An Exploration of the Kolmogorov-Smirnov Test as Competitor to Mutual Information Analysis

Carolyn Whitnall, Elisabeth Oswald, and Luke Mather

University of Bristol, Department of Computer Science,
Merchant Venturers Building, Woodland Road, BS8 1UB, Bristol, UK

Abstract. A theme of recent side-channel research has been the quest for distinguishers which remain effective even when few assumptions can be made about the underlying distribution of the measured leakage traces. The Kolmogorov-Smirnov (KS) test is a well known non-parametric method for distinguishing between distributions, and, as such, a perfect candidate and an interesting competitor to the (already much discussed) mutual information (MI) based attacks. However, the side-channel distinguisher based on the KS test statistic has received only cursory evaluation so far, which is the gap we narrow here. This contribution explores the effectiveness and efficiency of Kolmogorov-Smirnov analysis (KSA), and compares it with mutual information analysis (MIA) in a number of relevant scenarios ranging from optimistic first-order DPA to multivariate settings. We show that KSA shares certain 'generic' capabilities in common with MIA whilst being more robust to noise than MIA in univariate settings. This has the practical implication that designers should consider results of KSA to determine the resilience of their designs against univariate power analysis attacks.

1 Introduction

Differential power analysis (DPA) is a form of side-channel analysis which employs some type of statistic (the *distinguisher*) to identify a correct hypothesis about (part of) the secret key from within a set of possible alternative hypotheses. Popular distinguishers include the Pearson correlation coefficient, the distance-of-means test, and mutual information analysis (MIA). Mutual information (MI) measures the total dependency between two random variables, and was first proposed for use as a distinguisher at CHES 2008 ([5]). MIA's selling point is *genericity*: it is capable of key recovery even when the underlying leakages satisfy few assumptions.

Previous work such as [17] and [14] demonstrated that the (notoriously problematic) estimation of the leakage probability density functions for different key-dependent models is of decisive importance to the performance of MIA in practice. The authors of [17] suggested two alternative distinguishers based on statistics which are conceptually similar to MI but do not require explicit density estimation: the (two-sample) Kolmogorov-Smirnov (KS) test and the Cramér-von-Mises criterion. Each essentially computes some notion of a 'distance' between two distributions. Evaluations of these (and other similar) methods can be found in the statistical literature (for example, [16]): whilst the Cramér-von-Mises statistic performs particularly well (i.e. better than KS) for certain specific distributions, the KS statistic is found to perform well across the board and therefore represents the most generic, distribution-free method.

In this submission we demonstrate how the KS test statistic adapts to the purposes of DPA and investigate the properties and practical performance of such attacks. Alongside, we present an equivalent analysis of MIA—an ideal comparator because of its established role in the existing literature as well as its conceptual similarity to Kolmogorov-Smirnov Analysis (KSA). We assess the distinguishers as applied to key-recovery attacks against implementations of DES in four practically relevant leakage scenarios. Our results are interesting for academics and practitioners alike: from an academic point of view it is interesting to investigate how a conceptually similar approach such as the KS test performs in comparison to MIA. From a practical point of view we are providing information about how to choose the most appropriate distinguisher in certain settings. Specifically, in the setting where the actual power model of a device is unknown to the attacker and does not correspond to a 'nice' Hamming weight leakage, and where a substantial amount of noise distorts the data-dependent signal, we show that KSA actually outperforms MIA and hence is the best choice of a distinguisher (in this setting) at present. This setting is practically relevant as it resembles what can be expected when attacking devices that implement cryptography in hardware and have measures in place to increase the level of noise.

This submission starts by introducing differential power analysis (DPA) attacks in Sect. 2. To explain our comparison criterion we outline some key concepts related to the outcomes of DPA attacks (i.e. the distinguishing vectors) in Sect. 3. We then explain how the KS test adapts to DPA attacks (including considerations for higher-order attacks) in Sect. 4. Section 5 consists of our results. We conclude thereafter in Sect. 6.

1.1 Our Contributions

In Sect. 3 we adapt the ideas presented in [18] to our purposes and introduce the measure of *nearest-rival distinguishability* to compare distinguishers. We argue that this measure is relevant for practical considerations as it strongly influences the number of traces required for successful key recovery: the smaller the nearest-rival distinguishability score, the more traces will be necessary before the correct key stands out from the alternative hypotheses when the vector comes to be estimated in practice.

In Sect. 4 we show how the KS test statistic can be used to construct a distinguisher for power analysis attacks. We briefly include relevant results of the statistical literature and show how to apply them in the context of univariate and multivariate attacks. An interesting conclusion that we can draw is that whilst KSA shares many properties with MIA in the univariate setting, its extension to general multivariate settings is problematic (([12], ([4])).

In Sect. 5 we analyse the application of the KS distinguisher to four relevant scenarios. An important phenomenon that we observe is that KSA is consistently more robust to noise. Our results give conclusive evidence that it outperforms MIA in univariate scenarios (our study ranges from the optimistic Hamming weight assumption to realistic leakages including the assumption of an unknown highly-nonlinear function). Interesting observations result from our study of bivariate extensions of KSA: here it clearly underperforms MIA both in the masked and unmasked case, irrespective of noise. Our contribution thus gives a balanced view of KSA; it shows both its strengths and weaknesses.

2 Differential Power Analysis

In this article we consider a 'standard DPA attack' scenario as defined in [9]. We assume that the power consumption T of the target cryptographic device depends on some internal value (or state) $f_{k^*}(x)$. The state is a function of some part of the plaintext $x \in \mathcal{X}$, as well as some part of the secret key $k^* \in \mathcal{K}$. Consequently, we have that $T = L \circ f_{k^*}(X) + \varepsilon$, where X is a random variable taking values in $\mathcal{X}$, L is some function which describes the data-dependent component and ε comprises the remaining power consumption which can be modeled as independent random noise. The attacker has N power measurements corresponding to encryptions of N known plaintexts $x_i \in \mathcal{X}$, $i = 1, \ldots, N$ and wishes to recover the secret key k^*. The attacker can accurately compute the internal values as they would be under each key hypothesis $\{f_k(x_i)\}_{i=1}^N$, $k \in \mathcal{K}$ and uses whatever information he possesses about the true leakage function L to construct a prediction model $M : f(\mathcal{X}) \longrightarrow \mathcal{M}$.

DPA is based on the intuition that the modeled power traces corresponding to the correct key hypothesis should bear more resemblance to the true power traces than the modeled traces corresponding to incorrect key hypotheses. An attacker is thus concerned with comparing the degree of similarity between the true and modeled traces. A range of comparison tools—'distinguishers'—can be used, of which mutual information (MI) is an example. MI measures, in bits, the total information shared between two random variables, and is most intuitively expressed in terms of entropies via Shannon's formula: $\mathrm{I}(A;B) = \mathrm{H}(A) - \mathrm{H}(A|B)$.

It is employed as an attack distinguisher to compare the measured traces T with the hypothesis-dependent predictions $M_k = M \circ f_k(X)$:

$$D_{\mathrm{MI}}(k) = \mathrm{I}(T; M_k) = \mathrm{H}(T) - \mathrm{H}(T|M_k) = \mathrm{H}(T) - \underset{m \in \mathcal{M}}{\mathbb{E}} \left[\mathrm{H}(T|M_k = m) \right], \qquad (1)$$

and because the 'unexplained' entropy (the second term) is smallest when the predictions are good, we expect (1) to be maximised for the correct key hypothesis $k = k^*$.

MI is particularly appealing for use in DPA because it compares distributions in a general way, detecting not just linear relationships but nonlinear relationships too. Thus MIA has been promoted as a 'generic' distinguisher which potentially remains effective even in the absence of a good power model. It also has natural multivariate extensions, by which it can be straightforwardly adapted to higher-order attacks (see [2] for an overview). However, estimation of MI is notoriously problematic ([11]); all known estimators are biased and no 'ideal' estimator exists (different estimators perform differently depending on the underlying structure of the data). Consequently, MIA outcomes are highly sensitive to the estimation procedure and parameters chosen by the attacker.

3 Evaluation Methodology

The aim of our paper is to compare KSA with MIA in practically relevant scenarios. It is imperative to understand that we are seeking to compare statistical procedures and not attacks or devices: our methodology hence operates on hypothetical leakages simulated on the basis of clearly defined, practically relevant characteristics such as cryptographic function (we use as examples a non-linear substitution box taken from the DES standard, as well as the Boolean exclusive-or), device leakage model (we use as examples Hamming weight, the sum of unevenly weighted bits, and a highly-nonlinear function), and noise (we use Gaussian noise). Our results will be relevant for all devices which share above mentioned characteristics.

Our approach is based on the recent work published in [18] which proposes to study 'complicated' distinguishers such as MIA by computing and estimating (respectively) so-called theoretic and practical distinguishing vectors. The motivation for this is that distinguishers like MIA do not conform to the easily understood behaviours of 'simple' distinguishers such as correlation, which has a known sampling distribution and responds to noise in a well-understood fashion ((e.g. see Chapters 4 and 6 in [8]). We have mentioned before that estimation is notoriously difficult [11]. Studying only practical distinguishing vectors does not in many cases (as illustrated by previous work such as [2]) allow us to draw any definite conclusions about MIA because it is unclear from the practical vectors whether it is a lack of good estimators or an inherent weakness of MIA that causes its sometimes disappointing performance in practice. By contrast, by studying both theoretic and practical vectors we can assess whether MI itself is the problem or simply the estimation process.

Another contribution of [18] is that of defining measures for distinguishability. This is motivated by the fact that the larger the theoretic (true) margins by which the correct key is distinguished, the fewer traces we expect to require to detect this difference in practice ([7]). We use the following subsections to further elaborate on the key concepts relevant to our study (theoretic and practical distinguishing vectors, distinguishability).

3.1 Theoretic vs. Practical Distinguishing Vectors

We adopt the notation of [18], which defines the theoretic attack distinguisher as $\mathbf{D} = \{D(k)\}_{k \in \mathcal{K}} = \{D(L \circ f_{k^*}(X) + \varepsilon, M \circ f_k(X))\}_{k \in \mathcal{K}}$, where the plaintext input X takes values in $\mathcal{X}$ according to some known distribution (usually uniform). The distinguisher D is chosen as some function, e.g. MI. For a defined leakage function L and a power model M, the value $D(k)$ can be precisely calculated. It thus represents the 'true' value of the distinguisher given M, L, and key hypothesis k.

How to compute the 'true' distinguisher values. For each possible input $x \in \mathcal{X}$ to the cryptographic function we obtain a vector evaluating the (variance $\mathrm{Var}(\varepsilon)$) Gaussian density centred at the corresponding data-dependent leakage value $L \circ f_{k^*}(x)$. The average of these vectors, weighted by the input probabilities $\mathbb{P}(X = x)$, then gives the probability density of the power consumption evaluated over the full range of possible leakage values. Conditional densities, corresponding to each possible prediction value $m \in \mathcal{M}$ under each key hypothesis $k \in \mathcal{K}$, are constructed similarly. From these probability densities we are able to directly compute (via numerical integration) MIA distinguishing vectors as per equation (1). The same approach allows us to compute KSA distinguishing vectors (to be defined in Sect. 4, equation (2)).

In practice $\mathbf{D}$ must be estimated as the true distribution of T is unknown (in the unprofiled setting which we are examining). Suppose we have observations corresponding to the vector of inputs $\mathbf{x} = \{x_i\}_{i=1}^N$, and write $\mathbf{e} = \{e_i\}_{i=1}^N$ to be the observed noise (i.e. drawn from the distribution of ε). Then the estimated vector is $\hat{\mathbf{D}}_N = \{\hat{D}_N(k)\}_{k\in\mathcal{K}} = \{\hat{D}_N(L \circ f_{k^*}(\mathbf{x}) + \mathbf{e}, M \circ f_k(\mathbf{x}))\}_{k\in\mathcal{K}}$.

The theoretic distinguishing vector $\mathbf{D}$ can thus be seen as representing the 'best' result one could hope to achieve when performing an analysis in practice.

3.2 Notion of Distinguishability

It follows clearly from the working principle of the distinguishers (as explained in previous sections) that the results of each will be on very different scales: MI is measured in bits and takes values between zero and the total entropy of the measured traces, whereas the KS statistic measures the (absolute) difference between probability distributions and therefore takes values in $[0, 1]$. In order to make meaningful comparisons we need to define an outcome measure which is independent of the numerical results of distinguishers. One approach is to look at how well the correct key hypothesis 'stands out'. Previous work has introduced measures for 'standing out', e.g. a "DPA signal-to-noise ratio" was defined in [6]. We seek to represent, more directly than the "DPA signal-to-noise ratio", the margin to be detected by a practical attack. Thus we look at the distance of the correct key hypothesis from its nearest rival, and to scale this by an appropriate normalising constant. Consequently, we define the *nearest-rival distinguishability* score as the difference between the true-key distinguisher value and the highest incorrect-key value, divided by the standard deviation of the 'optimal' distinguishing vector: the theoretic output of an attack in a noise-free setting with a known power model.

$$\textit{Nearest-rival distinguishability}(\mathbf{D}) = \frac{D(k^*) - \max\{D(k)|k \neq k^*\}}{\sqrt{\mathrm{Var}\{D(L \circ f_{k^*}(X), L \circ f_k(X))\}_{k\in\mathcal{K}}}}.$$

We stress again that this measure of theoretic distinguishability is a meaningful indicator of the practical efficiency of an attack as statistical theory (for example, [7]) teaches us that the sample size required to detect a difference is strongly related to the true size of that difference: the lower the score, the more traces we expect to require for a successful attack in practice.

4 The Kolmogorov-Smirnov Distinguisher

The Kolmogorov-Smirnov (KS) test has been mentioned in [17] as a seemingly attractive alternative to MIA: it is similarly able to generically compare the distributions of two samples but achieves this without explicit estimation of their probability density functions (PDFs). It also extends fairly straightforwardly to bivariate distributions which makes it adaptable to second-order DPA attacks, although (unlike MI) it becomes problematic in higher dimensions ([4]).

In this paper we are particularly interested in how KSA compares with MIA, in 'typical' scenarios and in some of the more specific scenarios for which MIA has been promoted, namely unknown power model and higher-order attacks. The remainder of this section introduces the KS test and discusses its application to univariate and bivariate (second-order) DPA attacks.

4.1 Kolmogorov-Smirnov Based DPA Attacks

The (two-sample) KS test statistic measures the distance between the empirical cumulative distribution functions (CDFs) of two samples $\mathbf{A} = \{A_i\}_{i=1}^n$ and $\mathbf{B} = \{B_j\}_{j=1}^m$, in order to test whether they have been drawn from the same distribution. It is defined as $\sup_{x\in\mathbf{A}\cup\mathbf{B}} |F_A(x) - F_B(x)|$ where F_A, F_B are the empirical CDFs, i.e. $F_A(x) = \frac{1}{n}\sum_{i=1}^n I_{\{A_i\leq x\}}$ ($I_{\{A_i\leq x\}}$ is the indicator function, taking the value 1 if $A_i \leq x$ and 0 otherwise).

Just as MIA can be understood to operate by comparing the global traces T with the hypothesis-dependent conditional traces $T|M_k$—via the expected change in entropy—a KS-inspired distinguisher measures the maximum distance between the global and the conditional trace distributions, as averaged over the prediction space:

$$D_{\mathrm{KS}}(k) = \mathbb{E}[K(T||T|M_k)] = \mathop{\mathbb{E}}_{m\in\mathcal{M}}\left[\sup_t |F_T(t) - F_{T|M_k=m}(t)|\right]. \tag{2}$$

In case of the correct key hypothesis we expect the test statistic to return a large difference.

The particular appeal of the KS statistic as an alternative to mutual information is that it does not require the explicit estimation of densities, but only the calculation of empirical cumulative distribution functions.

Example: We illustrate the working principle of the KS test via a very simple example consisting of a DES implementation leaking the Hamming weight (HW) of the first S-Box with a signal-to-noise ratio (SNR, defined as $\frac{\mathrm{Var}(L\circ f_{k^*}(M))}{\mathrm{Var}(\varepsilon)}$) of 8. For each key hypothesis we estimate the empirical CDFs of the traces as conditioned on the model predictions and compare them with the 'global' CDF of the traces by computing the expected largest difference between them according to (2).

The left panel of Fig. 1 shows (in red) the conditional CDFs under the correct key hypothesis, where the 'weight' of the lines indicates the relative contribution of the prediction-specific KS statistics to the expectation which comprises the KS distinguisher as in equation (2). The difference—and most pertinently the maximum (vertical) distance—between these conditional CDFs and the global CDF (in blue) is visibly substantial. By comparison the right panel shows the same conditional CDFs as induced by an incorrect key hypothesis. These more closely resemble the global CDF; it is clear to see that the expected maximum distance will be substantially smaller. The same behaviour can be observed for all other incorrect key hypotheses, hence providing the rationale for our KS-inspired distinguisher: we expect only the correct key hypothesis to produce a large average difference.

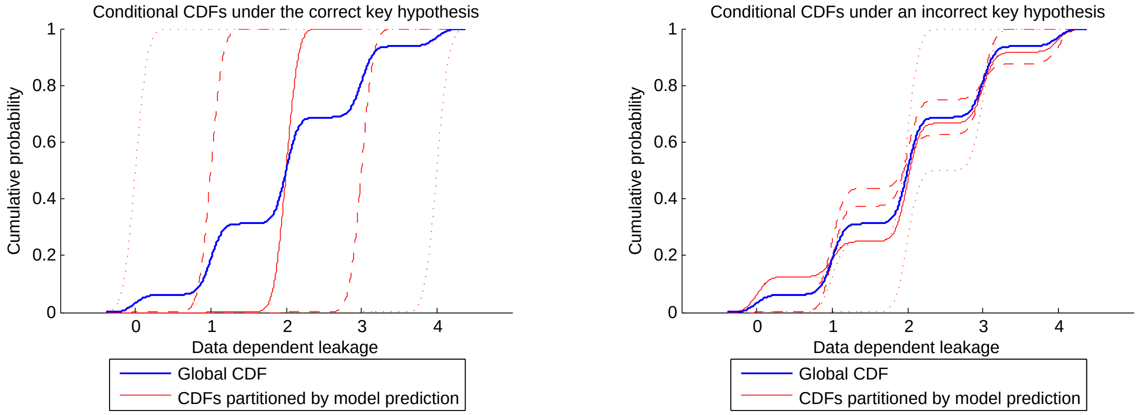

Fig. 1: *The KS test is based on the largest distance between the CDFs of two samples. The left and right panels show the CDFs as conditioned on the model predictions under the correct key hypothesis and an incorrect key hypothesis, respectively.*

Note that, by design, the test is very sensitive to *any* distributional difference; this is one of the features which makes it popular as a general, non-parametric method of comparison. But for the purposes of DPA there is a potential downside to this sensitivity: the statistic will detect even the subtle differences induced by the incorrect hypotheses, to the detriment of the margin by which the correct key is distinguished.

4.2 Multivariate Extensions

Standard first-order DPA attacks apply a distinguisher to a single point in a trace. It is appealing to suppose that including more than one data point might be beneficial. In the case of attacks against unprotected implementations this could produce better results as more data points potentially imply that more information can be exploited (this has been argued specifically for template attacks [3]). In the case of masked implementations it could provide a way to defeat the masking scheme as the joint distributions of two or more trace points might be related to unmasked model values.

Peacock ([12]) introduces a bivariate KS test statistic for comparing two-dimensional samples $(\mathbf{A_1}, \mathbf{A_2}) = \{(A_{1,i}, A_{2,i})\}_{i=1}^{n}$ and $(\mathbf{B_1}, \mathbf{B_2}) = \{(B_{1,j}, B_{2,j})\}_{j=1}^{m}$, which he defines as:

$$\sup_{(x,y)\in(\mathbf{A_1}\cup\mathbf{B_1})\times(\mathbf{A_2}\cup\mathbf{B_2})} |F_{A_1,A_2}(x,y) - F_{B_1,B_2}(x,y)|.$$

However, this extension is more problematic than the univariate case as it requires a meaningful construction of bivariate empirical CDFs.

The distribution-free property of the KS test rests on being able to map any distribution function on to any other distribution function using a transformation that preserves the ordering of the data. In the one-dimensional case this is trivially fulfilled: there are only two ways of ordering data, namely $\mathbb{P}(A \geq x)$ and $\mathbb{P}(A \leq x)$. As we have that $\mathbb{P}(A \geq x) = 1 - \mathbb{P}(A \leq x)$ the choice is in fact arbitrary.

In higher dimensions the empirical CDF can be defined as:

$$F_{A_1,A_2}(x,y) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} I_{A_{1,i}\leq x, A_{2,j}\leq y}$$

for all pairs (x, y). However, in the general case the choice of ordering now *does* affect the test statistic: there is no direct way to map (e.g.) between $\mathbb{P}(A_1 \leq x, A_2 \leq y)$ and $\mathbb{P}(A_1 \geq x, A_2 \leq y)$. In fact for d different random variables, there are 2^d possible orderings we need to consider. The simplest solution to this problem, as suggested by [12], is to find the maximum distributional difference arising from all 2^d possible orderings. The computational complexity of this approach is exponential in the number of variables ($O(2^d * n^d)$). Peacock shows in his work that a bivariate KS test statistic according to his suggestion is close enough to being distribution-free to be useful in practice.

Fasano and Franceschini [4] propose an optimisation whereby the test statistic is evaluated only at the points which are observed in the sample, i.e. at every $(x, y) \in (\mathbf{A_1}, \mathbf{A_2}) \cup (\mathbf{B_1}, \mathbf{B_2})$ rather than every $(x, y) \in (\mathbf{A_1} \cup \mathbf{B_1}) \times (\mathbf{A_2} \cup \mathbf{B_2})$. They are able to show that this leads to a linear increase in speed without compromising on the power of the test or the distribution-free property.

We next explain how this bivariate extension of the test statistic can be adapted to DPA attacks in which two trace points are exploited, and present analogous distinguishers based on multivariate extensions to mutual information. Note that, whilst the latter has natural extensions to dimensions greater than 2, the KS statistic is shown to be problematic in higher dimensions. The authors of ([4]) *do* present a three-dimensional test but this is not achieved without some difficulty and a substantial increase in complexity (now 2^3 orderings need to be considered); as such we choose not to make use of it ourselves.

Extensions for Masked Implementations In a second-order attack against a masked implementation we make univariate leakage predictions based on the (unmasked) target value and then exploit what this 'tells' us about the joint distribution of the mask and the target value combined. For the KS distinguisher this means that we are comparing the global joint CDF of the traces with the conditional joint CDFs as partitioned by the model predictions under each key hypothesis:

$$D_{\mathrm{2OKS}}(k) = \mathbb{E}[K(T_1,T_2||T_1,T_2|M_k)] = \mathop{\mathbb{E}}_{m\in\mathcal{M}}\left[\sup_{t_1,t_2}\left\{|F_{T_1,T_2}(t_1,t_2) - F_{T_1,T_2|M_k=m}(t_1,t_2)|\right\}\right]. \tag{3}$$

Previous work (such as [2]) has explored the various ways in which mutual information generalises to higher orders and how these different notions can be adapted to the purposes of DPA. For the purposes of comparison we focus on the extension which is most analogous to the KS distinguisher, namely the information shared between the *pair* of trace points taken jointly and the model prediction, as follows:

$$D_{\mathrm{2OMI}}(k) = \mathrm{I}((T_1,T_2);M_k) = \mathrm{H}(T_1,T_2) - \mathrm{H}(T_1,T_2|M_k). \tag{4}$$

Extensions for Unprotected Implementations In an unprotected implementation we can use multivariate extensions of our distinguishers to exploit the joint leakage of two target values simultaneously, for example key addition and the output of the first DES S-Box.[1] This approach makes use of a bivariate model prediction and thus calls for slightly different constructions of the distinguishers to those employed in the context of masked implementations.

For the KS distinguisher we simply condition the joint CDFs by the bivariate prediction and proceed as before:

$$\begin{aligned} D_{\mathrm{MKS}}(k) &= \mathbb{E}[K(T_1,T_2||T_1,T_2|(M_1,M_2)_k)] \\ &= \mathop{\mathbb{E}}_{(m_1,m_2)\in\mathcal{M}_1\times\mathcal{M}_2}\left[\sup_{t_1,t_2}\left\{|F_{T_1,T_2}(t_1,t_2) - F_{T_1,T_2|(M_1,M_2)_k=(x_1,x_2)}(t_1,t_2)|\right\}\right]. \end{aligned} \tag{5}$$

Analogously we consider the MI between the pair of trace values and the pair of predictions:

$$D_{\mathrm{MMI}}(k) = \mathrm{I}((T_1,T_2);(M_1,M_2)_k) = \mathrm{H}(T_1,T_2) - \mathrm{H}(T_1,T_2|(M_1,M_2)_k). \tag{6}$$

5 Results

For each scenario that follows we first analyse theoretic KSA and MIA vectors for varying levels of Gaussian noise. These are derived from (respectively) true distributional differences and true entropies, computed directly from the trace density functions as explained in Sect. 3. We complement this theoretic analysis—which gives an indication of the underlying potential of a distinguisher—by estimating 'practical' attack vectors against simulated traces and reporting on trace requirements (again as noise varies).[2]

5.1 Optimistic Scenario: DES S-Box With (Known) Hamming Weight Leakage

We first consider the simple and often-studied scenario in which the power consumption comprises a data-dependent component proportional to the Hamming weight of the (first) DES S-Box plus some independent Gaussian noise. Assuming Hamming-weight leakage is realistic for implementations on simple micro-controllers (e.g. [8] use this as their running example).

[1] This choice is meaningful as the model predictions are in this case statistically independent.

[2] For MIA estimations we employ the heuristic rule favoured by the literature, and estimate PDFs via histograms with the number of bins equal to the cardinality of the power model image (i.e. 5 for the HW power model, 16 for the identity power model). Therefore, though these are not 'definitive' results (as no universally 'best' estimator exists) they do represent an established methodology and, as such, a meaningful basis for comparison with KSA.

Theoretic Outcomes

Pure-Signal Leakage: Figure 2 shows the theoretic distinguishing vectors for MIA and KSA attacks using a Hamming weight (HW) power model against noise-free Hamming weight leakage of the first DES S-Box. It also illustrates our notion of distinguishability. Both distinguishers are capable of identifying the correct key; MIA achieves a slightly higher distinguishability score of 5.6 compared with 4.2 for KSA. Equivalent attacks using the identity (ID) power model were less distinguishing, with scores of 3.8 and 3.1 for MIA and KSA respectively: evidently, the generic capabilities of the distinguishers are not useful in this 'known power model' scenario.

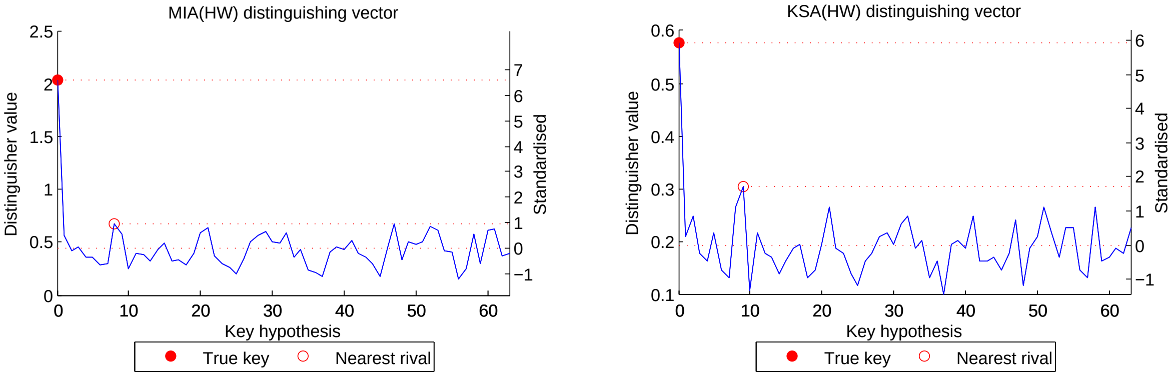

Fig. 2: *Theoretic distinguishing vectors for MIA(HW) and KSA(HW) in attacks against HW leakage of the first DES S-Box with zero noise.*

As SNR Varies: Figure 3 shows how the distinguishability scores vary with the strength of the data-dependent signal (relative to the Gaussian noise). The KSA attacks, though less distinguishing than their MIA counterparts in strong-signal scenarios, are more robust to noise and therefore attain a theoretic advantage in weak-signal scenarios.

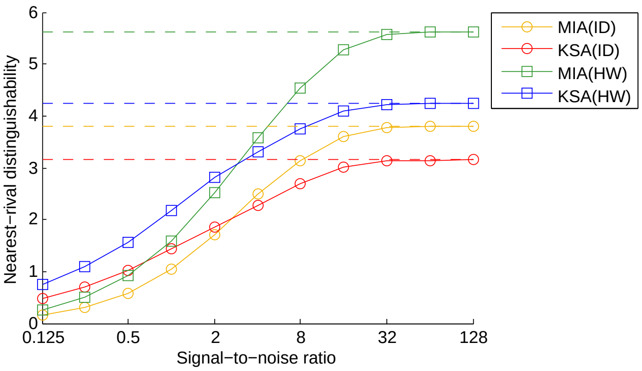

Fig. 3: *Theoretic distinguishing power as SNR varies, for attacks against the first DES S-Box with HW leakage.*

Practical Outcomes (Simulations) The first panel of Figure 4 shows the mean number of traces needed to recover the key; the second panel shows the 90^{th}-percentile, i.e. the number needed to achieve a 90% success rate. KSA(HW) performs almost identically to MIA(HW) (as could be expected from the theoretic vectors), with some evidence of a small advantage in weak-signal settings (again in keeping with the theoretic vectors). The ID attacks are more data intensive in both cases, but KSA(ID) exhibits consistently better performance than MIA(ID), probably due to the heavy estimation overhead incurred by the large number of bins required by the latter.

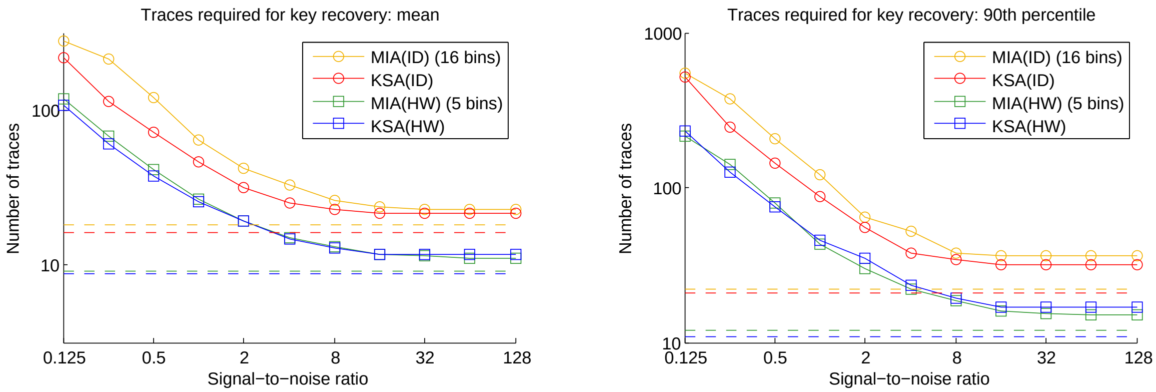

Fig. 4: *Mean and* 90^{th} *percentile of the trace requirement for key recovery, in repeated experiments against simulated HW leakage of the first DES S-Box, as SNR varies.*

5.2 Realistic Scenario: DES S-Box With Unknown Power Model

We next consider the performance of the two distinguishers in the case that the attacker does not have a precise power model. As motivated by [1] we focus on the case that the device leaks—instead of the Hamming weight—an unevenly weighted sum of the bits. This is realistic for typical micro-processors especially in the low-cost range (as reported by [1]). In our experiments, we assume that the least significant bit dominates in the leakage function with a relative weight of 10; in the experiments of [17] this was sufficient distortion to render MIA more effective than correlation DPA. To extend this analysis we also consider theoretic vectors assuming a highly non-linear power model[3]. This is relevant for hardware implementations, e.g. often non-linear functions are implemented via combinational logic in hardware and it is well known (see [10], [15]) that such implementations show leakage characteristics which are unrelated to linear leakage models.

Theoretic Outcomes

Pure-Signal Leakage: Both the HW and the generic ID variants of KSA are theoretically successful in a noise-free environment, but once again are slightly disadvantaged relative to MIA with distinguishing scores of 2.8 and 3.4 compared with 4.8 and 4.8 respectively.

As SNR Varies: The impact of noise is more marked than that observed for the known power model scenario, as can be seen in Figure 5; all attacks require a stronger signal before converging to their noise-free outcomes.

It is particularly notable that in high-noise settings the KSA attacks are actually more distinguishing than their MIA counterparts. Also of interest is the fact that the ID variants exhibit stronger outcomes and greater robustness to large amounts of noise than attacks using the (now imprecise) HW power model. Thus we confirm the existence of conditions under which KSA has the same ‘generic’ potential as MIA.

[3] To achieve a high-degree of non-linearity we use the Hamming weight of output of the AES SubBytes function.

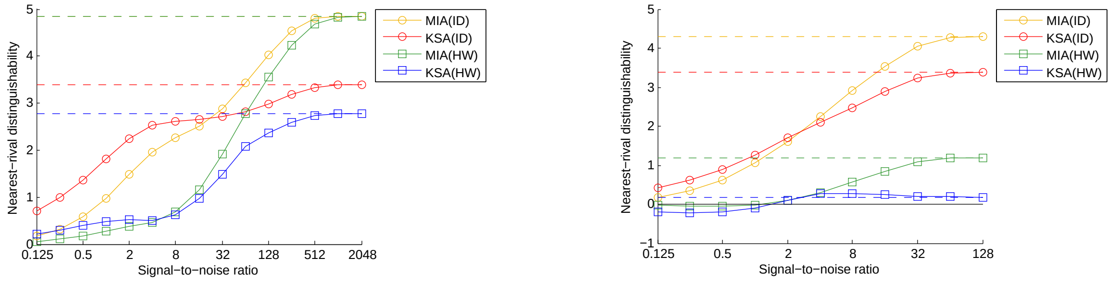

Fig. 5: *Theoretic distinguishing power as SNR varies for attacks against the first DES S-Box where the LSB dominates in the leakage with a relative weight of 10 (left panel) and were the leakage is a highly non-linear function (right panel)*

Practical Outcomes (Simulations) The theoretic KSA vectors show more distinguishing power than MIA so we have sufficient reason to expect that this translates to a practical advantage in terms of trace requirements, which we confirmed by deriving the practical distinguishing vectors via simulations.

Figure 6 plots the results (w.r.t. number of traces needed) of the practical distinguishing vectors as estimated from simulated traces with Gaussian noise. These tally well with the results of the theoretic vectors: ID attacks substantially outperform HW attacks when the leakage signal is weak, but this advantage is less clear in high-signal settings. KSA(ID) is particularly effective relative to MIA(ID) as estimated with 16 bins (we note that this does not necessarily represent the best-case capabilities of MIA but it is consistent with what one expects given the theoretic distinguishing vectors). KSA(HW) performs similarly to MIA(HW).

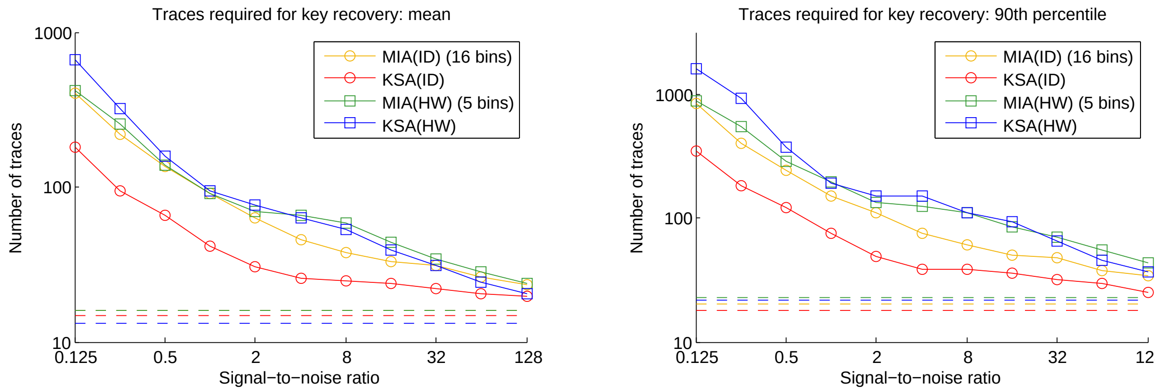

Fig. 6: *Mean and 90th percentile of the trace requirement for key recovery, in repeated experiments against simulated S-Box leakage in which the LSB dominates with a relative weight of 10.*

5.3 Higher-Order Scenario: Second-Order Attacks Against a Masked Implementation

As our first example of a multivariate application, we consider second-order attacks on a masked implementation of DES leaking the HW of the mask and the HW of the S-Box output, each with independent Gaussian noise. The second-order extensions for KSA and MIA distinguishers are as described in Sect.4.2.

Theoretic Outcomes

Pure-Signal Leakage: The noise-free distinguishing score of second-order KSA is just 0.6, compared with 3.2 for the MIA analogue. Thus both are capable of identifying the correct key, though with substantially reduced distinguishability relative to their first-order counterparts in unprotected scenarios, particularly in the case of KSA, as Fig. 7 illustrates.

As SNR Varies: Mark once more in Figure 7 that the KSA variant of the second-order attack exhibits greater noise robustness, so that in low-signal settings it shares comparable theoretic distinguishing power with MIA.

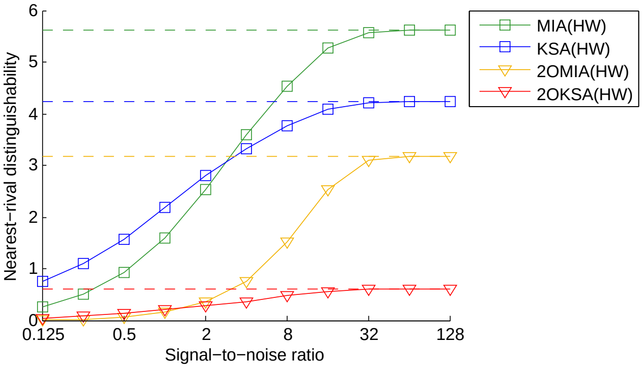

Fig. 7: *Theoretic distinguishing power as SNR varies, for second-order attacks against a masked implementation of DES with HW leakage.*

Practical Outcomes (Simulations) The first panel of Figure 8 shows the success rates for attacks against a masked DES implementation with noise-free leakage. The second-order KSA attack requires on average 150 traces, with a 90^{th}-percentile of 325, whilst second-order MIA is markedly more efficient, requiring on average only 30 traces with a 90^{th}-percentile of 45.

The remaining three panels show the same for scenarios in which small but increasing amounts of Gaussian noise are added. Even with an SNR as high as 128 the impact on success is substantial for both attack methods but (proportionately) more so for MIA. For an SNR of 32 (the lowest we attempted) the mean and 90^{th}-percentile of the trace requirement for KSA to be successful were 2,450 and 5,500 respectively; the equivalent figures for MIA were 1,440 and 3,200.

The heavy computational demands of the second-order KSA distinguisher mean that, as more noise is added, such attacks quickly become infeasible without enhanced computing power. Our theoretic analysis, and our practical results in other scenarios, indicate that it *could* achieve a small advantage over MIA (in terms of data complexity) when the signal is weak enough, but we are not able to test this and the advantage would likely be far outweighed by the relative computational costs.

5.4 Bivariate Extensions for an Unprotected Implementation

We next investigate whether or not attack outcomes can be improved by the incorporation of a second trace measurement corresponding to a different target function. In particular, we consider exploiting the joint leakage of key addition and the first DES S-Box, in the case that this is comprised of the Hamming weight of the target values plus some independent Gaussian noise.

Theoretic Outcomes

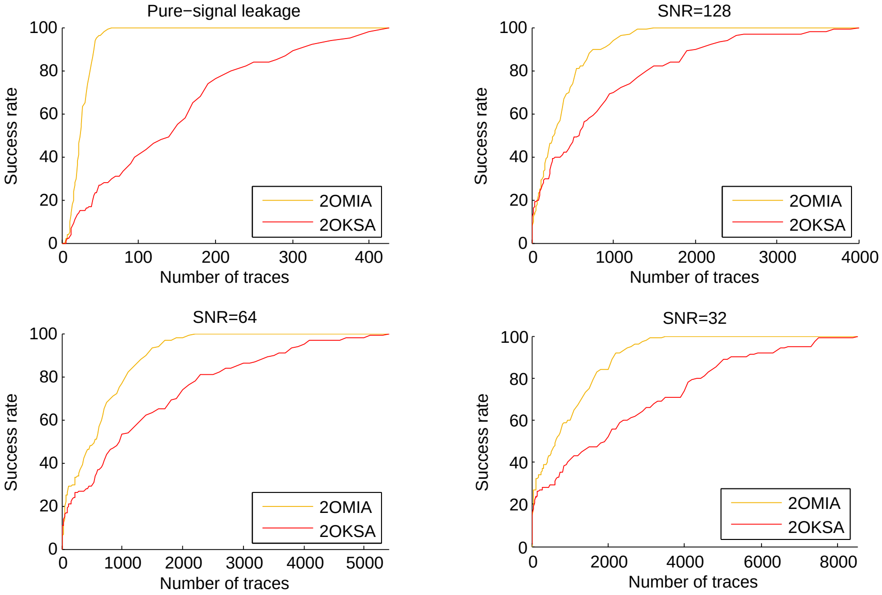

Fig. 8: *Success rates of attacks against a masked implementation of DES with HW leakage, as the number of traces increases.*

Pure-Signal Leakage: The noise-free distinguishability scores of bivariate MIA and KSA attacks are 3.6 and 1.7 respectively, compared with 5.6 and 4.2 for the equivalent univariate S-Box attacks. Thus, both methods are actually weakened by the incorporation of key addition leakage; KSA more so than MIA.

However, it is well documented that the resistance of a function to DPA has an inverse relationship with its resistance to cryptanalysis ([13]). In particular, the linearity of key addition makes it hard for DPA to distinguish between similar keys: small changes to the input produce small changes in the output. S-Boxes, on the other hand, are specially designed so that the converse is true, which makes them particularly vulnerable to DPA.

It is not, then, so surprising that key addition information detracts from attack distinguishability. If the leakage of two suitably nonlinear functions could be jointly targeted, our bivariate enhancement may prove more useful—we leave this as an open question.

As SNR Varies: Figure 9 shows the distinguishing scores of the bivariate attacks as compared with the univariate S-Box attacks, for varying levels of Gaussian noise. As with the univariate attacks, the bivariate KSA distinguisher is more robust to noise so that in very low-signal settings it exhibits a slight advantage over the bivariate (and indeed the univariate) MIA distinguisher. As in the application to the masked implementation, for all noise levels (i.e. including the noise-free setting) the bivariate distinguishing vectors are considerably less distinguishing than their univariate counterparts.

Practical Outcomes (Simulations) Figure 10 depicts the performance of practical bivariate attacks (against the DES S-Box and key addition jointly) as compared with univariate attacks against the DES S-Box alone. The lower theoretical distinguishing power, coupled with the additional complexity of estimation, mean that the bivariate attacks require more traces to be successful, in all tested noise settings. As with the univariate attacks, bivariate KSA performs very similarly to bivariate MIA.

These results (w.r.t. both theoretic and practical distinguishing vectors) are an important reminder that it is not the *quantity* of information which contributes to attack outcomes so much as the *quality*: identifying the most vulnerable targets is more likely to be fruitful than combining information from targets with differing degrees

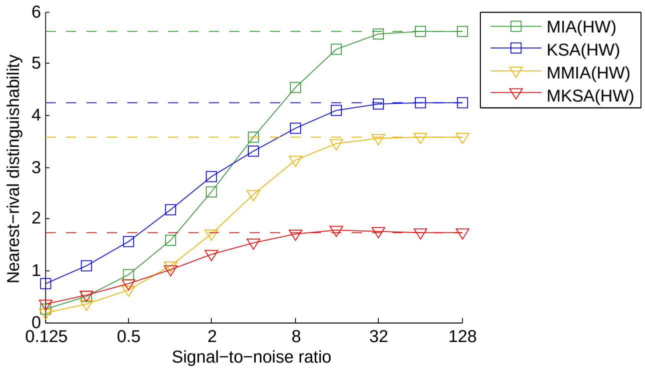

Fig. 9: *Theoretic distinguishing power as SNR varies, for bivariate attacks against DES with HW leakage.*

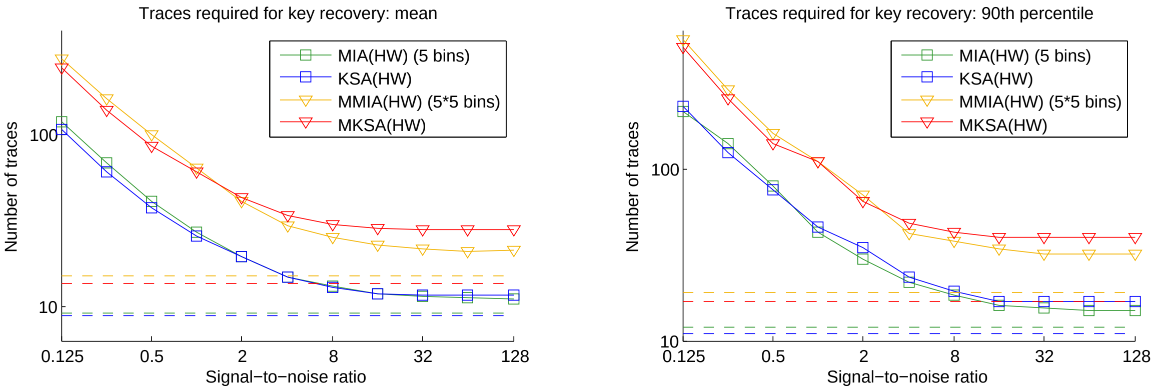

Fig. 10: *Mean and 90^{th} percentile of the trace requirement for key recovery, in repeated experiments against simulated HW leakage of AddRoundKey and the first DES S-Box jointly, as SNR varies.*

of DPA resistance. Moreover, univariate attacks remain less demanding in terms of computational complexity and the sample size required for estimation.

6 Conclusion

We have shown that the (two-sample) KS test statistic can be adapted to the purposes of DPA in a manner which bears considerable resemblance to MI-based DPA. We explored the theoretic and practical distinguishing vectors of KSA as compared with MIA, with a particular focus on scenarios that are relevant for practice.

Our findings showed that in noise-free or strong-signal univariate settings MIA was consistently the more distinguishing and more efficient attack, but when the signal was sufficiently weak the noise-robustness of KSA enabled it to gain an advantage.

The KSA distinguisher was found to share those characteristics of MIA which make it to some extent ‘power model free’; each can be adapted to use the identity power model in the case that an attacker lacks precise knowledge of the true data-dependent leakage (provided the target function is non-injective).

We also showed how a bivariate version of the (two-sample) KS test statistic enables extension to second-order KSA in order to defeat a masking scheme. However, here it was quite substantially outperformed by MIA in strong-signal settings and was so computationally complex as to be unfeasible in weak-signal settings. Moreover, whereas multivariate MI quite naturally incorporates additional data points, extensions of the KS test beyond 2 dimensions quickly become problematic so that there is little scope for third- or higher-order KSA.

A interesting question for future work is whether or not the known distribution of the KS test statistic could be used to formally derive the number of traces required for an attack to be successful, as has been accomplished in the case of correlation DPA (see §6.4 of [8]). Whilst the distribution of the KS test statistic is known it is unclear how it could be used to derive that of the KSA distinguisher (recall that this is defined as an average over several KS test statistics).

References

1. Akkar, M., Bevan, R., Dischamp, P., Moyart, D.: Power Analysis, What is Now Possible... In: Okamoto, T. (ed.) Advances in Cryptology ASIACRYPT 2000, Proceedings. pp. 489–502. Lecture Notes in Computer Science (2000)
2. Batina, L., Gierlichs, B., Prouff, E., Rivain, M., Standaert, F.X., Veyrat-Charvillon, N.: Mutual Information Analysis: A Comprehensive Study. Journal of Cryptology pp. 1–23 (2010)
3. Chari, S., Rao, J., Rohatgi, P.: Template Attacks. In: Kaliski, B., Koï£¡, e., Paar, C. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2002, Lecture Notes in Computer Science, vol. 2523, pp. 51–62. Springer Berlin / Heidelberg (2003)
4. Fasano, G., Franceschini, A.: A Multidimensional Version of the Kolmogorov-Smirnov Test. Monthly Notices of the Royal Astronomical Society 225, 155–170 (1987)
5. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual Information Analysis: A Generic Side-Channel Distinguisher. In: Oswald, E., Rohatgi, P. (eds.) Cryptographic Hardware and Embedded Systems – CHES 2008, Proceedings. Lecture Notes in Computer Science, vol. 5154, pp. 426–442. Springer-Verlag Berlin (2008)
6. Guilley, S., Hoogvorst, P., Pacalet, R.: Differential Power Analysis Model and Some Results. Smart Card Research and Advanced Applications Vi pp. 127–142 (2004)
7. Kraemer, Helena C., Thiemann, Sue: How Many Subjects?: Statistical Power Analysis in Research. Sage Publications, Inc, 1 edn. (September 1987), `http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0803929498`
8. Mangard, S., Oswald, E., Popp, T.: Power Analysis Attacks: Revealing the Secrets of Smart Cards. Springer (2007)
9. Mangard, S., Oswald, E., Standaert, F.X.: One for All - All for One: Unifying Standard DPA Attacks. IET Information Security (2011), to appear
10. Mangard, S., Pramstaller, N., Oswald, E.: Successfully Attacking Masked AES Hardware Implementations. In: Rao, J.R., Sunar, B. (eds.) Cryptographic Hardware and Embedded Systems – CHES 2005, 7th International Workshop, Edinburgh, UK, August 29 - September 1, 2005, Proceedings. Lecture Notes in Computer Science, vol. 3659, pp. 157–171. Springer (2005)
11. Paninski, L.: Estimation of Entropy and Mutual Information. Neural Computation 15(6), 1191–1253 (2003)
12. Peacock, J.: Two-Dimensional Goodness-of-Fit Testing in Astronomy. Monthly notices of the Royal Astronomical Society pp. 615–627 (1983)
13. Prouff, E.: DPA Attacks and S-Boxes. Fast Software Encryption 3557, 424–441 (2005)
14. Prouff, E., Rivain, M.: Theoretical and Practical Aspects of Mutual Information Based Side Channel Analysis. Applied Cryptography and Network Security 5536, 499–518 (2009)
15. Renauld, M., Standaert, F.X., Veyrat-Charvillon, N., Kamel, D., Flandre, D.: A formal study of power variability issues and side-channel attacks for nanoscale devices. In: Paterson, K. (ed.) Advances in Cryptology: EUROCRYPT 2011, Lecture Notes in Computer Science, vol. 6632, pp. 109–128. Springer Berlin / Heidelberg (2011)
16. Stephens, M.A.: EDF Statistics for Goodness of Fit and Some Comparisons. Journal of the American Statistical Association 69(347), 730–737 (September 1974)
17. Veyrat-Charvillon, N., Standaert, F.X.: Mutual Information Analysis: How, When and Why? In: Clavier, C., Gaj, K. (eds.) Cryptographic Hardware and Embedded Systems – CHES 2009, Proceedings. Lecture Notes in Computer Science, vol. 5747, pp. 429–443 (2009)
18. Whitnall, C., Oswald, E.: A Comprehensive Evaluation of Mutual Information Analysis Using a Fair Evaluation Framework. In: Proceedings of Crypto 2011. Springer (2011)