Comparative Evaluation of Rank Correlation Based DPA on an AES Prototype Chip*

Lejla Batina[1], Benedikt Gierlichs[1], and Kerstin Lemke-Rust[2]

[1] K.U. Leuven, ESAT/SCD-COSIC and IBBT
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium
firstname.lastname@esat.kuleuven.be
[2] T-Systems GEI GmbH
Rabinstr. 8, 53111 Bonn, Germany
kerstin.lemke-rust@gmx.de

Abstract. We propose a new class of distinguishers for differential side-channel analysis based on nonparametric statistics. As an example we use Spearman's rank correlation coefficient. We present a comparative study of several statistical methods applied to real power measurements from an AES prototype chip to demonstrate the effectiveness of the proposed method. Our study shows that Spearman's rank coefficient outperforms all other univariate tests under consideration. In particular we note that Pearson's correlation coefficient requires about three times more samples for reliable key recovery than the method we propose. Further, multivariate methods with a profiling step which are commonly assumed to be the most powerful attacks are not significantly more efficient at key extraction than the attack we propose. Our results indicate that power models which are linear in the transition count are not optimal for the attacked prototype chip.

Keywords: Differential side-channel analysis, AES hardware, DPA, Rank correlation, Template attacks, Stochastic model.

1 Introduction

Side-channel attacks are a very active research area ever since the fundamental publications of Kocher et al. [9,10]. Especially with the evolving low-cost applications, *i.e.* pervasive security applications such as RFIDs and sensor nodes, side-channel attack resistance has become a matter of paramount importance. There are many practical attacks published and, at the same time, a firm line of work on theoretical aspects considering models for attackers, countermeasures *etc.*

It is widely believed that a correlation coefficient is the best statistical test for most power models to expose the right key among all the candidates. For

* This work was supported in part by the IAP Programme P6/26 BCRYPT of the Belgian State (Belgian Science Policy), by FWO projects G.0475.05, and G.0300.07, by the European Comission through the IST Programme under Contract IST-2002-507932 ECRYPT NoE, and by the K.U. Leuven-BOF.

T.-C. Wu et al. (Eds.): ISC 2008, LNCS 5222, pp. 341–354, 2008.

this purpose the common choice is Pearson's correlation coefficient [5] in conjunction with the Hamming weight or distance model [3]. On platforms like microcontrollers, where the relationship between the transitions on a data bus and the observable power dissipation is strikingly linear, this choice is theoretically founded. However, other parts of a microcontroller, *e.g.* registers, and different platforms such as ASICs and FPGAs do not necessarily follow this simple and linear relationship. We found that there are better matches for the function. Relaxing the assumption to simply a monotonic function led us to a new set of distinguishers based on nonparametric statistics. The results of our study show improvements with respect to efficiency, measured in the number of samples required, when we compare to the methods under consideration.

The contribution of our work is fourfold: i) We introduce a new class of side-channel distinguishers based on nonparametric statistics. We demonstrate the effectiveness of our approach by applying Spearman's rank correlation coefficient in a comparative study. ii) We show that rank correlation reaches the highest success rate amongst all univariate methods and in particular outperforms Pearson's correlation coefficient on this platform. Therefore it must be considered as an important distinguisher. iii) We give a detailed comparison of well known and adopted attacks on an AES hardware module. To the best of our knowledge the only related work was published by Mangard et al. in [13] and applied DPA to unprotected and to masked CMOS, but they varied the attacked intermediate results of AES and not the statistical distinguisher. iv) We present the first comparative study of templates and stochastic models on an AES hardware module. The work in [2] also discusses template attacks but the test platform is a DES hardware module.

This paper is organized as follows. Section 2 summarizes previous and related work. Section 3 describes the architecture of the targeted AES hardware module. In Section 4 we introduce a new class of side-channel distinguishers based on nonparametric statistics and in particular Spearman's rank correlation coefficient. Section 5 briefly explains the attacks and distinguishers used in our study. Experimental results from an unprotected prototype chip in standard CMOS (sCMOS) technology are provided in Section 6. Section 7 concludes the paper and outlines future work.

2 Previous Work

A decade ago Kocher et al. introduced successful attacks by measuring the power consumption during the execution of cryptographic algorithms [10]. It was demonstrated that one can use the physical leakage to easily recover secret keys if no countermeasures were deployed in the implementation. The demonstrated attack known as Differential Power Analysis (DPA) was applied against implementations of cryptographic algorithms running on smart cards. The surprising results gave rise to a new research area and there have been many contributions on both theoretical and practical aspects of power analysis. Other side-channels were also introduced such as as electromagnetic emanation [6,17], timing [9], acoustics [21] *etc.*

DPA attacks as introduced in [10] use a so-called selection function to sort a set of power consumption samples into subsets. The authors proposed simple boolean partitioning to divide the power samples in two subsets. However, the selection function can be extended to more bits and accordingly the power samples are sorted into multiple subsets. In this case we speak about a multi-bit DPA [14]. Selection functions are defined on an intermediate value of the cryptographic algorithm under attack that can be predicted using a key hypothesis and known data. It is afterwards a statistical question to find a key hypothesis that results in the highest correlation between the predicted values of a selection function and the sampled power consumption. Kocher et al. suggested to apply the difference of means test to find the right key. To such tests one usually refers as side-channel distinguishers. Other distinguishers referred to in the literature are Pearson's correlation coefficient [5], Mutual Information [7], Bayesian classification, e.g. template attacks introduced by Chari et al. [4] and the stochastic model by Schindler et al. [20]. Distinguishers are also sometimes used to assist other side-channel attacks. For example, Rechberger and Oswald proposed to use a DPA attack to find interesting points in time for templates in [18].

In this paper we introduce a new class of distinguishers based on nonparametric statistics, and compare them with other widely adopted techniques. We show that Spearman's rank correlation coefficient outperforms all other univariate methods under consideration on an AES ASIC implementation in sCMOS.

A similar comparative study of templates and stochastic models was performed by Gierlichs et al. [8], but they attacked an AES software implementation. To our best knowledge the only practical side-channel attacks on real AES chips were published by Örs et al. [15] and by Mangard et al. [13]. In [15] Pearson's correlation coefficient was used to perform a DPA attack on an unprotected implementation. The authors of [13] performed extended DPA by focusing on different choices for the selection function and not on statistical tests. However, DPA attacks on both unprotected and protected CMOS were performed. The important result was that the use of algorithmic masking in hardware does not increase the side channel resistance substantially in the presence of glitches. Another example where an ASIC platform was attacked can be found in [2]. The authors applied a template attack on a DES implementation focusing on the key schedule and they used a special power model.

3 Architecture of the AES Hardware Module

Our experimental platform is an AES hardware module from the SCARD chip. The chip is an outcome of the "Side-Channel Analysis Resistant Design Flow - SCARD" project led by the European Commission [22]. It contains an 8051 microcontroller with AES-128 co-processor in $0.13\,\mu m$ sCMOS and several secured logic styles.

In the sequel we focus on the AES module which is implemented in standard CMOS logic and includes no countermeasures against side-channel attacks. The AES module supports AES-128 [1] encryption and decryption in ECB mode.

The implementation uses four parallel one-stage pipelined implementations of the AES S-Box. A similar implementation is described in [12]. The module includes the following parts: data unit, key unit, and interface. The most important part is the data unit (see Fig. 1), which includes the AES operation. It is composed of 16 data cells ($C_{i,j}$, where $i, j \in \{0, 1, 2, 3\}$) and four S-Boxes. A data cell consists of flip-flops (able to store 1 byte of data) and some combinational logic in order to perform AddRoundKey operations. Load data is done by shifting the input data column by column into the registers of the data cells. The initial AddRoundKey transformation is performed in the fourth clock cycle together with the load of the last column. To calculate one round, the bytes are rotated vertically to perform the S-box and the ShiftRows transformation row by row. In the first clock cycle, the S-Box transformation starts only for the fourth row. Because of pipelining the result is stored two clock cycles later in the first row. S-boxes and the ShiftRows transformations can be applied to all 16 bytes of the state within five clock cycles due to pipelining. The architecture is very compact and suitable for smartcards and other wireless applications, which makes the attacks extremely relevant.

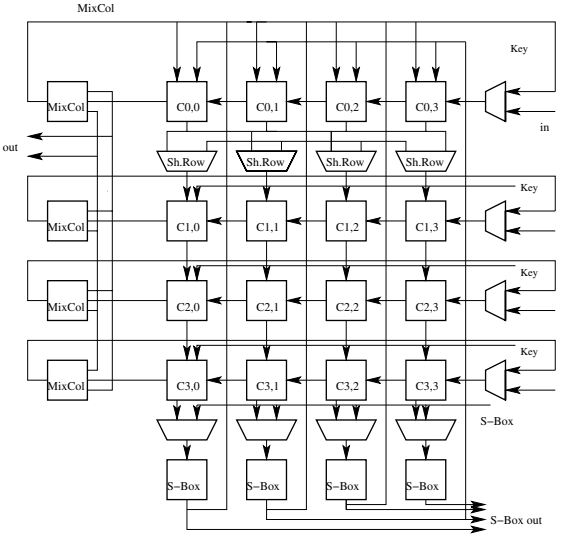

Fig. 1. The architecture of the AES module

The S-Boxes in the AES module are implemented by using composite field arithmetic, *i.e.* $GF(2^8)$ is considered as an extension field of $GF(2^4)$ as proposed by Wolkerstorfer et al. [23]. The original idea comes from Rijmen [19] as he suggested using subfield arithmetic in the crucial step of computing an inverse in the Galois Field.

We note here that the specifics of architecture do not cause the effectiveness of the attack proposed. The only fact about the platform that our distinguisher

takes advantage of is that the power model is not strictly linear in the transition count. This results in the attack performing better than other known methods.

4 Rank Correlation

Here we discuss some techniques which are usually referred to as nonparametric statistics [11] in the literature. Nonparametric equivalents to the standard correlation coefficient (*i.e.* Pearson's ρ) are Spearman's ρ, Kendal's τ, and Γ coefficient. These are also sometimes called nonparametric correlation coefficients. We demonstrate that in our experiments Spearman's correlation coefficient performs much better than the one of Pearson. This result suggests that one should consider alternative statistical tests in order to improve an attack's efficiency with respect to the number of required samples. This issue is also heavily platform-related so the influence of a power model is the most relevant one.

Figure 2 (left) shows the mean and the standard deviation of the power consumption as a function of the Hamming weight derived from a microcontroller moving data over its internal bus. The graph indicates that the relationship between power and the data's Hamming weight is very close to linear and that the empirical standard deviation is low. The plot on the right side of Fig. 2 on the other hand shows that the dependency between power consumed by a register update in the AES module and the Hamming distance of two subsequently stored data words, *i.e.* transition count, is not so close to linear. It is linear over small intervals but overall we can only say that it is a monotonic function. The large standard deviation can be caused either by algorithmic noise, *i.e.* it could reflect the power dissipation of the processing of other bits in parallel, or it can indicate that register updates with identical transition count do not lead to similar power consumption. The graph suggests that there might be a more suitable model than the strictly linear one. In general, one speaks also about the level of measurement, which can be interval, ordinal etc. In that case, one should look into nonparametric statistics, *i.e.* rank correlation, instead of Pearson's correlation coefficient.

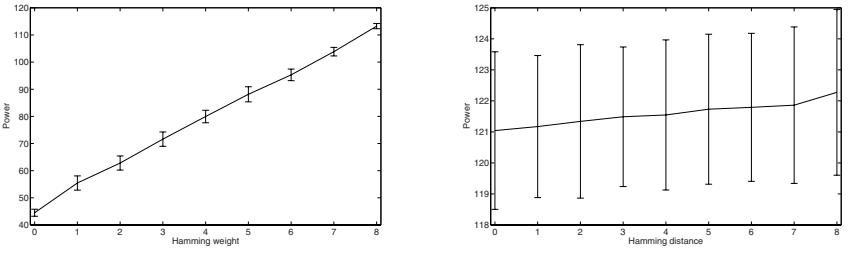

Fig. 2. Power dissipation of a micro-controller when accessing the data bus over Hamming weight (left); Power dissipation of the AES module when updating an 8bit register over Hamming distance (right)

The Pearson correlation coefficient is also known as the product-moment coefficient and it shows linear fits to (sometimes) noisy data. Pearson's ρ requires more information in the data than Spearman's coefficient, because it assumes the data is interval or ratio scaled, while Spearman's coefficient only expects it to be ordinal scaled. Data measured at the interval level are called interval scaled data, and data given with rank orders are called ordinal variables or rank variables.

Unlike Pearson's ρ, rank correlation does not assume a linear relationship between variables. A nonparametric (distribution-free) rank statistic is proposed by Spearman in 1904 and it can be also used as test of independence between two variables. More precisely, it is a measure of the strength of the association between two variables [11]. The Spearman rank correlation coefficient is a measure of monotonic relationship, which means that it can be used also if the relationship is non-linear. It was mainly meant to be used when the distribution of the data make Pearson's correlation coefficient unsuitable or misleading.

Let n be the number of pairs of values for variables X and Y defined on the (discrete) spaces $\mathcal{X}$ and $\mathcal{Y}$ and let d_i be the difference between each rank of corresponding values of X and Y. The formula to compute Spearman's rank correlation is:

$$\rho = 1 - \frac{6\sum_i {d_i}^2}{n(n^2-1)} \quad \text{or} \quad \rho = \frac{\sum_i (R_i - \overline{R})(S_i - \overline{S})}{\sqrt{\sum_i (R_i - \overline{R})^2 (S_i - \overline{S})^2}}, \tag{1}$$

where R_i and S_i are the ranks of variables X and Y. The latter formula is preferable if tied ranks exist, *i.e.* if the data to be ranked contains more than one value. In this case, Spearman's coefficient is actually computed as Pearson's correlation between ranks. The limited computational overhead is therefore given by the ranking process.

In Sect. 6 we show that this new side-channel distinguisher performs much better than Pearson's coefficient on our CMOS AES module. This is likely due to the specifics in the power consumption properties of the device.

Spearman's coefficient is, however, still insensitive to some types of dependence. Kendall's rank correlation gives a better measure of correlation and is also a better two sided test for independence. The Gamma statistic is preferable to both, Spearman or Kendall when the data contains many tied observations, but comes with the cost of increased computational complexity.

5 Established Side-Channel Attacks and Distinguishers

In this section we briefly recall known attacks which we apply to the AES hardware module.

5.1 Single-bit and Multi-bit DPA

(Single-Bit) DPA as proposed in [10] computes the DPA bias signal

$$\Delta_t = \frac{\sum_i p_{i,t} l_i}{\sum_i l_i} - \frac{\sum_i p_{i,t}(1-l_i)}{\sum_i (1-l_i)} \tag{2}$$

as the difference between the average of all measurements for which the so called selection function l_i evaluates to 1 and the average of all measurements for which the selection function evaluates to 0. The summations are taken over the q samples and the bias signal has to be computed for each time slice t within the power measurements p.

In [14] Messerges proposes to use selection functions based on several bits of the targeted intermediate value. He suggests to compute the DPA bias signal from the two subsets of power samples for which the selection function evaluates to maximal distance. For a selection function considering three bits for example, one would compute the difference of means of the subset "000" and the subset "111".

5.2 Pearson Correlation

In [3] Brier et al. suggest to estimate the Pearson correlation coefficient between a vector of power consumption samples p and a vector of power consumption predictions l

$$\rho_t = \frac{q\sum_i p_{i,t}l_i - \sum_i p_{i,t}\sum_i l_i}{\sqrt{q\sum_i p_{i,t}^2 - (\sum_i p_{i,t})^2}\sqrt{q\sum_i l_i^2 - (\sum_i l_i)^2}}\,. \tag{3}$$

The summations are taken over the q measurements and the correlation coefficient has to be estimated for each time slice t within the power curves p.

5.3 Multivariate Analysis

For multivariate analysis, it is assumed that the measurement vector $\boldsymbol{z} \in \mathbb{R}^m$ is distributed according to an m-variate Gaussian density

$$\mathcal{N}(\boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^m|\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\right] \tag{4}$$

where $\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{\Sigma}$ the covariance matrix of the normally distributed random variable $\boldsymbol{Z}$ with $\boldsymbol{\Sigma} = (\sigma_{uv})_{1\le u,v\le m}$ and $\sigma_{uv} := \mathbb{E}(Z_uZ_v) - \mathbb{E}(Z_u)\mathbb{E}(Z_v)$, $|\boldsymbol{\Sigma}|$ denotes the determinant of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$ its inverse. A Gaussian distribution is completely determined by its parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Both parameters can depend on the data processed, therefore enabling side channel leakage.

Both template attacks as well as stochastic methods consist of two-stages with different assumptions. The first stage is a profiling phase at which both key and plaintext or ciphertext are assumed to be known to the adversary. As result of profiling, the adversary obtains an m-variate Gaussian characterization of the key dependent physical leakage. The second stage is the key recovery stage (or classification) at which the adversary knows the plaintext or ciphertext, but not the key. At the second stage, the adversary's objective is key recovery.

Template Attacks. Roughly summarizing, there are three steps for building templates in the profiling stage. Firstly, the adversary computes the mean vector

$\boldsymbol{\mu}_k$ for each key dependency k. Secondly, m points in time are selected where significant differences are recognized among the mean vectors for different key dependencies. Finally, for each key dependency k the m-variate estimation of the noise is carried out resulting in the Gaussian distribution $\mathcal{N}(\boldsymbol{z}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. For a more detailed description of the algorithms we refer to [4,18,8].

Template classification computes the maximum likelihood, i.e., given n' measurements the adversary decides for the key hypothesis k^* that maximizes

$$\alpha_k := \prod_{i=1}^{n'} \mathcal{N}(\boldsymbol{z}_i, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{5}$$

among all k. Note that for practical purposes the log-likelihood is more adequate.

Stochastic Methods. Stochastic methods are an alternative approach for m-variate side channel analysis and have been introduced in [20] from which we only consider the so called 'maximum likelihood principle' in this paper.

In contrast to templates, stochastic methods estimate only one covariance matrix $\boldsymbol{\Sigma}$ that is used for all key dependent Gaussian densities $\mathcal{N}(\boldsymbol{z}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma})$. Furthermore, stochastic methods estimate the mean vector $\boldsymbol{\mu}_k$ by using general linear least squares targeting one key dependent and predictable intermediate result of the cryptographic implementation based on a power model. The power model used determines the vector subspace for the linear regression. Besides the Hamming weight model, a common power model is the bit-wise coefficient model saying that each bit of an intermediate result contributes to the overall power consumption. For example, for an 8-bit data item one uses a nine-dimensional vector subspace, spanned by the constant function 1 and eight single bits of the data item in the bit-coefficient model and a two-dimensional vector subspace spanned by the constant function 1 and the Hamming weight of the data item in the Hamming weight model. For a more detailed explanation of the applied algorithms at profiling we refer to [20,8].

Classification computes the maximum likelihood, i.e., given n' measurements the adversary decides for the key hypothesis k^* that maximizes

$$\alpha_k := \prod_{i=1}^{n'} \mathcal{N}(\boldsymbol{z}_i, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}) \tag{6}$$

among all k. As the covariance matrix $\boldsymbol{\Sigma}$ is identical, this is equivalent to minimizing the term $\sum_{i=1}^{n'} (\boldsymbol{z}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_i - \boldsymbol{\mu}_k)$.

6 Experimental Results

Our experimental platform is the sCMOS AES hardware module from the SCARD chip. The architecture of the AES co-processor is discussed in detail in Sect. 3. We obtained 50 000 power measurements p_i $(i = 1, \ldots, 50\,000)$ by sampling the voltage drop over a 50Ω resistor inserted in the chip's Vdd line at

a rate of 2 GS/s while the coprocessor was encrypting randomly chosen plaintext messages.

Let $x_i \in \{0,1\}^8$ ($i \in \{0,1,\ldots,15\}$) denote the plaintext byte. Accordingly, let $k_i \in \{0,1\}^8$ be the corresponding AES key byte. By $S(\cdot)$ we denote the AES S-box. The intermediate result chosen is

$$\Delta_{ii'} = S(x_i \oplus k_i) \oplus S(x_{i'} \oplus k_{i'}) \tag{7}$$

with $i \neq i'$. This intermediate result $\Delta_{ii'}$ is for example given by the differential of two adjacent data cells in the studied AES hardware architecture. $\Delta_{ii'}$ depends on two 8-bit inputs to the AES S-box $(x_i \oplus k_i, x_{i'} \oplus k_{i'})$. For the comparison of statistical tests, the targeted data cells are C0,0 and C0,1 of Fig. 1 in the remainder.

6.1 Difference of Means

The difference of means distinguisher failed at our scenario. We tested single-bit and multi-bit (two, three, and four bits) selection functions and considered up to $q = 25\,000$ power samples. No parameter combination led to key discovery.

6.2 Correlation Coefficients

Figure 3 shows the results we obtain for Pearson's and Spearman's correlation coefficient when using $q = 50\,000$ measurements and the correct key hypothesis. An attack with all 2^{16} key hypotheses still indicates the two correct key bytes when we reduce the number of measurements to $q = 5000$. Therefore we use at most 5000 measurements for the following comparison of Pearson's and Spearman's coefficient. To reduce computational complexity we assume in the remainder that the key byte k_i is known and test, whether the correct value of key byte $k_{i'}$ can be recovered. The number of key hypotheses is reduced to 2^8.

Figure 4 shows the efficiency of Pearson's correlation coefficient in detecting the correct key value from a given number q of power samples. We plot the maximum positive and minimum negative correlation (y-axis) over the number q of samples (x-axis) that we obtained for each key hypothesis on the overall time section. The correlation trace for the correct key hypothesis is plotted in

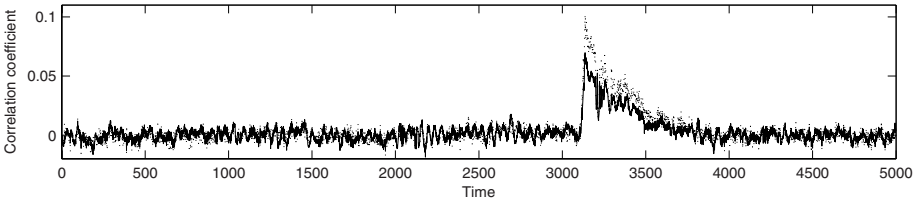

Fig. 3. Correlation coefficients for 8-bit Hamming distance as a function of time. Pearson (solid) and Spearman (dotted).

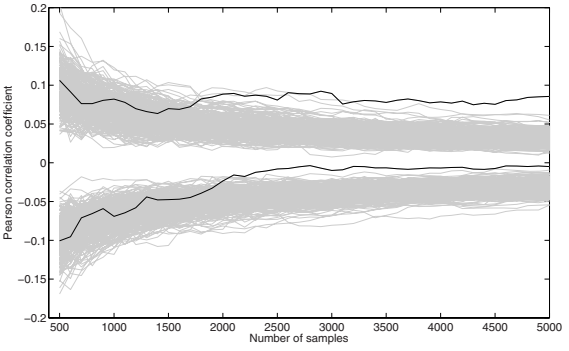

Fig. 4. Min and max Pearson correlation coefficient over number of samples; the black traces correspond to the correct key hypothesis

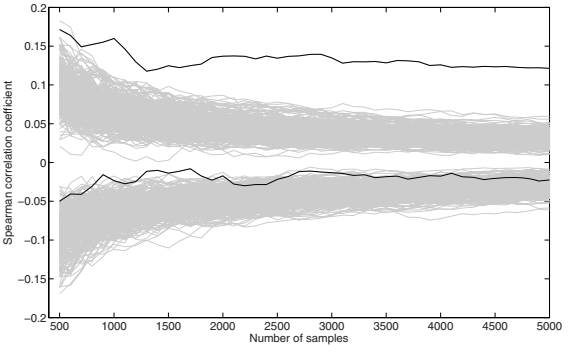

Fig. 5. Min and max Spearman Rank correlation coefficient over number of samples; the black traces correspond to the correct key hypothesis

black. One can observe that approximately $q = 4000$ power samples are required for key recovery.

Figure 5 depicts the performance of the Spearman rank correlation coefficient in the same manner as for Fig. 4. Obviously, significantly less samples (about $q = 1300$ or roughly 30% of the measurements needed by Pearson's correlation coefficient) are required for key recovery. Note that all numbers in this comparison have been confirmed by an experiment with a second data set and targeting a different cell in the hardware architecture. We report on the attacks' success rates as a function of the number of measurements in Sect. 6.5.

6.3 Stochastic Methods

Stochastic methods are applied in the bit-wise coefficient model, *i.e.* a nine-dimensional vector subspace is used for the estimation of the intermediate result

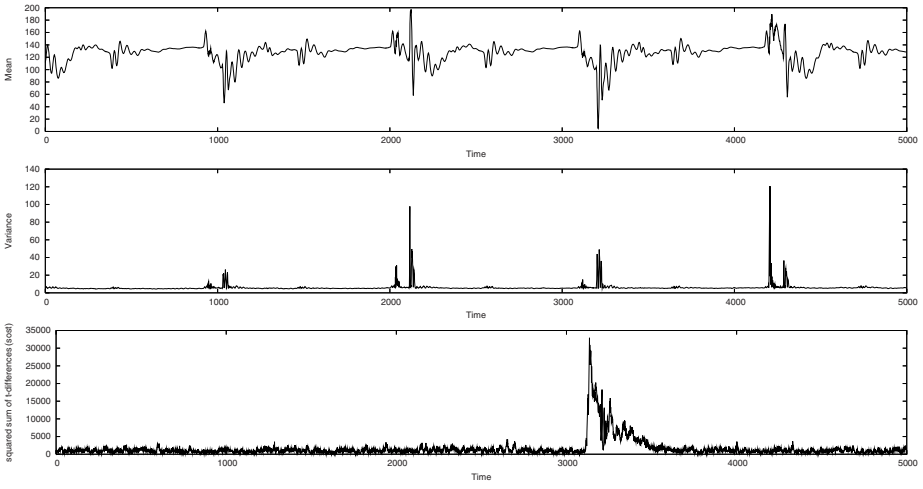

Fig. 6. Average curve (top), variance curve (middle) and sost curve (bottom) for the relevant time section derived from 40 000 measurements

in (7). For the profiling phase and classification phase we use complementary sets of measurements. In total, 40 000 measurements are used for profiling and 10 000 measurements for classification purposes.

Fig. 6 shows the mean and variance vector in the time frame for which we observed correlation peaks in the previous experiments. We chose the squared sum of t-differences (*sost*) trace (cf. [8]) for the identification of contributing points in time that is also shown in Fig. 6. For the computation of the *sost* trace the data dependent coefficients for the intermediate result (7) were estimated with 40 000 measurements. As result of this estimation one can compute the mean vector for each possible value of (7).

After identification of points of interest, the estimation of the mean vectors is repeated with 20 000 measurements and the estimation of the covariance matrix at the selected points in time is done with the other disjunctive set of 20 000 measurements.

Classification success rates are about 73% for $n' = 1000$ measurements, 97% for $n' = 2000$ measurements, and 100% for $n' = 3000$ measurements using ten selected points of interest ($m = 10$).

6.4 Template Attack

As for the stochastic method, we use a set of $n = 40\,000$ measurements for the profiling phase and a complementary set of $n' = 10\,000$ measurements for the classification phase. After the estimation of the mean vectors $\boldsymbol{\mu}_k$ we compute the *sost* trace (cf. [8]) which indicates interesting points in time. As one can see in Fig. 7 (left) the *sost* trace points toward a very narrow time window. The *sost* trace within this time window, see Fig. 7 (right), looks very similar to

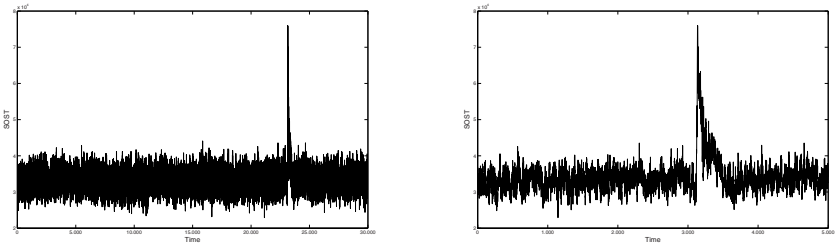

Fig. 7. *sost* trace template attack for the overall (left) and the relevant time frame (index 20 000 to 25 000, right)

the *sost* trace shown in Fig. 6. Again we experiment with the number and the distribution of points of interest. For the sake of comparison we report about the best results we could achieve using $m = 10$ points.

Once the points of interest are chosen, we estimate the covariance matrices $\mathbf{\Sigma}_k$ from the same set of 40 000 measurements. It turns out that classification of samples from the remaining set of measurements leads only to negligible success rates. We assume that the failure is caused by the number of measurements we use. If the number of measurements is too small, the estimations of the $\boldsymbol{\mu}_k$ and in particular of the $\mathbf{\Sigma}_k$ are bad. Since the stochastic method achieves reasonable success rates, we decide to estimate only a single, key-independent covariance matrix $\mathbf{\Sigma}$. But again, the template attack achieves only minor success rates. For a final test, we follow the suggestion of [16] and do not estimate the covariance matrix $\mathbf{\Sigma}$ at all, but simply set it to the unity matrix. This choice reflects the assumption that the side-channel leakage at the selected points in time is independent. This setting leads to classification success rates of about 32% for $n' = 1000$ samples, 63% for $n' = 2000$ samples, and 82% for $n' = 3000$ samples.

6.5 Overall Comparison

The complete results for the comparison are given in Table 1. The success rates refer to various numbers of measurements, ranging from 500 to 3000 curves, that were used for an attack and are derived from 500 experiments each using a set of randomly chosen measurements. It is obvious that Spearman's coefficient outperforms all other univariate distinguishers in all cases.

When comparing the performance of the template attack and the stochastic method, we conclude that in our scenario the stochastic method leads to better success rates and is the method of choice. The authors of [8] observed that the stochastic method can lead to better results than the template attack if the number of measurements for the profiling step is not sufficiently large. To enable the template attack on this AES hardware module, key-dependent covariance matrices $\mathbf{\Sigma}_k$ need to be replaced with a single matrix $\mathbf{\Sigma}$ and furthermore this matrix has to be set to the identity map. This fact might deserve further research

Table 1. Success rates for the distinguishers for the given number of measurements: distance of means, Pearson correlation, Spearman rank correlation, template attack with a single covariance matrix set to the unity matrix, stochastic model

No.	DoM	Pearson Corr.	Sp. Rank. Corr.	Template Attack	Stochastic Model
500	-	13.6%	39.6%	15.6%	41.4%
1000	-	29.8%	77.8%	31.8%	73.4%
2000	-	64.2%	99.0%	63.2%	96.8%
3000	-	84.0%	100.0%	82.4%	100%

on the application of template attacks if the target of evaluation is a hardware module. A more detailed investigation of this matter is beyond the scope of this paper but will be part of our future work.

7 Conclusions

We propose a new class of side-channel distinguishers based on nonparametric statistics. We compare the efficiency of Spearman's rank correlation coefficient to that of other known attack methods when extracting the key from an AES-128 prototype chip. The results allow two conclusions. Spearman's rank correlation coefficient performs best amongst the univariate methods we apply. In particular, it outperforms Pearson's correlation coefficient by far, requiring only about 30% of the number of samples. This observation indicates that a power model which is linear in the transition count is suboptimal. The observation is naturally bound to the targeted device and different platforms can lead to different results. Moreover, multivariate methods with a profiling step which are commonly considered the most powerful attacks require much more measurements and do not perform significantly better than the proposed distinguisher in this experiment. A detailed investigation of this matter is beyond the scope of this paper, but part of our future research.

References

1. FIPS 197: Announcing the Advanced Encryption Standard (AES) (November 2001), `http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf`
2. El Aabid, M.A., Guilley, S., Hoogvorst, P.: Template Attacks with a Power Model. Cryptology ePrint Archive, Report, 2007/443 (2007), `http://eprint.iacr.org/`
3. Brier, E., Clavier, C., Olivier, F.: Correlation power analysis with a leakage model. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 16–29. Springer, Heidelberg (2004)
4. Chari, S., Rao, J.R., Rohatgi, P.: Template Attacks. In: Kaliski, B.S., Koç, Ç., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 13–28. Springer, Heidelberg (2003)
5. Coron, J.-S., Kocher, P.C., Naccache, D.: Statistics and secret leakage. In: Frankel, Y. (ed.) FC 2000. LNCS, vol. 1962, pp. 157–173. Springer, Heidelberg (2001)
6. Gandolfi, K., Mourtel, C., Olivier, F.: Electromagnetic Analysis: Concrete Results. In: Koç, Ç.K., Naccache, D., Paar, C. (eds.) CHES 2001. LNCS, vol. 2162, pp. 251–261. Springer, Heidelberg (2001)

7. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual Information Analysis - A Generic Side-Channel Distinguisher. In: Oswald, E., Rohatgi, P. (eds.) CHES 2008. LNCS, vol. 5154, pp. 426–442. Springer, Heidelberg (2008)
8. Gierlichs, B., Lemke-Rust, K., Paar, C.: Templates vs. Stochastic Methods. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 15–29. Springer, Heidelberg (2006)
9. Kocher, P.C.: Timing attacks on implementations of Diffie-Hellman, RSA, DSS and other systems. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 104–113. Springer, Heidelberg (1996)
10. Kocher, P.C., Jaffe, J., Jun, B.: Differential Power Analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999)
11. Lehman, E.L., D'Abrera, H.J.M.: Nonparametrics: Statistical Methods Based on Ranks. Prentice-Hall, Englewood Cliffs (1998)
12. Mangard, S., Aigner, M., Dominikus, S.: A Highly Regular and Scalable AES Hardware Architecture. IEEE Trans. Computers 52(4), 483–491 (2003)
13. Mangard, S., Pramstaller, N., Oswald, E.: Successfully Attacking Masked AES Hardware Implementations. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 157–171. Springer, Heidelberg (2005)
14. Messerges, T.S.: Power Analysis Attacks and Countermeasures on Cryptographic Algorithms. PhD thesis (2000)
15. Örs, S.B., Gürkaynak, F., Oswald, E., Preneel, B.: Power-analysis attack on an ASIC AES implementation. In: Proceedings of the International Conference on Information Technology (ITCC), Las Vegas, NV, USA, April 5-7 (2004)
16. Oswald, E., Mangard, S.: Template Attacks on Masking – Resistance is Futile. In: Abe, M. (ed.) CT-RSA 2007. LNCS, vol. 4377, pp. 243–256. Springer, Heidelberg (2006)
17. Quisquater, J.-J., Samyde, D.: ElectroMagnetic Analysis (EMA): Measures and Couter-Measures for Smard Cards. In: Attali, I., Jensen, T.P. (eds.) E-smart 2001. LNCS, vol. 2140, pp. 200–210. Springer, Heidelberg (2001)
18. Rechberger, C., Oswald, E.: Practical Template Attacks. In: Lim, C.H., Yung, M. (eds.) WISA 2004. LNCS, vol. 3325, pp. 440–456. Springer, Heidelberg (2005)
19. V. Rijmen.: Efficient Implementation of the Rijndael SBox, `http://www.iaik.tugraz.at/RESEARCH/krypto/AES/old rijmen/rijndael/sbox.pdf`
20. Schindler, W., Lemke, K., Paar, C.: A Stochastic Model for Differential Side Channel Cryptanalysis. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 30–46. Springer, Heidelberg (2005)
21. A. Shamir, E. Tromer.: Acoustic cryptanalysis, `http://theory.csail.mit.edu/~tromer/acoustic/`
22. The SCARD project, `http://www.scard-project.eu/`
23. Wolkerstorfer, J., Oswald, E., Lamberger, M.: An ASIC Implementation of the AES SBoxes. In: Preneel, B. (ed.) CT-RSA 2002. LNCS, vol. 2271, pp. 67–78. Springer, Heidelberg (2002)