysis using Hamming weight model

corresponding to 64 key hypothesis are traced. The differential peak corresponding to the correct key (i.e., curve with star points) becomes the highest one and the index I passes from 0.6 to 1.3. Hence, our proposed selective weights multi-bit power analysis can allow to detect the secret key from the Hamming weight information.

Secondly, we investigate the performance of the statistical analysis based attack presented in Section 4.2. Recall that the goal of this method is not to find out the optimal set of weights but to use all sets of weights to find out the key hypothesis of the highest frequency F_{key} of being designated as the correct key. As $\Sigma_{D_{abs}}(\mathcal{B})$ is used in the statistical analysis instead of $\Sigma_D(\mathcal{B})$, the weights α_i are positive. We thus fix $\alpha_1 = 1$ and vary three other weights from 0.2 to 2 with step of 0.2. We have then 1000 sets of weights. The frequency of being guessed as correct key when using these 1000 different sets of weights is depicted for the 64 key hypothesis in Fig. 6. The key corresponding to the highest F_{key} is exactly the correct key (hypothesis number 1) with $F_{key} = 58\%$. The result shows that the statistical analysis can help us to eliminate many wrong key hypothesis and indicate exactly the correct key. As the suitable weights for multi-bit power analysis is device category dependant and noise dependant, the statistical analysis becomes then a powerful attack method.

In short, by exploiting Hamming weight of data, the original multi-bit power analysis does not allow to detect the secret key. The choice of the weights in multi-bit combination is shown to play an important role. The weights should to be carefully selected and cryptographic device type aware. By using a reference device to select the suitable weights, the attack performance is significantly improved. Otherwise, our proposed statistical analysis is shown very effective to detect the secret key. The big advantage of this method is that it does not require any prior knowledge about the attacked device category and the optimal set of weights in multi-bit combination.

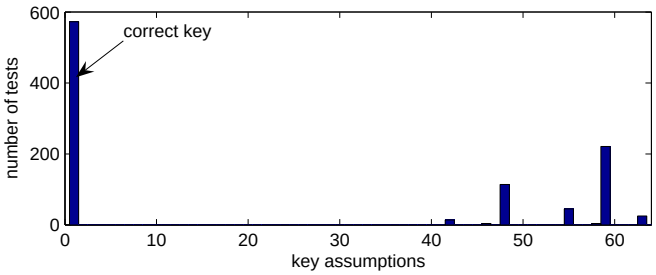

Fig. 6. Statistical multi-weight power analysis using Hamming weight model with 3000 texts

5.4 Hamming distance model

The purpose of this subsection is to show that our proposed methods still permit to improve the Hamming distance multi-bit power analysis. With the same 3000 electromagnetic signals used in the previous experiment, we can detect the correct key using the mono-bit DPA resulted from four bits b_1, b_2, b_3, b_4. The index I and the number of texts needed for key detection of each mono-bit DPA as well as for Bevan's multi-bit DPA is summarized in Table 2.

	b_1	b_2	b_3	b_4	Multi-bits
Index I	1.6	1.4	1.9	2.0	3.0
text number	1000	2000	700	600	150

Table 2. Attack-efficient index and number of texts for key detection

As the signs of all four differential peaks are the same, the identical weights multi-bit DPA proposed by Bevan has performed efficiently as shown in Fig. 7. If we apply the selective weights method with $\{\alpha_1 = 1, \alpha_2 = 2, \alpha_3 = 0.2, \alpha_4 = 1.8\}$, the index I has increased from 3.0 to 3.5. The attack performance has been enhanced.

The secret key is also easily detected by employing the statistical analysis. We always fix $\alpha_1 = 1$ and vary three other weights from 0.2 to 2 with a step of 0.2. According to this analysis, 100% of the tests guess the first hypothesis as the secret key. Figure 8 depicts the height of differential peaks corresponding to each set of weights and each key hypothesis. We observe that the differential peaks related to the key hypothesis number 1 are always the highest ones. This result validates the mathematical analysis presented in section 4.2 and confirms the effectiveness of our proposed method.

Besides the possibility to detect the secret key, the effectiveness of an attack is also measured by the number of required texts (i.e. number of side channel

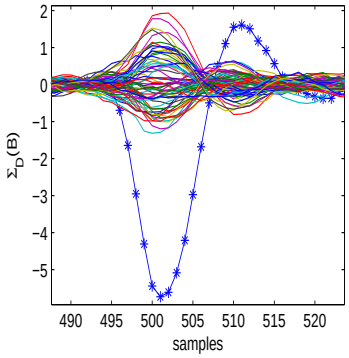

Fig. 7. Bevan's 4-bit power analysis with 3000 texts using Hamming distance model

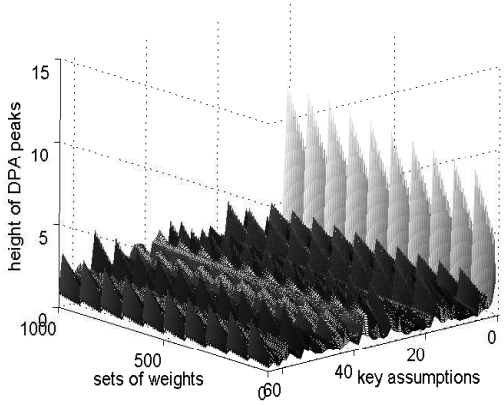

Fig. 8. Statistical multi-weight power analysis with 3000 texts using Hamming distance model

signals) for key detection. According to our experiments, both original (and selective) multi-bit power analysis and the statistical analysis can allow to retrieve the secret key with only 150 texts (see Fig. 9 and Fig. 10). In Fig. 9, the differential peak of the correct key is slightly higher than another differential peak of a wrong key. In fact, the attack-efficient index is only equal to $I = 1.05$. The distinction between the correct key and a wrong key is not much significant. On the contrary, with the same number of used text, the correct key is well distinguished with the wrong key if we use the statistical analysis method. More precisely, the F_{key} values corresponding to the correct key and most potential wrong key are 57% and 28% respectively.

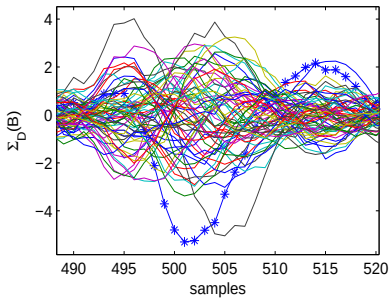

Fig. 9. Original 4-bit power analysis with 150 texts using Hamming distance model

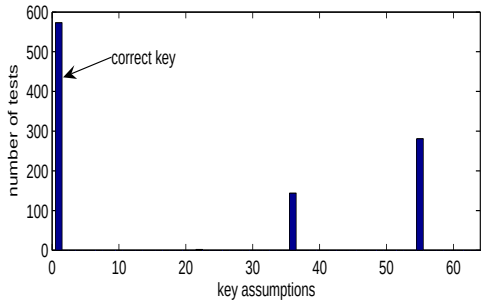

Fig. 10. Statistical multi-weight power analysis using Hamming distance with 150 texts

Note that, in case of wrong reference state the four differential peaks do not have the same sign, the identical weights will not be efficient anymore. The selective weights and the statistical analysis thus become the efficient solutions.

6 Conclusions

Firstly, this paper shows the relation between the correlation attack CPA and the consumption model and emphasizes the importance to use a precise model. However, finding a general consumption model for all kinds of devices is not a simple task because the model depends on each device and even on each targeted bits. Also, as noted in introduction for an hardware algorithm, the model needs generally to be reduced to a fairly small number of bit to restrict the key hypothesis.

Therefore, we proposed an alternative, which adapts more easily to a new device. This new method is to tune the coefficients of PPA rather than defining a precise power model.

In our paper, the bit-asymmetry hypothesis has also been introduced. Note that this hypothesis is valuable in many types of device due to some imbalance conception effects. The Hamming distance model, as introduced for CPA, which computes only the number of flipped bits, can not adopt this hypothesis. Instead of defining a new model, we proposed the selective weight multi-bit power analysis, which takes into account the non-equivalence of bits by using a reference device. The experimental results show that this multi-bit power analysis can be greatly improved by employing the suitable weights. Furthermore, we proposed a novel method based on statistical analysis of key detection results according to different possible sets of weights. The latter has also been demonstrated by experimental evaluation as an efficient and reliable attack method.

References

1. 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.
2. M.L. Akkar, L. Goubin: A Generic Protection Against High-Order Differential Power Analysis. *In proceedings of FSE 2003*, LNCS 2887, pp. 192 - 205, Springer-Verlag, 2003.
3. M.L. Akkar, R. Bevan, P. Dischamp, D. Moyart: Power Analysis, What Is Now Possible... *In proceedings of ASIACRYPT 2000*, LNCS 1976, pp. 489 - 502, Springer-Verlag, 2000.
4. R. Bevan, E. Knudsen: Ways to Enhance DPA. *In proceedings of ICISC 2002*, LNCS 2587, pp.327-342, Springer-Verlag, 2003.
5. R. Bevan: Estimation statistique et sécurité des cartes à puces, évaluation d'attaques DPA évolués. OCS, rapport de thèse, 2004.
6. E. Brier, C. Clavier, F. Olivier: Correlation Power Analysis with a Leakage Model, *In proceedings of CHES 2004*, LNCS 3156, pp. 16-29, Springer-Verlag, 2004.

7. C. Canovas, J. Clédière: What do S-boxes Say in Differential Side Channel Attacks? *Cryptology ePrint Archive*, http://eprint.iacr.org/, Report 20085/311, 2005.
8. S. Chari, C.S. Jutla, J.R. Rao, P. Rohatgi: Towards Sound Approaches to Counteract Power Analysis Attacks. *In proceedings of CRYPTO 1999*, LNCS 1666, pp. 348-412, Springer-Verlag, 1999.
9. S. Chari, J.R. Rao, P. Rohatgi: Template Attacks. *In proceedings of CHES 2002*, LNCS 2523, pp. 13-28, Springer-Verlag, 2002.
10. J.S. Coron, L. Goubin: On Boolean and Arithmetic Masking Against Differential Power Analysis. *In proceedings of CHES 2000*, LNCS 1965, pp. 231-237, Springer-Verlag, 2000.
11. J.S. Coron, P. Kocher, D. Naccache: Statistics and Secret Leakage. *In proceedings of Financial Cryptography*, LNCS 1972, pp. 157-173, Springer-Verlag, 2000.
12. P.N. Fahn, P.K. Pearson: IPA: A New Class of Power Attacks. *In proceedings of CHES 1999*, LNCS 1717, pp. 173-186, Springer-Verlag, 1999.
13. K. Gandolfi, C.Mourtel, F.Olivier: Electromagnetic Attacks: Concrete Results. *In proceedings of CHES 2001,* LNCS 2162, pp. 252-261, Springer, 2001.
14. L. Goubin, J. Patarin: DES and Differential Power Analysis: The Duplication Method. *In proceedings of CHES 1999*, LNCS 1717, pp. 158-172, Springer-Verlag, 1999.
15. S. Guilley, P. Hoogvorst, R. Pacalet: Differential Power Analysis Model and some Results *In proceedings of CARDIS 2004,* Kluwer Academic Publishers, pp. 127-142, 2004.
16. P. Kocher, "Timing attack on implementation of Diffe-Hellman, RSA, DSS and other systems," in *Advances in Cryptology - Crypto96*. New York: LNCS 1109, Springer Verlag.
17. P. Kocher, J. Jaffe, B. Jun: Differential Power Analysis. *In proceedings of CRYPTO 1999*, LNCS 1666, pp. 388-397, Springer-Verlag, 1999.
18. T.H. Le, J. Clédière, C. Servière, J.L. Lacoume, "Higher Order Statistic for Power Analysis Enhancement", *In proceedings of e-Smart 2006*, September 2006.
19. T.H. Le, J. Clédière, C. Canovas, C. Servière, J.L. Lacoume, B. Robisson, "A proposition for CPA enhancement", *In proceedings of CHES 2006*, October 2006.
20. R. Mayer-Sommer: Smartly Analysing the Simplicity and the Power of Simple Power Analysis on Smartcards. *In proceedings of CHES 2000*, LNCS 1965, pp. 78-92, Springer-Verlag, 2000.
21. T. S. Messerges, E. A. Dabbish, R. H. Sloan: Investigations of Power Analysis Attacks on Smartcards. *In proceedings of the USENIX Workshop on Smart Card Technology 1999*, `http://www.usenix.org/`, 1999.
22. T. S. Messerges, "Using Second-Order Power Analysis to Attack DPA Resistant Software", *In proceedings of CHES 2000*, LNCS 1965, pp. 238-251, Springer, 2000.
23. T. S. Messerges, E. A. Dabbish, R. H. Sloan: Examining Smart-Card Security under the Threat of Power Analysis Attacks. *IEEE Transactions on Computers*, Vol. 51, N5, pp. 541-552, May 2002.
24. E.Peeters, F-X. Standaert, J-J. Quisquater, "Power and Electromagnetic Analysis: Improved Model, Consequences and Comparisons", *Elsevier Science*, January 2006.
25. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery: Numerical Recipes in C++. *Cambridge University Press,* Second Edition, 1002pp, New York, 2002.
26. J.J. Quisquater, D. Samyde: Electromagnetic Analysis (EMA): Measures and Countermeasures for Smart Cards. *In proceedings of e-Smart 2001,* LNCS 2140, pp. 200-201, Springer, 2001.

27. J.R. Rao, P. Rohatgi: EMpowering Side-Channel Attacks. *Cryptology ePrint Archive,* http://eprint.iacr.org/, Report 2001/037, 2001.
28. J. R. Rao, P. Rohatgi, H. Scherzer, S. Tinguely : Partitioning Attacks : Or How to Rapidly Clone Some GSM Cards. *In proceedings of the 2002 IEEE Symposium on Security and Privacy*, pp. 31-41, IEEE Computer Society, 2002.
29. M. Rivain: Boîtes-S et attaques par canaux auxiliaires. *Rapport de Master "Cryptologie, Sécurité et Codage de l'Information"*, Universités INPG-UJF, Grenoble, 2006.
30. F.-X. Standaert, F. Mace, E. Peeters, J.-J. Quisquater: Updates on the Security of FPGAs Against Power Analysis Attacks. *In proceedings of ARC 2006*, LNCS 3985, pp. 335-346, Springer-Verlag, 2006.

A Appendices

We denote $\Delta_D^c(b_i)$ and $\Delta_D^w(b_i)$ the mono-bit differential curve values, computed from N power consumption signals $W_j = W(C_j), j = 1...N$, corresponding to the correct and the wrong key hypothesis respectively. We note v_i the value of the bit b_i given by the selection funtion $f(C_i, K_s)_{\mathcal{B}}$, $\tilde{b_i}$ its real value for the correct key hypothesis, and b_i' its previous value. In case of Hamming weight model, the group G_0 (or G_1) contains the power consumption signals corresponding to the texts of which v_i is equal to 0 (or 1). If the Hamming distance model is analysed, the groups G_1 and G_0 are constructed based on whether the bit b_i is or is not flipped, i.e. if v_i equals b_i' or not.

In case of using the correct key, all N consumption curves are correctly distributed in the two groups G_0 and G_1; the number of curves for each group is N_0 and N_1 respectively. If a wrong key is used, there are n_i curves which are wrongly placed in G_0 and G_1. The differential computings for Hamming weight and Hamming distance of a bit b_i at the instant τ where b_i is manipulated, are shown as following:

A.1 Differential curve value for the correct key in Hamming weight case

$$\begin{aligned}
\Delta_D^c(b_i) &= \frac{1}{N_1} \sum_{j=1, v_i=\tilde{b_i}=1}^{N_1} W_j - \frac{1}{N_0} \sum_{j=1, v_i=\tilde{b_i}=0}^{N_0} W_j \\
&= \frac{1}{N_1} \left(\sum_{j=1,0\to 1}^{N_1/2} W_j + \sum_{j=1,1\to 1}^{N_1/2} W_j \right) - \frac{1}{N_0} \left(\sum_{j=1,0\to 0}^{N_0/2} W_j + \sum_{j=1,1\to 0}^{N_0/2} W_j \right) \\
&= \frac{1}{N_1} \left(\frac{N_1}{2}(p_i + \delta_i) + 0 \right) - \frac{1}{N_0} \left(0 + \frac{N_0}{2} p_i \right) \\
\Delta_D^c(b_i) &= \frac{\delta_i}{2}
\end{aligned}$$

A.2 Differential curve value for a wrong key in Hamming weight case

$$
\begin{aligned}
\Delta_D^w(b_i) &= \frac{1}{N_1}\sum_{j=1,v_i=1}^{N_1} W_j - \frac{1}{N_0}\sum_{j=1,v_i=0}^{N_0} W_j \\
&= \frac{1}{N_1}\left(\sum_{j=1,v_i=\tilde{b_i}=1}^{N_1-n_i} W_j + \sum_{j=1,v_i=1,\tilde{b_i}=0}^{n_i} W_j\right) - \frac{1}{N_0}\left(\sum_{j=1,v_i=\tilde{b_i}=0}^{N_0-n_i} W_j + \sum_{j=1,v_i=0,\tilde{b_i}=1}^{n_i} W_j\right) \\
&= \frac{1}{N_1}\left(\sum_{j=1,0\to 1}^{(N_1-n_i)/2} W(C_j) + \sum_{j=1,1\to 1}^{(N_1-n_i)/2} W(C_j) + \sum_{j=1,0\to 0}^{n_i/2} W(C_j) + \sum_{j=1,1\to 0}^{n_i/2} W(C_j)\right) \\
&\quad -\frac{1}{N_0}\left(\sum_{j=1,0\to 0}^{(N_0-n_i)/2} W(C_j) + \sum_{j=1,1\to 0}^{(N_0-n_i)/2} W(C_j) + \sum_{j=1,0\to 1}^{n_i/2} W(C_j) + \sum_{j=1,1\to 1}^{n_i/2} W(C_j)\right) \\
&= \frac{1}{N_1}\left(\frac{N_1-n_i}{2}(p_i+\delta_i) + 0 + 0 + \frac{n_i}{2}p_i\right) - \frac{1}{N_0}\left(0 + \frac{N_0-n_i}{2}p_i + \frac{n_i}{2}(p_i+\delta_i) + 0\right) \\
\Delta_D^w(b_i) &= (1 - \frac{n_i}{N_1} - \frac{n_i}{N_0})\frac{\delta_i}{2}
\end{aligned}
$$

A.3 Differential curve value for the correct key in Hamming distance case

$$
\begin{aligned}
\Delta_D^c(b_i) &= \frac{1}{N_1}\sum_{j=1,b_i'\neq v_i=\tilde{b_i}}^{N_1} W_j - \frac{1}{N_0}\sum_{j=1,b_i'=v_i=\tilde{b_i}}^{N_0} W_j \\
&= \frac{1}{N_1}\left(\sum_{j=1,0\to 1}^{N_1/2} W_j + \sum_{j=1,1\to 0}^{N_1/2} W_j\right) - \frac{1}{N_0}\left(\sum_{j=1,0\to 0}^{N_0/2} W_j + \sum_{j=1,1\to 1}^{N_0/2} W_j\right) \\
&= \frac{1}{N_1}\left(\frac{N_1}{2}(p_i+\delta_i) + \frac{N_1}{2}p_i\right) - \frac{1}{N_0}(0+0) \\
\Delta_D^c(b_i) &= p_i + \frac{\delta_i}{2}
\end{aligned}
$$

A.4 Differential curve value for a wrong key in Hamming distance case

$$
\Delta_D^w(b_i) = \frac{1}{N_1}\sum_{j=1,b_i'\neq v_i}^{N_1} W_j - \frac{1}{N_0}\sum_{j=1,b_i'=v_i}^{N_0} W_j
$$

$$
\begin{aligned}
&= \frac{1}{N_1}\left(\sum_{j=1,b'_i\neq v_i=\tilde{b_i}}^{N_1-n_i} W_j + \sum_{j=1,b'_i\neq v_i\neq\tilde{b_i}}^{n_i} W_j\right) - \frac{1}{N_0}\left(\sum_{j=1,b'_i=v_i=\tilde{b_i}}^{N_0-n_i} W_j + \sum_{j=1,b'_i=v_i\neq\tilde{b_i}}^{n_i} W_j\right) \\
&= \frac{1}{N_1}\left(\sum_{j=1,0\to1}^{(N_1-n_i)/2} W(C_j) + \sum_{j=1,1\to0}^{(N_1-n_i)/2} W(C_j) + \sum_{j=1,0\to0}^{n_i/2} W(C_j) + \sum_{j=1,1\to1}^{n_i/2} W(C_j)\right) \\
&\quad -\frac{1}{N_0}\left(\sum_{j=1,0\to0}^{(N_0-n_i)/2} W(C_j) + \sum_{j=1,1\to1}^{(N_0-n_i)/2} W(C_j) + \sum_{j=1,0\to1}^{n_i/2} W(C_j) + \sum_{j=1,1\to0}^{n_i/2} W(C_j)\right) \\
&= \frac{1}{N_1}\left(\frac{N_1-n_i}{2}(p_i+\delta_i) + \frac{N_1-n_i}{2}p_i + 0 + 0\right) - \frac{1}{N_0}\left(0+0+\frac{n_i}{2}p_i + \frac{n_i}{2}(p_i+\delta_i)\right) \\
\Delta_D^w(b_i) &= (1-\frac{n_i}{N_1}-\frac{n_i}{N_0})(p_i+\frac{\delta_i}{2})
\end{aligned}
$$