Analysis of Nonparametric Estimation Methods for Mutual Information Analysis

Alexandre VENELLI[1,2]

[1] IML - ERISCS Université de la Méditerranée,
Case 907, 163 Avenue de Luminy
13288 Marseille Cedex 09, FRANCE
[2] Vault-IC France, an INSIDE Contactless Company
Avenue de la Victoire, Z.I. Rousset,
13790 Rousset, FRANCE
avenelli@insidefr.com

Abstract. Mutual Information Analysis (MIA) is a side-channel attack introduced recently. It uses mutual information, a known information theory notion, as a side-channel distinguisher. Most previous attacks use parametric statistical tests and the attacker assumes that the distribution family of the targeted side-channel leakage information is known. On the contrary, MIA is a generic attack that assumes the least possible about the underlying hardware specifications. For example, an attacker should not have to guess a linear power model and combine it with a parametric test, like the Pearson correlation factor. Mutual information is considered to be very powerful however it is difficult to estimate. Results of MIA can therefore be unreliable and even bias. Several efficient parametric estimators of mutual information are proposed in the literature. They are obviously very efficient when the distribution is correctly guessed. However, we loose the original goal of MIA which is to assume the least possible about the attacked devices. Hence, nonparametric estimators of mutual information should be considered in more details and, in particular, their efficiency in the side-channel context. We review some of the most powerful nonparametric methods and compare their performance with state-of-the-art side-channel distinguishers.

Keywords: Side-channel analysis, mutual information analysis, entropy estimation, nonparametric statistics

1 Introduction

Side-channel analysis is a technique that uses information leaked by a physical implementation of cryptographic algorithms. The concept of using side-channel information to break a cryptosystem was introduced by Kocher [12]. In his paper, Kocher analyses differences in the computation time of certain cryptographic operations that depend on a secret. On embedded devices, monitoring the power consumption or recording the electromagnetic radiations is easy to realize and

is very revealing of the computations executed by the system. Statistical tests are used in side-channel cryptanalysis so that an attacker does not need to know precise implementation details in order to extract secret keys. Generally, once the side-channel information is recorded from a device, the attacker post processes it and evaluates it by some statistical analysis. In 1999, Kocher et al. [13] introduced the concept of Differential Power Analysis (DPA), a side-channel attack that uses the difference of means as statistical test. The attacker makes a key hypothesis and partitions side-channel measurements into two sets depending on the value of a key-dependent computation in the cryptographic algorithm. Then, the adversary computes the difference of means of the two sets for each key hypothesis. If the difference shows distinct peaks, the corresponding key hypothesis is assumed to be correct.

A lot of research in the side-channel domain consist in proposing relevant statistical tests to enhance the results of these attacks. In 2004, Brier et al. [4] introduce the use of the Pearson correlation factor as statistical test. The corresponding side-channel attack is called Correlation Power Analysis (CPA). This correlation factor seems to give the best results on the vast majority of embedded devices. This is mostly due to the technology Complementary Metal Oxyde Semiconductor (CMOS) that is used in the industry to build smart cards. It is commonly assumed that the number of bits of a bus or an internal register, that flip at a given time, is linearly proportional to the current absorption of the device [17]. This supposition seems correct on most CMOS systems. As the CPA finds the linear dependencies between power consumption curves and a leakage function based on a key guess and a plaintext value, it is very powerful.

Making an assumption on the power consumption characteristic details of a device can be considered a strong hypothesis for an attacker. In 2008, Gierlichs et al. [8] propose a side-channel attack effective without any knowledge or restrictive assumption about the power model of the device, i.e. the relationships between the power consumption of the device and its processed data. The attack is called Mutual Information Analysis (MIA) and uses mutual information as a side-channel distinguisher. When the CPA only records linear relations, the estimation of mutual information does not need to have assumptions about the dependencies of the variables. Even if the MIA is more generic, in practice the CPA often performs better on CMOS logic. However for devices using special types of logic, as the dual-rail logic [10, 5], the assumption that the relation between processed data and power consumption is linear should not hold.

The poor performance of the MIA compared to the CPA [8, 20, 24, 29] may not be inherent to its properties but due to its inefficient estimation in most cases. The MIA, as presented in the original paper [8], uses histograms, the less effective method to estimate mutual information. Different authors [24, 16] propose more efficient techniques to estimate mutual information, however the techniques make assumptions on the power model of the device, i.e. parametric techniques. In this paper, we focus our analysis on nonparametric methods for mutual information estimation. We then review and evaluate the efficiency of each method when applied with MIA. Finally, we compare state-of-the-art side-

channel distinguishers with the most performant nonparametric estimators of mutual information.

Section 2 summarizes the fundamentals of information theory as well as introduces generalized mutual information. In Section 3, we study some of the most used statistical tests of the the side-channel literature. Section 4 reviews classical methods of estimation of mutual information, and more particularly, nonparametric methods. We evaluate the different techniques of estimation in the context of side-channel analysis on various setups in Section 5. Section 6 concludes the article.

2 Information Theory Framework

Shannon in [26] laid down foundations of information theory in communication systems. The entropy in a signal corresponds to the quantity of information it contains. In the context of cryptanalysis and more particularly side-channel attacks, one is interested in how much information is generated from a cryptographic device. If the device leaks information when it processes a secret, an attacker could recover the leakage through side-channel analysis and hence obtain information, e.g. bits of the secret. Mutual information is a measure closely related to entropy. It is a special case of the notion of relative entropy which records something close to a distance between two distribution functions.

2.1 Basics on Probability Theory

Let X be a random variable which takes on a finite set of values $\{x_1, x_2, \ldots, x_n\}$. Let $\mathbb{P}(X = x_i)$ be the probability distribution of X. Hence, the function $f : x \mapsto \mathbb{P}(X = x)$ is often called the probability density function (pdf) of X. Similarly, we define the function $F : x \mapsto \mathbb{P}(X \leq x)$ as the cumulative distribution function (cdf) of X.

The entropy of X is defined as

$$H(X) = -\sum_x f(x) \log(f(x)).$$

Let $H(X)$ and $H(Y)$ be the entropy of X and Y respectively. The joint entropy of X and Y is defined as

$$H(X, Y) = -\sum_{x,y} \mathbb{P}(X = x, Y = y) \log(\mathbb{P}(X = x, Y = y)).$$

The conditional entropy of X given Y, noted $H(X|Y)$, is defined as

$$H(X \mid Y) = \sum_y \mathbb{P}(Y = y) H(X \mid Y = y), \text{ with}$$
$$H(X \mid Y = y) = -\sum_x \mathbb{P}(X = x, Y = y) \log(\mathbb{P}(X = x, Y = y)).$$

The mutual information $I(X;Y)$ quantifies the amount of information between two variables X and Y. It is defined as

$$I(X;Y) = H(X) - H(X \mid Y).$$

Mutual information is in fact a special case of the Kullback-Leibler (KL) divergence [15]. This divergence measures the dissimilarity between two distributions. Let f and g be two pdf of a random variable X. The KL divergence, also called relative entropy, is then defined as

$$D_{\mathrm{KL}}(f \parallel g) = \sum_x f(x) \log \frac{f(x)}{g(x)}.$$

The mutual information can then be described as

$$I(X;Y) = D_{\mathrm{KL}}(f(x,y) \parallel f(x)f(y)).$$

2.2 Generalized Mutual Information

Let X be a discrete random variable as previously defined. The Rényi entropy [25] of order α is defined as

$$H_\alpha(X) = \begin{cases} \frac{1}{1-\alpha} \log \sum_x f(x)^\alpha & \text{for } \alpha \geq 0, \alpha \neq 1 \\ -\sum_x f(x) \log f(x) & \text{for } \alpha = 1. \end{cases}$$

The entropy of Shannon corresponds to $H_1(X)$. With the previous definition of Rényi entropy, we can introduce the quantity

$$I_\alpha(X;Y) = H_\alpha(X) + H_\alpha(Y) - H_\alpha(X,Y).$$

The quantity I_α has the following property:

$$I_\alpha \geq 0 \quad \text{if and only if } \alpha = 0 \text{ or } 1.$$

The value I_α only corresponds to the classical definition of mutual information in these two cases. However in [23, Basic Theorem, Ch. 3], the authors consider the case $\alpha = 2$. Using the collision entropy H_2, they call the quantity $I_2(X;Y)$ Generalized Mutual Information (GMI) where either the random variable X or Y is uniformly distributed. In this case, the GMI and the classical mutual information are both strictly positive and measure both the independence between two variables. The GMI is particularly interesting as there is a more efficient method of estimation based on kernel estimators (Sec. 4.3) [22].

3 Classical Side-Channel Distinguishers

3.1 Differential Side-Channel Model

Let K be a random variable representing a part of the secret. Let X be a random variable representing a part of the input, or output, of the cryptographic

algorithm. Suppose an attacker wants to target an intermediate value computed with the function F that takes as parameters X and K. Let L be a random variable representing the side-channel leakage generated by the computation of $F(X, K)$. In practice, the attacker is only able to obtain N realizations of the random variable L, noted $V_L = (l_1, \ldots, l_N)$, as he inputs N different values of X, noted $V_X = (x_1, \ldots, x_N)$. Using a distinguisher function D, he combines these two vectors plus an hypothesis on the value of the secret k'. If the distinguisher D is relevant and if the leakage vector V_L brings enough information on $F(X, K)$, then the correct value k taken by K can be recovered. In the literature, some worked on creating a model for $F(X, K)$. For example, taking the Hamming weight of the output of F [18], the Hamming distance [4] or simply its value [8] was considered. Other researches were conducted on the distinguisher function D that plays a fundamental role in the attack. Depending on the choice, the function is able to extract more or less information from the side-channel leakages. We briefly review in the following the statistical tests used as function D proposed in the literature.

3.2 Difference of Means

Kocher et al. [13] proposed the concept of differential side-channel attack in 1999. In their original paper, the authors use a Difference of Means (DoM) as distinguisher function. It is in fact a simplified student T-test, a well-known statistical test. For simplicity reasons, we suppose the function $F(X, K)$ only outputs the least significant bit of the result. Let k' be an hypothesis on the secret. The attacker can form two sets:

$$G_0 = \{L \mid F(x_j, k') = 0\} \quad \text{and} \quad G_1 = \{L \mid F(x_j, k') = 1\}.$$

Finally, he computes the difference of means between the two partitions as:

$$\Delta_{k'} = \frac{\sum_{l \in G_0} l}{|G_0|} - \frac{\sum_{l \in G_1} l}{|G_1|}.$$

If the attacker detects a significant difference between the two sets, he can suppose that the hypothesis k' is correct.

3.3 Pearson Correlation Factor

Introduced by Brier et al. [4] in the context of side-channel analysis, the Pearson correlation factor, also called Pearson rho or product-moment correlation, measures linear dependencies between two variables X and L. The authors called the attack Correlation Power Analysis (CPA). In practice, if the attacker is only able to obtain N realizations of the leakage function, then the formula is:

$$\rho_{k'}(X, L) = \frac{N \sum_i l_i F(x_i, k') - (\sum_i l_i \sum_i F(x_i, k'))}{\sqrt{N \sum_i l_i^2 - (\sum_i l_i)^2} \sqrt{N \sum_i F(x_i, k')^2 - (\sum_i F(x_i, k'))^2}}.$$

Pearson correlation calculations are based on the assumption that both X and L values are sampled from a normal distribution. Hence, Pearson rho is part of parametric tests. On the contrary, methods that do not assume a particular distribution family for the data are said to be nonparametric.

3.4 Cluster Analysis

Differential Cluster Analysis (DCA) was introduced in [3]. It uses classical cluster analysis statistics in the side-channel analysis context. The principle of cluster analysis is to group similar objects into respective categories, i.e. clusters, and then use a statistical method in order to discover structures in the observed data. In side-channel analysis, the clusters often correspond to the outputs of the attacked intermediate value, which is similar to the mutual information technique presented Sec. 4. Amongst the statistical function used to characterize clusters proposed in [3], the use of variance seems particularly suited.

3.5 Nonparametric Correlation Statistics

Nonparametric tests make no assumptions about the distribution parameters of the variables. They do not rely on the estimation of parameters such as the mean or the standard deviation. Therefore, they are often called parameter-free or distribution-free methods. The most commonly used nonparametric equivalents to Pearson correlation factor are Spearman R, Kendall tau and coefficient Gamma. The coefficient Gamma [9] is similar to Kendall tau and is not very relevant in our analysis.

The use of the Spearman R has been proposed in [2]. Spearman R assumes that the variables are on a rank ordered scale. If several values of the variables are equal, which is the case in the context of side-channel analysis, the formula for Spearman R is the same as for Pearson's rho. The rank of identical values is the mean of their respective ranks.

Kendall tau [11] is similar in terms of results to Spearman R. However, its computation and its statistical meaning is different. Kendall measures the degree of relationships between variables whereas Pearson and Spearman test the null hypothesis that there is no relationships between variables. There is different versions of Kendall statistic. In our context, one should use the coefficient that makes adjustments for tied values:

$$\tau_b = \frac{N_c - N_d}{\sqrt{(N(N-1)/2 - t)(N(N-1)/2 - u)}},$$

where N_c is the number of pairs ranked in the same order on both variables, N_d is the number of pairs ranked differently on the variables, t is the number of tied values in the first variable, u is the number of tied values in the second and N is the number of observations.

In [29], the authors propose to use other nonparametric statistics: the Kolmogorov-Smirnov (K-S) test and the Cramér-von Mises (CVM) test. These

tests are very similar to the DCA and the mutual information analysis. Indeed, the data is placed in different clusters, each typically covering a range of values of the attacked intermediate value. Let $F_X(x)$ and $F_L(x)$ be the empirical cumulative distribution functions of the sample populations X and L. The K-S test between the variables X and L is:

$$D_{\mathrm{KS}}(X \parallel L) = \sup_x |F_X(x) - F_L(x)| .$$

The CVM test is defined similarly as:

$$D_{\mathrm{CVM}}(X \parallel L) = \sum_x (F_X(x) - F_L(x))^2.$$

4 Estimators of Mutual Information

Gierlichs et al. in [8] propose the use of mutual information as side-channel distinguisher in an attack called Mutual Information Analysis (MIA). The authors present this method as an interesting alternative to the powerful CPA as the attacker does not have to assume a particular power consumption model for the targeted device (Sec. 3.1). Indeed, mutual information records both linear and non-linear relationships between variables while CPA only measures linear ones. In theory, MIA should be considered more generic as the attacker makes less assumptions about the device. However in practice, the results of MIA are not good compared to CPA [29, 24, 20]. In fact, the efficiency of MIA is closely related to its chosen estimator of mutual information. Some authors studied parametric estimation methods and their efficiency combined with MIA [24, 7, 16]. On the contrary, nonparametric estimators are not thoroughly researched [28], although they fit the original purpose of MIA more suitably.

4.1 Parametric vs Nonparametric Estimation

There are two basic approaches to estimation: parametric and nonparametric. In this paper, we restrict ourselves to the nonparametric field. Parametric estimation makes assumptions about the regression function that describes the relationship between dependent variables. Therefore, the density function will assume that the data are from a known family of distributions, such as normal, and the parameters of the function are then optimized by fitting the model to the data set. Nonparametric estimation, by contrast, is a statistical method that has no meaningful associated parameters. There is often no reliable measure used for the choice of the parameters. However, this type of estimation seems more suitable to the original purpose of the MIA: a generic side-channel attack that makes the less assumptions possible. Hence, this paper seeks to introduce efficient nonparametric pdf estimation methods in the context of side-channel analysis.

4.2 Histogram-based Estimator

The most simple and time efficient method to estimate pdf is using histograms. An histogram consists in a partition of the range of values of each variables into b discrete bins of equal length. The pdf of each bin is estimated by the relative frequency of occurrence of samples in the bin. Let X be a random variable with N realizations. The b partitions are defined as: $a_i = [o + ih, o + (i + 1)h]$ where o the value of the origin, h is the width of the bins and $i = 0, \ldots, b - 1$. Let k_i be the number the measurements of X that lie in the interval a_i. The pdf f_i of X can be approximated as

$$\hat{f}_i = \frac{k_i}{N}.$$

As this method is nonparametric, its parameters are not easily determined. The choice of the number of bins b or their width can be non-trivial. In any case, the partitioning must be the same for both variables. Even if Histogram-based Estimation (HE) is computationally efficient, its results contain more statistical errors than other methods.

4.3 Kernel Estimator

Kernel Density Estimation (KDE) constructs a smooth estimate of the density by centering kernel functions at data samples [19]. The kernels weight the distance of each points in the sample to the reference point depending on the form of the kernel function and according to a given bandwidth h. In KDE, h plays a similar role as b in HE. In fact, the uniform kernel function forms an histogram. Gaussian kernels are most commonly used and we use them as well in this study. Let $\{x_1, \ldots, n_N\}$ be N realizations of the random variable X. The pdf estimate using a Gaussian kernel is given by:

$$\hat{f}(x) = \frac{1}{N} \frac{1}{h\sqrt{2\pi}} \sum_{i=1}^{N} \exp\left(-\frac{(x - x_i)^2}{2h^2}\right).$$

This estimation method is quite costly in computational time. Kernel estimators are considered to be very good for density estimation of one-dimensional data however it is not always the case for mutual information estimation.

4.4 *k*-Nearest Neighbor Estimator

Kraskov et al. [14] present a new estimator based on distances of k-Nearest Neighbors (KNN) to estimate densities. The authors consider a bivariate sample and, for each reference point, a distance length is computed so that k neighbors are within this distance length noted $\epsilon(i)$ for a reference point i. The number of points with distance $\epsilon(i)/2$ gives the estimate of the joint density at the point i. The distance is then projected into each variable subspace to estimate the marginal density of each variable. The estimation of MI using KNN depends on

the choice of k. In [14] the authors explain that statistical errors increase when k decreases. In practice, we should use $k > 1$, however if k is too large, systematic errors can outweigh the decrease of statistical errors. KNN gives good results with less statistical errors than previous methods but with a computationally heavy algorithm [21].

4.5 B-Spline Estimator

In [6] Daub et al. introduce the use of B-spline functions as entropy estimators. A B-spline curve is a generalized Bézier curve. It is specified by the parameters:

- the degree d, or order $k = d + 1$, so that each segment of the piecewise polynomial curve has degree d or less,
- a sequence of $m + 1$ numbers, $t_0, \ldots, t_m$, called knot vector, such that $t_i \leq t_{i+1}, \forall i \in \{1, \ldots, m-1\}$,
- control points, $b_0, \ldots, b_n$.

A B-spline curve is defined in terms of B-spline basis functions. The i-th basis function of degree d, noted $B_{i,d}$, defined by the knot vector $t_0, \ldots, t_m$ is defined by the Cox-de Boor recursion formula as follows:

$$B_{i,0}(z) = \begin{cases} 1 & \text{if } t_i \leq z < t_{i+1} \\ 0 & \text{otherwise.} \end{cases}$$

$$B_{i,d}(z) = \frac{z - t_i}{t_{i+d} - t_i} B_{i,d-1}(z) + \frac{t_{i+d+1} - z}{t_{i+d+1} - t_{i+1}} B_{i+1,d-1}(z),$$

for $i = 0, \ldots, n$ and $d \geq 1$. Finally, the property:

$$\sum_{i=0}^{n} B_{i,d}(z) = 1,$$

for any value of z, makes B-spline basis functions suitable as a pdf estimator. This estimator is noted BSE. More details on the use and advantages of BSE in the side-channel context are available in [28].

5 Experimental Analysis

We analyze in this section the practical efficiency of nonparametric estimators of mutual information in the context of side-channel attacks. We compare their performances with state-of-the-art proposed side-channel distinguishers:

- classical parametric test, CPA (Sec. 3.3),
- nonparametric tests, SPE (Sec. 3.5), CVM (Sec. 3.5),
- cluster analysis, DCA with variance as criterion function (Sec. 3.4),
- mutual information with parametric estimation, Cumulant-based Estimator (CE) [16] which is the most powerful parametric estimator,

– mutual information with nonparametric estimation, GMIA (Sec. 2.2), HE (Sec. 4.2), KDE (Sec. 4.3), KNN (Sec. 4.4), BSE (Sec. 4.5).

In order to compare the efficiency of side-channel attacks, we use common metrics proposed in the literature [27]. Guessed Entropy (GE) is the average position of the correct key hypothesis in the sorted vector of hypothesis at the end of the attack. Results using another metric are presented in Appendix A.

Attacks are performed on two different setups: the publicly available power curves of DPA Contest 2008/2009 [30] of a DES implementation and curves acquired on an Atmel STK600 board with an Atmel AVR ATmega2561 [1] of a multi-precision multiplication algorithm.

On DPA Contest 2008/2009 curves of a DES, the intermediate value targeted is the output the SBox in the last round. For the attacks using mutual information with nonparametric estimation, we consider no power model, i.e. the value of the data. The Hamming weight model is used for the other attacks. Each attack is performed on 135 sets of 600 power curves in order to average the results. We evaluate each distinguishers and present the results in Fig. 1. Attacks can be assigned to different groups depending on their efficiency. The best seems to be CE, CPA and SPE which are all parametric tests. The following attacks are KDE, DCA and CVM amongst which are the first mutual information nonparametric ones. The BSE is next, followed by KNN, GMIA and HE.

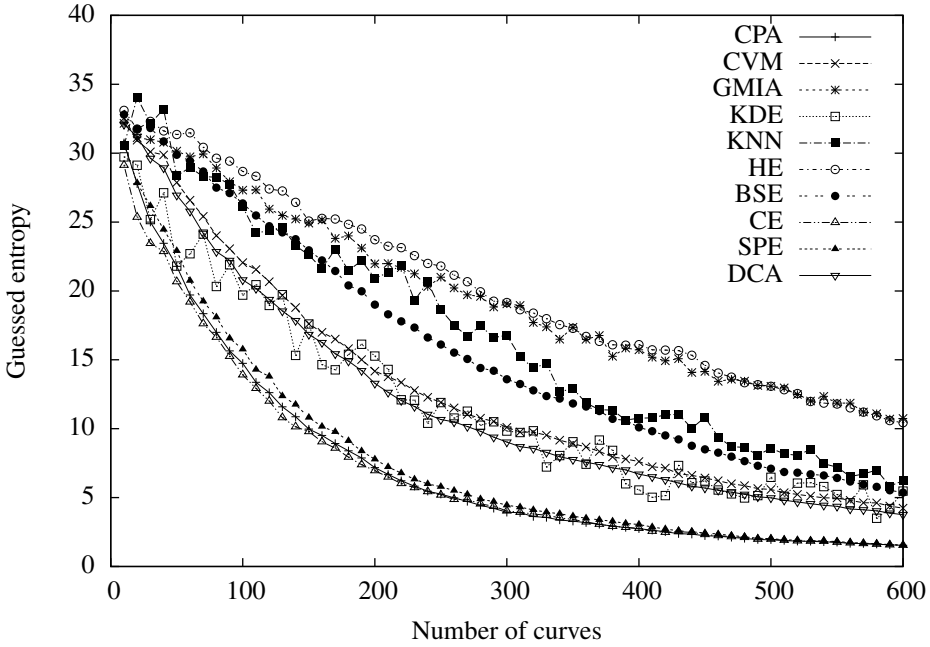

Fig. 1: Guessed entropy results on DPA Contest 2008/2009 curves of a DES.

The same attacks are also performed on curves acquired on a STK600 development board with a 8-bit Atmel AVR ATmega2561. This setup is not particularly well suited to perform side-channel attacks. Therefore the power traces contains significantly more noise than the DPA Contest ones. We attack a column-wise multi-precision multiplication algorithm implemented in software. The targeted values are the intermediate 8-bit multiplications $x_i \times y_j$ considering one of the multiplicand known by the attacker. As with the previous setup, we do not consider a power model for the attacks using mutual information with nonparametric estimation. The other attacks assume the Hamming weight model. Each attack is performed on 20 sets of 2000 power curves. We obtain a slightly different performance from several attacks (Fig. 2). As previously, the most powerful attacks are still CE, CPA and SPE. However BSE seems to perform much better and is at the same level as CVM, DCA and KDE. The estimators BSE and KDE are the most efficient nonparametric methods but BSE is more computationally efficient than KDE, hence an more interesting choice.

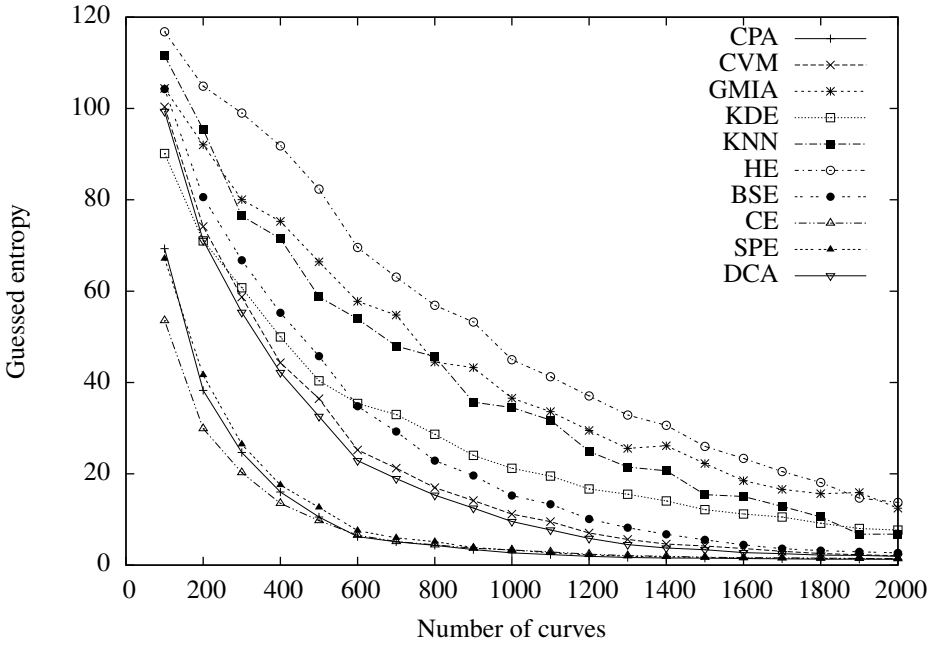

Fig. 2: Guessed entropy results on STK600 curves of a multi-precision multiplication.

With this overall comparison of state-of-the-art side-channel distinguishers, we can note differences between classical statistical tests performance and their efficiency in the side-channel context. For example, the KNN estimator should be less subject to statistical errors than BSE or KDE. However it performs

worse in this scenario. Classical parametric tests are still amongst the most powerful in most cases. In particular the recently presented Cumulant-based Estimator [16], a parametric estimator of mutual information, is very interesting. These experimental analysis also show the gain obtained when using efficient nonparametric estimators of mutual information. Even if the MIA attack is not the most powerful, its performance is greatly improved compared to the classical histogram estimator that has been used in the literature as a reference.

6 Conclusion

In this paper, we review some of the most statistically powerful nonparametric estimators of mutual information in the context of side-channel analysis. The distinction between parametric and nonparametric methods is important and should be clearly made when comparing side-channel distinguishers efficiency. Depending on the supposed knowledge of the adversary, one of these two classes of attacks needs to be considered. We also note that, in terms of performance, nonparametric estimation in MIA is not as bad as previously thought. The KDE and BSE estimators perform quite well for an acceptable computational overhead in the case of BSE. Even if the study is done on CMOS devices, we can expect a similar improvement of performance on different types of logic when using efficient nonparametric methods.

References

1. ATMEL: ATmega 2561 Data Sheet, `http://www.atmel.com/dyn/resources/prod_documents/doc2549.pdf`
2. Batina, L., Gierlichs, B., Lemke-Rust, K.: Comparative Evaluation of Rank Correlation Based DPA on an AES Prototype Chip. ISC 2008, LNCS 5222, 341–354 (2008)
3. Batina, L., Gierlichs, B., Lemke-Rust, K.: Differential Cluster Analysis. CHES 2009, LNCS 5747, 112–127 (2009)
4. Brier, E., Clavier, C., Olivier, F.: Correlation Power Analysis with a Leakage Model. CHES 2004, LNCS 3156, 135–152 (2004)
5. Chen, Z., Zhou, Y.: Dual-Rail Random Switching Logic: A Countermeasure to Reduce Side Channel Leakage. CHES 2006, LNCS 4249, 242–254 (2006)
6. Daub, C., Steuer, R., Selbig, J., Kloska, S.: Estimating Mutual Information Using B-spline Functions - an Improved Similarity Measure for Analysing Gene Expression Data. BMC Bioinformatics 5, 118 (2004)
7. Flament, F., Guilley, S., Danger, J., Elaabid, M., Maghrebi, H., Sauvage, L.: About Probability Density Function Estimation for Side Channel Analysis. In: COSADE 2010 (2010)
8. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual Information Analysis - A Generic Side-Channel Distinguisher. CHES 2008, LNCS 5154, 426–442 (2008)
9. Goodman, L., Kruskal, W.: Measures of Association for Cross Classifications. II: Further Discussion and References. Journal of the American Statistical Association 49, 732–764 (1954)

10. Guilley, S., Hoogvorst, P., Mathieu, Y., Pacalet, R.: The Backend Duplication Method. CHES 2005, LNCS 3659, 383–397 (2005)
11. Kendall, M.: A New Measure of Rank Correlation. Biometrika 30, 1–2 (1938)
12. Kocher, P.: Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. CRYPTO 1996, LNCS 1109, 104–113 (1996)
13. Kocher, P., Jaffe, J., Jun, B.: Differential Power Analysis. CRYPTO 1999, LNCS 1666, 388–397 (1999)
14. Kraskov, A., Stogbauer, H., Grassberger, P.: Estimating Mutual Information. Physical Review E 69, 66138 (2004)
15. Kullback, S., Leibler, R.: On Information and Sufficiency. The Annals of Matematical Statistics 22, 79–86 (1951)
16. Lee, T.H., Berthier, M.: Mutual Information Analysis under the View of Higher-Order Statistics. In: To appear in IWSEC 2010 (2010)
17. Messerges, T.S., Dabbish, E.A., Sloan, R.H.: Investigations of Power Analysis Attacks on Smartcards. In: USENIX Workshop on Smartcard Technology. pp. 151–162 (1999)
18. Messerges, T.S., Dabbish, E.A., Sloan, R.H.: Power Analysis Attacks of Modular Exponentiation in Smartcard. CHES 1999, LNCS 1717, 144–157 (1999)
19. Moon, Y.I., Rajagopalan, B., Lall, U.: Estimation of Mutual Information using Kernel Density Estimators. Physical Review E 52(3), 2318–2321 (1995)
20. Moradi, A., Mousavi, N., Paar, C., Salmasizadeh, M.: A Comparative Study of Mutual Information Analysis under a Gaussian Assumption. Information Security Applications, LNCS 5932, 193–205 (2009)
21. Papana, A., Kugiumtzis, D.: Evaluation of Mutual Information Estimators on Nonlinear Dynamic Systems. Nonlinear Phenomena in Complex Systems 11, 225–232 (2008)
22. Pompe, B., Heilfort, M.: On the Concept of the Generalized Mutual Information Function and Efficient Algorithms for Calculing it (1995)
23. Pompe, B., Physik, F.: Measuring Statistical Dependences in a Time Series. Journal of Statistical Physics 73, 587–610 (1993)
24. Prouff, E., Rivain, M.: Theorical and Practical Aspects of Mutual Information Based Side Channel Analysis. ACNS 2009, LNCS 5536, 499–518 (2009)
25. Rényi, A.: On Measures of Information and Entropy. In: Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability. vol. 1, pp. 547–561 (1961)
26. Shannon, C.: A Mathematical Theory of Communication. The Bell System Technical Journal 27, 379–423 (1948)
27. Standaert, F.X., Gierlichs, B., Verbauwhede, I.: Partition vs . Comparison Side-Channel Distinguishers: An Empirical Evaluation of Statistical Tests for Univariate Side-Channel Attacks against Two Unprotected CMOS Devices. ICISC 2008, LNCS 5461, 253–267 (2008)
28. Venelli, A.: Efficient Entropy Estimation for Mutual Information Analysis using B-splines. WISTP 2010, LNCS 6033, 17–30 (2010)
29. Veyrat-Charvillon, N., Standaert, F.: Mutual Information Analysis: How, When and Why? CHES 2009, LNCS 5747, 429–443 (2009)
30. VLSI research group and TELECOM ParisTech: The DPA Contest 2008/2009, `http://www.dpacontest.org`

A First-order Success Rate Results

We present here the results of the success rate metric on the two platforms detailed in Sec. 5. First-order success rate is, for a given number of curves, the probability that the correct key hypothesis is ranked first in the sorted vector of hypothesis. These results are consistent with the guessed entropy metric presented in Sec. 5.

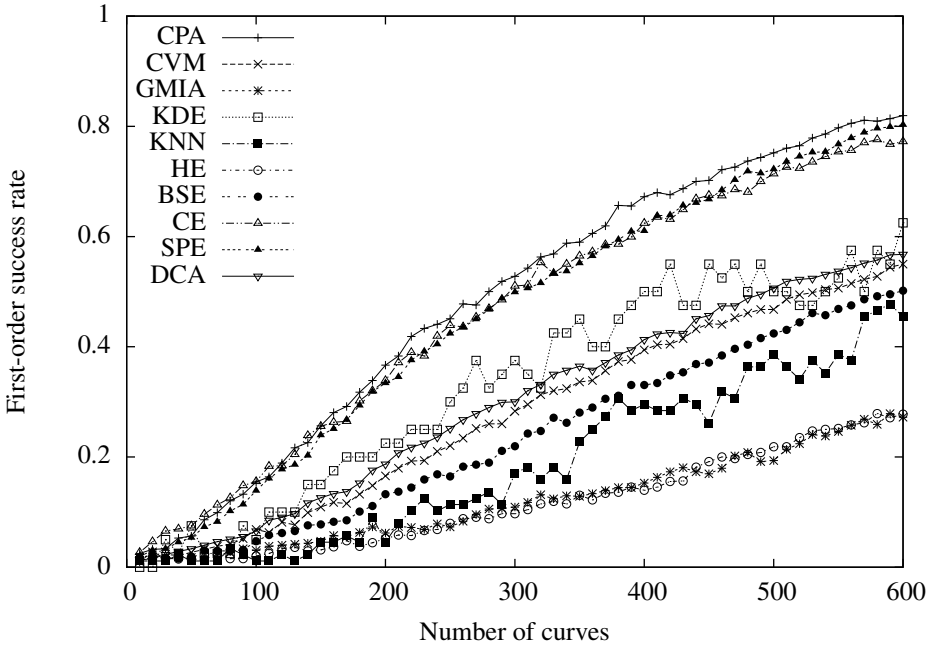

Fig. 3: First-order success rate on DPA Contest 2008/2009 curves of a DES.

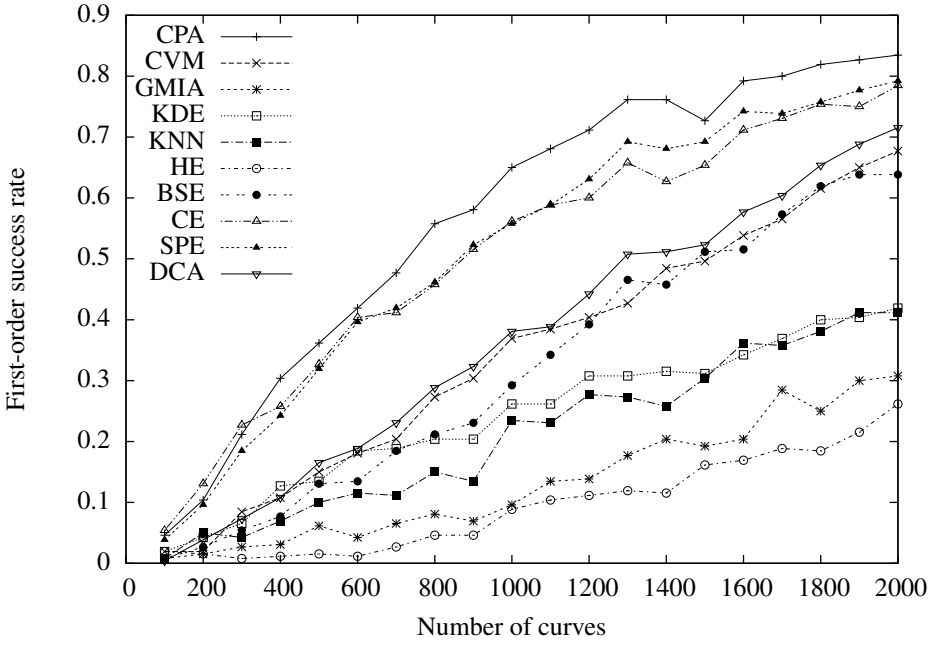

Fig. 4: First-order success rate results on STK600 curves of a multi-precision multiplication.