The myth of generic DPA... and the magic of learning

Carolyn Whitnall[1], Elisabeth Oswald[1], and François-Xavier Standaert[2]

[1] University of Bristol, Department of Computer Science,
Merchant Venturers Building, Woodland Road, BS8 1UB, Bristol, UK.
{carolyn.whitnall, elisabeth.oswald}@bris.ac.uk
[2] Université catholique de Louvain, UCL Crypto Group
Place du Levant 3, B-1348 Louvain-la-Neuve, Belgium.
fstandae@uclouvain.be

Abstract. A *generic* DPA strategy is one which is able to recover secret information from physically observable device leakage without any *a priori* knowledge about the device's leakage characteristics. Here we provide much-needed clarification on results emerging from the existing literature, demonstrating precisely that such methods (strictly defined) are *inherently* restricted to a very limited selection of target functions. Continuing to search related techniques for a 'silver bullet' generic attack appears a bootless errand. However, we find that a minor relaxation of the strict definition—the incorporation of some minimal non-device-specific intuition—produces scope for *generic-emulating* strategies, able to succeed against a far wider range of targets. We present stepwise regression as an example of such, and demonstrate its effectiveness in a variety of scenarios. We also give some evidence that its practical performance matches that of 'best bit' DoM attacks which we take as further indication for the necessity of performing profiled attacks in the context of device evaluations.

Keywords: side-channel analysis, differential power analysis

1 Introduction

Ever since Kocher et al. showed that differential power analysis (DPA) could be successful even with very little information about the target implementation [16], the research community has pursued 'generic' methods—informally, techniques able to recover secret information even in the total absence of knowledge about the attacked device's data-dependent power consumption. Recent suggestions include mutual information analysis (MIA) using an identity power model [12], distinguishers based on the Kolmogorov–Smirnov (KS) two-sample test statistic [30,35] and the Cramér–von Mises test [30], linear regression (LR)-based methods

This article is the author version of a paper appearing at CT-RSA 2014; the final publication is available from Springer, DOI: 10.1007/978-3-319-04852-9_10.

which can be seen as a sort of on-the-fly profiling [9,24], and an innovative approach using copulas [31].

However, all existing proposals share a common shortfall when applied to injective target functions: in order to distinguish between hypotheses the attacker must, after all, have some meaningful piece of knowledge by which to partition the measurements (in the case of MIA and KS-based DPA) or select the appropriate set of covariates (in the case of LR-based DPA) [31]. Unfortunately, this dependence on prior knowledge has been under-appreciated because of the apparent success of 'arbitrary' work-arounds such as the practice of partitioning intermediate variables according to their 7 least significant bits (sometimes called the 7LSB model). However, it is shown in [34] that this strategy is far from universally-applicable and only works to the extent that the seemingly indifferent partition captures something meaningful about the leakage after all. For example, noise on top of a typical CMOS Hamming weight consumption distorts the trace measurements *towards* the 7LSB model sufficiently for MIA to succeed, but this is not the case in general (i.e in arbitrary leakage scenarios). Such attacks can no longer be considered 'generic', a description which is earned primarily by virtue of the non-reliance on *a priori knowledge* rather than the chosen statistical methodology. The focus on defining universally-applicable *distinguishers* indicates a confusion about the role of the distinguisher and that of the power model in what has so far been only informally defined as 'generic' DPA. It also raises the fundamental question of whether *truly* 'generic' tools exist at all.

Establishing whether or not generic DPA attacks exist has fundamental consequences for the process of cryptographic device evaluation. The presence of generic attacks would imply that any device could potentially be attacked without any information about its internal functioning or leakage characteristics. Consequently, attacks based on profiling would only be 'better' in terms of efficiency (number of power traces needed)—not in terms of applicability. The absence of generic attacks would imply that there exist devices (leakage characteristics) which can only be evaluated soundly by performing profiled attacks—a practice which is not commonly undertaken at present (see, e.g., [19] Appendix F). In the following, we tackle this important question in the practically relevant context of standard DPA as investigated, e.g., in [9,12,16,24,30,35]. That is, we assume that the mean of the side-channel leakage distributions is key-dependent.

1.1 Our contribution

We first develop a theory of power models according to Stevens' 'levels of measurement' [28], enabling us to formally define what constitutes a *generic power model*. We show that different distinguishers require different types of power model and derive the notion of a *generic-compatible distinguisher* accordingly. The pairing of a *generic-compatible distinguisher* with the *generic power model* we call a *generic strategy*. These definitions provide a basis for making conclusive general statements about generic DPA. We show that the noninjectivity of the target function is a prerequisite for *any* first-order generic strategy to succeed,

proving the absence of a universally-applicable generic distinguisher in the context of first-order DPA! (Generic higher-order DPA can only be *more* difficult, so this conclusive statement naturally extends upwards). As a further finding we observe that noninjectivity alone is not sufficient for generic success, and investigate additional requirements on the target function. It is already known that there is an inverse relationship between performance against certain S-box criteria and susceptibility to DPA [21]; we demonstrate a sufficient condition for first-order generic success which is promoted (though not inevitably produced) by the desirable S-box property of *differential uniformity* [20].

Having ruled out the possibility of a universally-applicable generic distinguisher, we investigate *minimal* relaxations on the generic criteria producing theoretically plausible attack strategies. As a starting point we take the LR-based distinguisher [9,24], which (we show) qualifies as generic-compatible but returns more auxiliary information than other such methods when applied against an injective target. Hence, even though the keys remain indistinguishable in the ranking (as is consistent with the first half of this paper and with earlier studies [31]), the hypothesis-dependent model estimates—i.e. the estimated coefficients in the polynomial expression for the leakage—contain additional clues about the correct key. At this stage we introduce some 'non-device-specific intuition' regarding the simplicity of the leakage function relative to the cryptographic target function (typically an S-box). This extremely minimal assumption (which we will explain more formally in due course) allows us to exploit the model estimates, which we propose to do using the techniques of *stepwise regression*. Such a strategy is no longer strictly generic, but the general device-independent nature of the extra assumption prompts us to coin the description *generic emulating*. We verify that this proposed strategy truly is effective—even against injective target functions such as the AES and PRESENT S-boxes, and even as the true leakage becomes increasingly unusual or complex (high-degree polynomials, for example). We also show that the proposed strategy is efficient, albeit seemingly no better in performance than difference-of-means (DoM) based attacks.

2 Preliminaries

2.1 Differential power analysis

We consider a 'standard DPA attack' scenario as defined in [18], and briefly explain the underlying idea as well as introduce the necessary terminology here. We assume that the power consumption T of a cryptographic device depends on some internal value (or state) $F_{k^*}(X)$ which we call the *target*: a function $F_{k^*} : \mathcal{X} \to \mathcal{Z}$ of some part of the known plaintext—a random variable $X \overset{R}{\in} \mathcal{X}$—which is dependent on some part of the secret key $k^* \in \mathcal{K}$. Consequently, we have that $T = L \circ F_{k^*}(X) + \varepsilon$, where $L : \mathcal{Z} \to \mathbb{R}$ describes the data-dependent component and ε comprises the remaining power consumption which can be modeled as independent random noise (this simplifying assumption is common in the literature—see, again, [18]). 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 $M : \mathcal{Z} \to \mathcal{M}$.

DPA is motivated by the intuition that the model predictions under the correct key hypothesis should give more information about the true trace measurements than the model predictions under an incorrect key hypothesis. A distinguisher D is some function which can be applied to the measurements and the hypothesis-dependent predictions in order to quantify the correspondence between them. For a given such comparison statistic, D, the *theoretic* attack vector is $\mathbf{D} = \{D(L \circ F_{k^*}(X) + \varepsilon, M \circ F_k(X))\}_{k \in \mathcal{K}}$, and the *estimated* vector from a practical instantiation of the attack is $\hat{\mathbf{D}}_N = \{\hat{D}_N(L \circ F_{k^*}(\mathbf{x}) + \mathbf{e}, M \circ F_k(\mathbf{x}))\}_{k \in \mathcal{K}}$ (where $\mathbf{x} = \{x_i\}_{i=1}^N$ are the known inputs and $\mathbf{e} = \{e_i\}_{i=1}^N$ is the observed noise). Then the attack is *o-th order theoretically successful* if $\#\{k \in \mathcal{K} : \mathbf{D}[k^*] \leq \mathbf{D}[k]\} \leq o$ and *o-th order successful* if $\#\{k \in \mathcal{K} : \hat{\mathbf{D}}_N[k^*] \leq \hat{\mathbf{D}}_N[k]\} \leq o$.[3]

Definition 1 *A practical instantiation of a* standard univariate DPA attack *computes, given a set of power traces* $\mathbf{T}$*, a prediction model* M*, a set of inputs* $\mathbf{X}$*, and a comparison statistic* D*, the distinguishing vector* $\hat{\mathbf{D}}_N = \{\hat{D}_N(L \circ F_{k^*}(\mathbf{x}) + \mathbf{e}, M \circ F_k(\mathbf{x}))\}_{k \in \mathcal{K}}$*. A practical instantiation is said to be* o-th order successful *if* $\#\{k \in \mathcal{K} : \hat{\mathbf{D}}_N[k^*] \leq \hat{\mathbf{D}}_N[k]\} \leq o$.

2.2 Measuring DPA outcomes

Metrics to compare the *efficiency* of DPA attacks include the *(o-th order) success rate* and the *guessing entropy* of [27]—defined respectively as the probability of o-th order success and the expected number of key hypotheses remaining to test after a practical attack on a given number of traces. However, in the evaluation of generic strategies, the question of asymptotic feasibility takes precedence over that of efficiency. By the law of large numbers $\frac{1}{N}\sum_{i=1}^N L \circ F_{k^*}(x) + e_i \to L \circ F_{k^*}(x)$ as $N \to \infty$ (as long as the samples are independent and identically distributed). We can therefore discuss feasibility from the perspective of the *ideal* distinguishing vector $\mathbf{D}_{IDEAL} = \{D(L \circ F_{k^*}(X), M \circ F_k(X))\}_{k \in \mathcal{K}}$, noting that this no longer depends on the noise but only on the hypothesis-dependent power models relative to the true leakage. Indeed, averaging the trace measurements conditioned on the inputs is a popular pre-processing step in practice as it strips out irrelevant variance and reduces the dimensionality of the computations (see, for example, [1]); it is a sound approach as long as the side-channel information to be exploited originates in differences between the mean values of the leakage distributions, which *is* the case in our standard DPA scenario.

For the purposes of evaluating the theoretic capabilities of generic emulating and related strategies, we will focus on first-order asymptotic success, as captured

[3] Note that standard DPA attacks do not include collision-based attacks [25], which exploit information from several leakage points per observation, and do not require a power model at all.

by the (ideal) nearest-rival distinguishing margin (see [33,34]):
$NRMarg(\mathbf{D}_{IDEAL}) = \mathbf{D}_{IDEAL}[k^*] - \max\{\mathbf{D}_{IDEAL}[k] | k \neq k^*\}$. In Sect. 4.6, where we investigate the practical performance of our proposed generic emulating distinguisher, we report success rates for attacks against simulated leakages.

2.3 Boolean vectorial functions

We are often interested in the special case that the key-indexed functions F_k can be expressed as $F_k(X) = F(k * X)$ where $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ is an (n-m) Boolean vectorial function and $*$ denotes the *key combining* operator (e.g., XOR). It particularly pertains to the study of block ciphers, and their associated S-boxes.

Certain algebraic properties of such functions are known to be particularly important to the *cryptanalytic* robustness of a cipher system. We (very) briefly recall those concepts that will play a role in our later analysis; for a good basic introduction see [14] or, for a more comprehensive explanation, [6,7].

F is *affine* if it can be expressed as a linear map followed by a translation—that is, if there exists a matrix $M \in \mathbb{F}_2^{m \times n}$ and a vector $v \in \mathbb{F}_2^m$ such that $F(x) = Mx \oplus v$. *Nonlinearity* is defined as: $N_F = \min_{u \in \mathbb{F}_2^n, v \in \mathbb{F}_2^m \setminus \{\mathbf{0}\}} \sum_{x \in \mathbb{F}_2^n} u \cdot x \oplus v \cdot F(x)$.

F is *balanced* if the preimages in F of all singleton subsets of $\mathbb{F}_2^m$ are uniformly sized: that is, $\forall y \in \mathbb{F}_2^m$, $\#\{x \in \mathbb{F}_2^n | F(x) = y\} = 2^{n-m}$. This property applies to many functions used in block ciphers, particularly S-boxes [36] where any bias on the unobserved inputs is extremely undesirable.

Another desirable S-box property is *differential uniformity* [20]—that the derivatives of F with respect to $a \in \mathbb{F}_2^n$ (defined as $D_a F(x) = F(x) \oplus F(x \oplus a)$) be *as uniform as possible*. If there exists a vector $a \in \mathbb{F}_2^n$ such that $D_a F(x)$ is constant over $\mathbb{F}_2^n$ then a is called a *linear structure* of F and (as per [10]) can be exploited by a cryptanalyst. $\{a \in \mathbb{F}_2^n | D_a F = cst\}$ is the *linear space* of F.

3 Clarifying generic DPA

What does it mean for an attack to be 'generic'? The discussion in the literature has focused on appropriating, as distinguishers, statistics which 'require few distributional assumptions'—trawling the statistical literature for nonparametric, distribution-comparing procedures such as the Kullback-Leibler divergence (a.k.a. Mutual Information Analysis) [12], the Kolmogorov–Smirnov [30,35] and Cramér–von Mises [30] tests, and copulas [31]. However, the emphasis on finding 'distribution-free' statistics for use as distinguishers somewhat distracts from the essential defining feature of generic DPA which is that *no assumptions have been made about the device leakage*. Clearly, the (fairly common) practice of combining such distinguishers with an informed prior model does not produce a generic attack: we need to begin by establishing what constitutes a *generic power model*.

We first delineate the different types of model used in DPA attacks, and discuss which distinguishers are suitable in each instance. We can then define a generic power model, a generic-compatible distinguisher, and a generic DPA strategy. These definitions form the basis for a number of propositions that clarify

the cases in which any generic strategy is bound to fail (we spell out necessary conditions for success and discuss further the feasibility of generic DPA).

3.1 Delineating leakage assumptions

Firstly we must distinguish between assumptions about the *data-dependent* leakage, as captured by the power model, and assumptions about the *distribution of the noise*—which in most cases play a less visible role, but can affect how accurately or efficiently certain statistics may be estimated. Fig. 1 visualises this two-dimensional continuum, and indicates the suitability of popular distinguishers as assumptions vary.

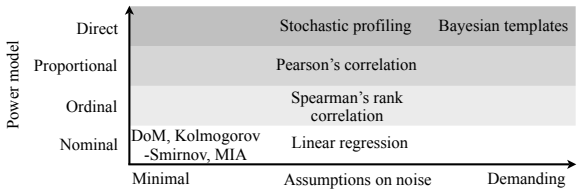

Fig. 1: *Types of leakage model and the assumptions required by common distinguishers.*

Assumptions about the noise range from *fully characterised distributions* as exploited (e.g.) by Bayesian template attacks, down to *no knowledge whatsoever*, when the robustness of nonparametric statistics such as mutual information and the Kolmogorov–Smirnov test may come in handy. Fortunately, the often reasonable assumption of approximate normality opens up a broad range of (semi-)parametric options, which are to be preferred as they are inherently less costly to estimate.

We now consider the *nature* of the power model, with which this paper is primarily concerned. Previous studies have talked about 'good' power models, in an arbitrary sense, and most have missed the very material distinction between different *levels* of model. As hinted towards in [2,11], the widely-accepted 'levels of measurement'—ratio, interval, ordinal, nominal—laid out by Stevens [28] present a natural framework for delineation. It is important to understand the appropriate (type-specific) notion of accuracy for a given model, and to select a compatible distinguisher; that is, one which (implicitly) interprets the model according to the correct type.

The type of power model exploited by profiled attacks (e.g. Bayesian templates [8] and stochastic profiling [24]) amounts to a *direct approximation* of the actual power consumed by processing the data, in contribution to the overall consumption. This requirement is the most demanding possible, expressed as $M \approx L$ (c.f. the 'ratio scale' of [28]). The outcome of an attack will depend on

how accurately the templates approximate the actual data-dependent consumption (as well as the noise distribution). The error sum-of-squares is a natural way of quantifying the appropriate notion of accuracy.

Less demanding is the requirement that the attacker has a power model which is a good approximation for L *up to proportionality*: $M \approx \alpha L$ (c.f. the 'interval scale' of [28]). Pearson's correlation coefficient provides a natural way to quantify accuracy and can be directly adapted for use as a distinguisher [4] (a popular strategy since, as a simple, moment-based statistic, it can usually be estimated very efficiently with respect to the number of trace measurements required).

Less demanding again is the requirement that M approximates L *up to ordinality*: $\{z|M(z) < M(z')\} \approx \{z|L(z) < L(z')\}\ \forall z' \in \mathcal{Z}$ (c.f. the 'ordinal scale' of [28]). Such a model could be exploited via a variant of correlation DPA using Spearman's rank correlation coefficient, as proposed in [2]. And, again, the accuracy of the model can be quantified via the rank correlation itself.

The least demanding requirement to place on a model is that it approximates the leakage function *up to nominality* only: $\{z|M(z) = M(z')\} \approx \{z|L(z) = L(z')\}\ \forall z' \in \mathcal{Z}$ (c.f. the 'nominal scale' of [28]). As ever, such a model must be paired with a statistic which interprets the values appropriately: that is to say, as arbitrary labels only. In fact, these correspond to the 'partition-based' distinguishers of [26]. Typical examples include statistics which are used to compare arbitrary distributions, such as MI [12] and the KS test statistic [30,35]. Kocher et al.'s original Difference-of-Means (DoM) test [16] also falls into this category, but is limited in how much information it is able to exploit as it is only able to operate with a two-way partition model. To produce this partition, either the value of a single bit is used (in which case the other bits act as algorithmic noise, increasing the data complexity of the attack), or combinations of multiple bits are used, which results in discarded traces (instances not fitting into either category).

Appropriate notions of accuracy for a nominal model are drawn from classification theory. *Precision* is the probability that items grouped according to the model really do belong together, whilst *recall* is the probability that items which belong together are identified as such (see, e.g. [17]).[4]

$$\begin{aligned} Precision(M) &= \mathbb{P}(L(z) = L(z')|M(z) = M(z')), \\ Recall(M) &= \mathbb{P}(M(z) = M(z')|L(z) = L(z')). \end{aligned}$$

3.2 Defining 'genericity'

We are now in a position to discuss the generic power model: what, in practice, does it mean to make *no* assumptions about the data-dependent leakage? Essentially, that we do no more than to assign a distinct label to each

[4] The classification theory literature more frequently states these definitions in terms of ratios of counts—practically convenient but less directly translatable across contexts. See [13] for a more explicit probabilistic interpretation; though in our case we are, of course, averaging over multiple classes.

value in the range of the target function. These labels can be seen to correspond to the key-dependent equivalence classes produced by the preimages of F_k: $[x]_k = F_k^{-1}[F_k(x)]\ \forall x \in \mathcal{X}$.

Definition 2 *The* generic power model *associated with key hypothesis $k \in \mathcal{K}$ is the nominal mapping to the equivalence classes induced by the key-hypothesised target function F_k.*

The 'identity' power model emphasised in previous literature is fine for this purpose as long as it is understood that the mapping is simply a convenient labelling system and should be interpreted *nominally* only. It is clear, then, that the *generic-compatible distinguishers* are precisely those (described in Sect. 3.1 above) which interpret hypothesis-dependent predictions as an approximation up to nominality of the data-dependent leakage.

Definition 3 *A distinguisher is* generic-compatible *if it is built from a statistic which operates on nominal scale measurements.*

This provides valuable clarification on previous work such as [3], which demonstrated successful attacks against Hamming weight leakage using correlation DPA with an 'identity' power model. The authors rightly remarked that this was possible precisely because, over $\mathbb{F}_2^4$, the identity is sufficiently accurate as a *proportional* approximation of the Hamming weight to produce a successful correlation attack. Far from operating generically, the identity mapping in such a strategy is interpreted as an interval scale model—not a perfect approximation but adequate in the specific case that L can be well-approximated by the Hamming weight. And even in this restricted case it is not, of course, invariant to permutation of the 'identity' labels.

Definitions 2 and 3 combine towards a natural notion of a 'generic strategy':

Definition 4 *A* generic strategy *performs a standard univariate DPA attack using the generic power model paired with a generic-compatible distinguisher.*

However, as previous work on 'partition-based' distinguishers (separately, e.g. [12,31,35], and collectively [26]) has consistently noted, not all (indeed, not many) scenarios are suited to a generic strategy.

3.3 Conditions for a generic strategy to succeed

All distinguishers operate by identifying the key hypotheses producing the most accurate model predictions for the actual measurements, according to the appropriate notion of accuracy for the model type (some are able to perform this comparison more effectively or from fewer trace measurements). In the generic setting each key hypothesis $k \in \mathcal{K}$ gives rise to a model M_k s.t. $M_k^{-1}[z] = F_k^{-1}[z]$ $\forall z \in F_k(\mathcal{X})$, and it is the comparative *nominal* accuracy which will determine key-recovery success. We can therefore explore the conditions necessary for a successful attack—independently of any particular distinguisher—by reasoning

directly about the accuracy of F_{k^*} and F_k, $\forall k \in \mathcal{K} \setminus \{k^*\}$ as nominal approximations for $L \circ F_{k^*}$. Recall the precision and recall measures introduced in Sect. 3.1 (with $\mathbb{E}$ to denote expectation):

$$\begin{aligned} Precision(M_k) &= \mathbb{P}(L \circ F_{k^*}(x) = L \circ F_{k^*}(x') | F_k(x) = F_k(x')) \\ &= \mathbb{E}_{x \in \mathcal{X}} \left[\frac{\#F_{k^*}^{-1}[L^{-1}[L \circ F_{k^*}(x)]] \cap F_k^{-1}[F_k(x)]}{\#F_k^{-1}[F_k(x)]} \right] \end{aligned}$$

$$\begin{aligned} Recall(M_k) &= \mathbb{P}(F_k(x) = F_k(x') | L \circ F_{k^*}(x) = L \circ F_{k^*}(x')) \\ &= \mathbb{E}_{x \in \mathcal{X}} \left[\frac{\#F_{k^*}^{-1}[L^{-1}[L \circ F_{k^*}(x)]] \cap F_k^{-1}[F_k(x)]}{\#F_{k^*}^{-1}[L^{-1}[L \circ F_{k^*}(x)]]} \right] \end{aligned}$$

Trivially, the precision of the generic model under the correct hypothesis is always maximal (the leakage preimage must contain the function preimage). By contrast, the recall depends additionally on the true leakage function, so that even under the correct hypothesis we do not get perfect recall unless it happens that L is also injective. The ability of a strategy to reject an *incorrect* alternative requires the corresponding model to be of inferior quality; whether this is so depends on features of F_k and L. An immediate and quite restrictive pre-requisite arises from the inherent nature of the generic power model:

Proposition 1. *No generic strategy is able to distinguish the correct key k^* from an alternative hypothesis k if F_{k^*} and F_k are injective.*

Proof. If F_{k^*}, F_k are injective then $\forall x \in \mathcal{X}$, $F_k^{-1}[F_k(x)] = F_{k^*}^{-1}[F_{k^*}(x)] = \{x\}$. Each hypothesis produces models of equivalent nominal accuracy—no generic-compatible distinguisher can separate the candidates.

Indeed, all of the known generic-compatible distinguishers, from the seminal CHES '08 paper on MIA [12] to the recent copula-based method presented at Crypto '11 [31], have individually been shown to fail whenever the *composition* of the target function and the power model is injective; the same observation was made for the entire class of 'partition-based' distinguishers described in [26]. The authors duly noted that some restriction was required on the power model in order for these distinguishers to operate against an injective target, but left as an open question the existence (or demonstrable non-existence) of an as-yet undiscovered method which would somehow circumvent this requirement. Demonstrating that the limitation is attributable directly to the generic power model rules out this possibility.

Noninjectivity is therefore a necessary condition, but not, as we next establish, a sufficient one. In the general case it is rather difficult to formulate useful, concrete observations so we will henceforth narrow down to the restricted but highly relevant case that F is a *balanced* (n-m) function and k is introduced by key addition (as described in Sect. 2.3). It then becomes fairly straightforward to draw out such function characteristics as will obstruct a generic strategy.

Proposition 2. *Suppose F is a balanced, non-injective (n-m) function, with k introduced by (XOR) key addition, i.e. $F_k(x) = F(x \oplus k)$. Then:*

(a) If F is affine then no generic strategy is able to distinguish the correct key k^* from any $k \in \mathcal{K} \setminus \{k^*\}$.
(b) If $a \in \mathbb{F}_2^n$ is a linear structure of F then no generic strategy is able to distinguish between k^* and $k^* \oplus a$.
(c) If, for some $a \in \mathbb{F}_2^n$ we have that $D_aF(x)$ depends on x only via *$F(x)$, then no generic strategy is able to distinguish between k^* and $k^* \oplus a$.*

The proof of Proposition 2 can be found in Appendix A. Part (a) arises from the fact that all key hypotheses produce indistinguishably 'good' models for the leakage; the distinguishing vector produced by such an attack would be flat and maximal across all hypotheses.

The implication of 2(b) is that $k^* \oplus a$ cannot be rejected if the derivative of F with respect to a is *constant* over the *domain* of F, i.e $\#D_aF(\mathbb{F}_2^n) = 1$. In such a case we would expect a practical attack to exhibit a *ghost peak* at $k^* \oplus a$ [4]; [21], notes a corresponding phenomenon for correlation DPA.

Part (c) can be otherwise expressed as the fact that $k^* \oplus a$ cannot be rejected if the derivative of F with respect to a is *constant* over *each singleton preimage* of F, i.e. $\#D_aF(F^{-1}[F(x)]) = 1 \; \forall x \in \mathbb{F}_2^n$. We have actually observed this property in the fourth DES S-box, for the key-offset $a = 47_{(10)} = 101111_{(2)}$: consequently, $k^* \oplus 47$ produces a 'ghost peak' in the distinguishing vector, with a nonetheless substantial margin between these *two* and the remaining hypotheses—a good example of an attack scenario with a low first-order, but high second-order, success rate [27]. Our observation is consistent with (and illuminates) past works such as [5] which recognised the unusual operation of DPA distinguishers confronted with this particular S-box/offset combination.

Thus emerges a minimal requirement for k^* to be distinguished from k:

Proposition 3. *Suppose F is a balanced, noninjective (n-m) function, with k introduced by (XOR) key-addition. A necessary condition for a generic strategy to distinguish k^* from k is: $\exists x \in \mathbb{F}_2^n$ such that $\#D_{k^* \oplus k}F(F^{-1}[F(x)]) \neq 1$. If L is injective then this becomes a sufficient condition.*

This is informally expressed as the requirement that there is at least one (singleton) preimage over which the derivative with respect to $k^* \oplus k$ is *not* constant. The proof follows from our reasoning in support of Proposition 2 and can be found in Appendix A along with a toy example to demonstrate that we can no longer claim sufficiency if L is noninjective.

Recall from Sect. 2.3 the idea that the derivatives of an S-box should ideally be close to uniform—thus maximising entropy; affine functions or functions with non-null linear spaces represent the extreme in terms of cryptanalytic vulnerability. The pursuit of such a design goal would not guarantee the minimal condition above, as even a perfectly balanced derivative could be so arranged as to be constant over the singleton preimages (which are of cardinality 2^{n-m} since F is also balanced). However, it would certainly seem to increase the chance that the condition be met for a given key-offset, as the more finely D_aF partitions $\mathbb{F}_2^n$, the fewer the possible *refinements* into 2^m (balanced) parts. Therefore, among the (already restricted) class of noninjective S-boxes we would expect

ghost peaks and indistinguishable keys to be a rarity—even more so as the size of the S-box increases.

4 Introducing generic-emulating DPA

Most existing generic-compatible distinguishers return only some sort of 'classification accuracy', leading them to fail against injective targets. But, on examination of the literature, LR-based attacks emerge as an interesting candidate for generic DPA: they can be used with a full basis of polynomial terms (equivalent, we shall show, to a generic power model), but possess additional features that may possibly be exploited. In particular, further to the distinguishing vector of goodness-of-fit values, LR-based DPA also returns the estimated model coefficients, which differ by key hypothesis. In this section we explore how the coefficients may be interpreted in the light of some simple, non-device-specific intuition to reveal the correct key, and show that the process can be automated straightforwardly using LR in a stepwise mode.

We begin by introducing (standard) LR-based DPA, explaining the mechanism by which it distinguishes the correct key, and demonstrating that it is among the class of generic-compatible distinguishers. We then present the 'generic-emulating' stepwise linear regression- (SLR-) inspired variant which exploits the non-device-specific intuition to successfully attack injective targets even with 'no' (other) prior knowledge. We finally demonstrate the effectiveness of these distinguishers against well-known (injective and noninjective) S-boxes, as the level of prior knowledge available varies from 'complete' to 'none'.

4.1 Introduction to linear regression-based DPA

The motivation for an LR-based approach begins with the observation that $L : \mathbb{F}_2^m \to \mathbb{R}$ can be viewed as a pseudo-Boolean vectorial function with a unique expression in numerical normal form [6]. That is to say, there exists coefficients $\alpha_u \in \mathbb{R}$ such that $L(z) = \sum_{u \in \mathbb{F}_2^m} \alpha_u z^u$, $\forall z \in \mathbb{F}_2^m$ (z^u denotes the monomial $\prod_{i=1}^m z_i^{u_i}$ where z_i is the i^{th} bit of z). Finding those coefficients amounts to finding a power model for L in polynomial function of the coordinate functions of F. As first observed in [24], and demonstrated in [9], linear regression can be adapted to non-profiled key-recovery: the true leakage function is estimated 'on-the-fly' and recovered synchronously with the true key.

Appendix B provides background on linear regression; in short, the LR-based attack uses ordinary least squares to estimate, for each $k \in \mathcal{K}$, the parameters of the model $L_{k^*}(X) + \varepsilon = \alpha_{\mathbf{0}} + \sum_{u \in \mathcal{U}} F_k(X)^u \alpha_u$ where $\mathcal{U} \subseteq \mathbb{F}_2^m \setminus \{\mathbf{0}\}$. The distinguishing vector comprises the R^2 measure of fit from each of these models: $D_{\mathrm{LR}}(k) = \rho(L_{k^*}(X) + \varepsilon, \hat{\alpha}_{k,0} + \sum_{u \in \mathcal{U}} F_k(X)^u \hat{\alpha}_{k,u})^2$ (where ρ denotes Pearson's correlation coefficient). It can be viewed as a generalisation of correlation DPA, where the power model M is known *a priori*: $D_\rho(k) = \rho(L_{k^*}(X) + \varepsilon, M \circ F_k(X))$. In each case, the value of k which produces the *largest* distinguisher value is selected as the key guess.

4.2 Linear regression is generic-compatible

In the way the distinguisher is naturally presented, the attacker's prior knowledge is contained within $\mathcal{U}$; it is not immediately obvious exactly what *is* the power model, or where it fits alongside the various types presented in Sect. 3.1. In fact, each $u \in \mathcal{U}$ could be seen to represent a *separate* power model which divides the traces into two nominal classes: $\{x \in \mathbb{F}_2^n | F_k(x)^u = 1\}$ and $\{x \in \mathbb{F}_2^n | F_k(x)^u = 0\}$.[5] Intuitively, as long as the power consumption really *does* differ systematically according to the bit-interaction term represented by u, then this 'approximation' has low precision but high recall under the correct key hypothesis, and loses accuracy under an incorrect hypothesis as long as the function F is such that changes to the input produce nonuniform changes to the output. In fact, this is the mechanism by which the original difference-of-means DPA [16] operates!

So the linear regression distinguisher could be viewed as an extension of difference-of-means DPA—a means of exploiting *multiple* (overlapping) nominal approximations, each of low precision (and therefore weak as standalone models) but in conjunction providing a refined description of the leakage.

Intuitively, the generic instantiation should correspond to $\mathcal{U} = \mathbb{F}_2^m \setminus \{\mathbf{0}\}$ (i.e., imposing no restrictions on the leakage form). But our previous reasoning about the operation of generic strategies supposed a *single* power model (F_k, interpreted nominally) and it is hard to see how we might begin to reason about the impact of multiple power models. Fortunately, in the $\mathcal{U} = \mathbb{F}_2^m \setminus \{\mathbf{0}\}$ case *only*, the operation of the distinguisher *can* be re-framed in terms of the generic power model as defined above, so that all of our prior reasoning applies.

Proposition 4. *The linear regression-based DPA attack with a full set of covariates $\mathcal{U} = \mathbb{F}_2^m \setminus \{\mathbf{0}\}$ constitutes a generic strategy.*

We sketch a proof as follows: If M_k is an arbitrary labelling on F_k, we can always map bijectively to $\mathbb{F}_2^m$ to acquire an arbitrary *permutation* of the function outputs $M'_k(x) = p \circ F_k(x)$. For each $u \in \mathbb{F}_2^m$, the associated monomial $M'_k(x)^u$ has a unique expression in numerical normal form $M'_k(x)^u = \sum_{v \in \mathbb{F}_2^m} b_v F_k(x)^v$, $b_v \in \mathbb{R}$ [6]. So the system of equations relating to an incorrect hypothesis k can be re-written in function of $F_k(x)$ by substituting in these expressions, expanding out and collecting up the terms. We end up with different values of α_u, $u \in \mathbb{F}_2^m$ whenever we reparametrise in this way, but, crucially, the terms in the equation *collectively* explain the measured traces equally well—and it is in this sense that linear regression DPA is *invariant to re-labelling* and therefore can be discussed alongside other generic-compatible strategies (though it is not usually used in this way—particularly as meaningful restrictions on $\mathcal{U}$ contribute to efficiency gains in the estimation stage).

As we would expect from Sect. 3, LR-based DPA fails against injective targets when used generically (i.e. with $\mathcal{U} = \mathbb{F}_2^m \setminus \{\mathbf{0}\}$). This failure can be better

[5] Note that the labeling is irrelevant since they are represented in the regression equation by dummy variables: the 1/0 assignment is arbitrary and will impact only the estimated coefficients, not the R^2.

understood when we consider that the data-dependent part of the power consumption can be expressed as a system of 2^n equations (in function of $F_k(x)$) with 2^n unknowns. Because this system is fully-determined and consistent under any key hypothesis it *always* has a perfect solution, so as to produce a flat distinguishing vector of maximal R^2s.[6]

4.3 Exploiting non-device-specific intuition

The unique opportunity presented by generic LR arises from the fact that it produces, not just the distinguishing vector of R^2 values (which are unable to discriminate between hypotheses when the target is injective), but also the hypothesis-dependent sets of estimated coefficients. When $k = k^*$ these give the correct expression for L in function of the output bits; the rest of the time, they give an expression for $L \circ F_k \circ F_{k^*}^{-1}$. If, then, the attacker was able to recognise the correct expression, he would be able to identify the secret key.

Thus motivated, we examine the correct and incorrect expressions for L in the case that the target function is an injective S-box (of size 8 bits in the case of AES, or 4 bits in the case of PRESENT) and that the true form of the leakage is the Hamming weight: $L(z) = \sum_{i=0}^{m} z^{2^i}$. Fig. 2 shows the coefficients, in the polynomial expression for L, on the covariates as produced by the true key k^* (in black) and on those as produced under an incorrect hypothesis k' (in grey). The high nonlinearity of the S-box functions ensure that, when viewed as a polynomial in $F_k(X)$ rather than $F_{k^*}(X)$, the leakage function L is also highly nonlinear in form.

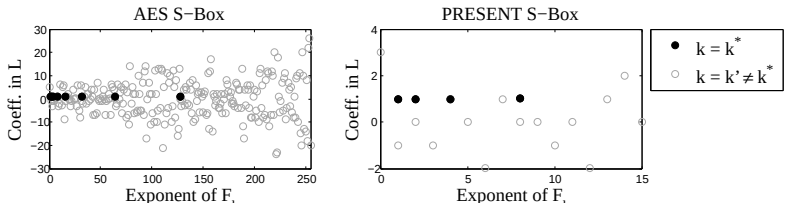

Fig. 2: *Coefficients, in the fitted expression for L, on the covariates as predicted under the correct and an alternative hypothesis.*

In the face of such evidence an attacker would be justified in favouring hypothesis k^* over k: intuitively, it seems more likely (especially given the known high nonlinearity of F) that the 'simpler' expression (i.e. the one corresponding to the black circles in Fig. 2) is the correct one. To exploit the extra information

[6] In the case of noninjective targets, the system is *overdetermined* (2^n equations, 2^m unknowns). Provided the target satisfies the criteria in Sect. 3.3 then this system is *only* consistent under the correct key hypothesis (thus only then does it have a perfect solution—there are only 2^m *linearly independent* equations).

represented by the coefficients, we therefore need to trust this intuition (which implicitly also assumes that $M_k = F_k$). This takes us a step away from the generic strategy—but since the intuition is not specific to any particular device it appears to be a very small step. That is, we just need to assume that the leakage function is 'sufficiently simple' compared to the target function. This is justified for a wide range of devices manufactured in CMOS technologies, including advanced 65-nanometer processes [23]. In fact, even for protected logic styles such as introduced by Tiri and Verbauwhede [29], it turns out that ensuring a complex (e.g. highly nonlinear) leakage function is a challenging task [22]. Besides, the results in Sect. 4.5 will also demonstrate that this 'simplicity constraint' on the leakage function can be quite relaxed.

Of course, comparing graphs is not ideal from a practical perspective, besides which the true leakage function may not always have so simple a form as to be visibly discernible: we would like to encapsulate the underlying reasoning into an automated and systematic procedure for testing hypotheses. In the next section we introduce a learning technique from data mining which uses our non-device-specific intuition about 'what the leakage should look like' to produce, in a wide range of leakage scenarios, asymptotically successful key recovery against injective targets *even when provided with the full set of covariates* $\mathcal{U} = \mathbb{F}_2^m \setminus \{\mathbf{0}\}$. Such a strategy, whilst not *generic*, may reasonably be described as *generic-emulating*.

4.4 A stepwise regression-based distinguisher

Stepwise regression [15] is a model-building tool whereby potential explanatory variables are iteratively added and removed depending on whether they contribute sufficient explanatory power to meet certain threshold criteria (see Appendix C for full details). The resulting regression model should therefore exclude 'unimportant' terms whilst retaining all of the 'significant' terms. In the context of LR-based DPA this equates to testing each of the multiple binary models represented by $u \in \mathcal{U}$ separately (conditioned on the current model) and then privileging those which appear most meaningful.

Under a correct key hypothesis, and *beginning with a full basis* $\mathcal{U} = \mathbb{F}_2^n \setminus \{\mathbf{0}\}$ we would expect to obtain a 'good' regression model which explains most of the variance in L, although with some minor terms absent if they do not meet our threshold criteria for statistical significance. The example depicted in Fig. 2 above justifies the hope that the model produced under an incorrect hypothesis might be 'less good': with the explanatory power being so much more dispersed, the contribution of any individual term decreases. These small contributions are prejudiced against in the model building process (depending on the threshold criteria) but their *actual* contributions are real and so, therefore, is the loss in excluding them. If the aggregate loss is sufficient then the resulting R^2 will be enough reduced relative to the true key R^2 to distinguish between the two.

We therefore explore next whether stepwise linear regression (SLR) can indeed be used as a 'generic-emulating' distinguisher, i.e. as generic compatible

distinguisher that only uses the additional non-device-specific intuition as introduced in this paper.

4.5 Theoretic distinguishing margins for SLR-based DPA

Fig. 3 shows the distinguishing margins achieved (asymptotically) against AES, PRESENT and DES S-boxes by our proposed generic-emulating SLR-based distinguisher (labelled 'GenEm SLR'). The strategy is effective against all three targets and remains so even as the degree of the leakage polynomial increases.

For comparison, we also show the margins for several related strategies. The optimal strategy is a correlation DPA with a known power model; as expected, this has the largest margins in all scenarios (the margins we report are for the *squared* correlation coefficients, so as to be directly comparable to the R^2-based margins reported for the LR variants). Generic LR-based DPA only succeeds against the (noninjective) DES S-box, where it can be seen to underperform relative to generic-emulating SLR. LR with an appropriately restricted basis (i.e. comprising terms up to and including the true order of the leakage function, labelled 'MaxDeg LR') succeeds (and outperforms generic-emulating SLR) against low-degree leakage but decreases in effectiveness as the degree increases, eventually coinciding with generic LR. Restricting the initial basis for SLR (again, up to the degree of the true leakage, labelled 'MaxDeg SLR') likewise produces increased distinguishing margins in low-degree settings, but of course can no longer be considered generic-emulating.[7]

The DoM distinguisher is considered sub-optimal as it only exploits the leakage of a single bit, but is generally seen as the 'best' an attacker can do without prior knowledge on the power model—a sort of 'last resort'. Therefore, it is an important baseline comparison for our proposed strategy. Since the DoM distinguisher is SCA-equivalent to correlation DPA with a single-bit power model (see [9])[8], what we actually report (labelled 'Best DoM') are the margins produced by the squared correlation coefficients for the best out of every possible single-bit partition (again, so as to place it on a like-for-like scale with our other distinguishers).

As can be seen from Fig. 3, the bit-by-bit DoM strategy does (on average) distinguish the key once an appropriate bit has been identified. However, it achieves this by smaller margins than the generic emulating SLR-based distinguisher, at least in the case of the AES and DES S-boxes. This is in line with our expectation that it is more informative to exploit the entire intermediate value than it is to exploit a single bit only. In the case of PRESENT, DoM and SLR appear close, with a slight advantage to DoM. We conjecture that this is due to the smallness of the S-box, which limits the attainable degree of cryptographic nonlinearity—the particular feature which SLR exploits.

[7] The asymptotic outcomes appear to be reliably consistent over the 500 repeated experiments—see Appendix D for more information.

[8] That is, the distinguishing vectors are exactly proportional so that the relative margins are identical. The result also matches that resulting from LR-based DPA with a single bit term in the regression equation, as should be obvious from Sect. 4.2.

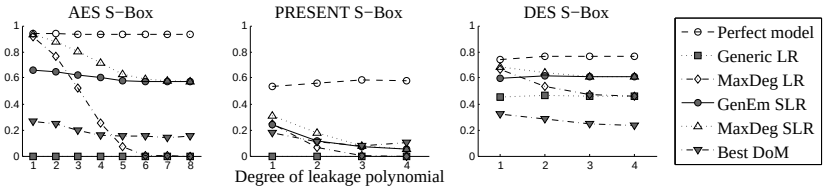

Fig. 3: *Median distinguishing margins of attacks against AES, PRESENT and DES S-boxes as the leakage degree increases (500 experiments with uniformly random coefficients between -10 and 10).*

It is perhaps surprising to note that the example attacks above succeed even when the leakage degree is maximal. The success of generic-emulating SLR rests on the comparative 'complexity' (in some sense) of $L \circ F_k \circ F_{k^*}^{-1}$ relative to L. Evidently, high *polynomial degree* is not a relevant criteria on L for predicting attack failure. We have constructed example failure cases (e.g. random permutations over $\{0, \ldots, 2^m - 1\}$, indicating that SLR fails if L has a high *cryptographic* nonlinearity when interpreted as a function over $\mathbb{F}_2^m$), but we leave as an open question the precise properties of L which will cause failures in general.

4.6 Practical success rate evaluation

The above analysis shows the AES S-box to be the most interesting scenario (of the three) for generic-emulating SLR : its large size ensures sufficiently high cryptographic nonlinearity (by contrast with PRESENT), and its injectivity means that it is not vulnerable to generic attacks (by contrast with DES). Therefore, in order to establish its effectiveness in practice, we performed experimental attacks against AES with (arbitrarily generated) degree-8 polynomial leakages—the most challenging of the leakage forms considered above. Fig. 4 shows the success rate as the number of traces increases, as compared with the success rates of DoM in the best case (the strongest of all 8 possible one-bit attacks) and the average case (the outcome of a single, randomly-chosen one-bit attack). In practice, an attacker does *not* know the best bit to attack, and so is in this latter scenario, where success is by no means guaranteed and the SLR strategy is far more likely to recover the key from a given number of traces. However, by trying each bit in turn (or all in parallel) an attacker can greatly improve their chances, and indeed the *best* DoM is consistently more data efficient than generic-emulating SLR despite the fact that the latter exploits the leaked information far more comprehensively. This is because of the increased estimation costs incurred by stepwise regression, which requires fitting a model with up to 2^8 unknown coefficients, whilst DoM amounts to the estimation of two means.[9]

[9] It is well-recognised that the data complexity of different statistical estimators varies widely; the subsequent gap between the theoretic and practical capabilities of DPA distinguishers is discussed in more detail in [33].

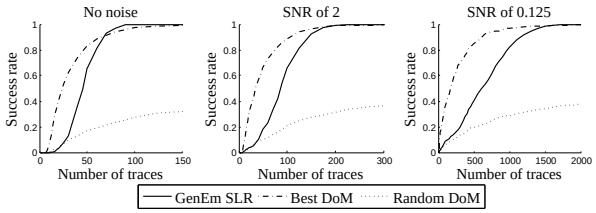

Fig. 4: *Success rates as the number of traces increases, for DoM and SLR attacks against the AES S-box with high degree leakage (500 experiments with uniformly random coefficients between -10 and 10).*

5 Conclusion

Implementers and evaluators routinely perform DPA attacks against devices to identify vulnerabilities. Yet the current state of the art, e.g. [19] Appendix F, is often based on incomplete understanding of the myriad attack methods and how they relate. Practitioners are rightly concerned about the increasingly unmanageable amount of work required for a thorough evaluation, e.g. [32]—but testing only a subset of methods risks overestimating security if the best possible strategy is omitted.

The non-existence of universally-applicable generic attacks—as shown in the first part of this paper—implies that profiled attacks are necessary in security evaluations. It also leads to questions about the existence of 'almost generic' methods that would connect worst-case security evaluations with (more realistic) non-profiled adversaries, as addressed in the second part of the paper. In the absence of a viable power model a usual strategy is to 'revert' back to single-bit models, e.g. using Kocher et al.'s DoM-based methods. However, using our non-device-specific intuition, we were able to define a novel tweak on the LR-based method that works in a generic-emulating manner and, for large enough (i.e. nonlinear enough) S-boxes, produces outcomes comparable to those attainable by single-bit strategies (based on the 'most leaky' bit). The practical advantage of generic-emulating SLR is unclear because of the substantial estimation costs involved; however, it greatly improves over the success rates of a randomly-selected DoM and is not too far behind the 'best' DoM, which looks to remain the most practically-effective known non-profiled distinguisher for use against unknown leakage distributions, by virtue of the minimal data complexity associated with estimating sample means.

Acknowledgements This work has been funded in part by the ERC project 280141 (acronym CRASH), and in part by the EPSRC via grant EP/I005226/1. François-Xavier Standaert is an associate researcher of the Belgian Fund for Scientific Research (FNRS-F.R.S.).

References

1. The DPA Contest. `http://www.dpacontest.org/`.
2. L. Batina, B. Gierlichs, and K. Lemke-Rust. Comparative Evaluation of Rank Correlation Based DPA on an AES Prototype Chip. In T.-C. Wu, C.-L. Lei, V. Rijmen, and D.-T. Lee, editors, *Information Security*, volume 5222 of *LNCS*, pages 341–354. Springer Berlin / Heidelberg, 2008.
3. L. Batina, B. Gierlichs, E. Prouff, M. Rivain, F.-X. Standaert, and N. Veyrat-Charvillon. Mutual Information Analysis: A Comprehensive Study. *Journal of Cryptology*, 24:269–291, 2011.
4. E. Brier, C. Clavier, and F. Olivier. Correlation Power Analysis with a Leakage Model. In M. Joye and J.-J. Quisquater, editors, *CHES 2004*, volume 3156 of *LNCS*, pages 135–152. Springer Berlin / Heidelberg, 2004.
5. C. Canovas and J. Clediere. What Do S-boxes Say in Differential Side Channel Attacks? Cryptology ePrint Archive, Report 2005/311, 2005.
6. C. Carlet. *Boolean Models and Methods in Mathematics, Computer Science, and Engineering*, chapter Boolean Functions for Cryptography and Error Correcting Codes, pages 257–397. Cambridge University Press, New York, NY, USA, 1st edition, 2010.
7. C. Carlet. *Boolean Models and Methods in Mathematics, Computer Science, and Engineering*, chapter Vectorial Boolean Functions for Cryptography, pages 398–469. Cambridge University Press, New York, NY, USA, 1st edition, 2010.
8. S. Chari, J. Rao, and P. Rohatgi. Template Attacks. In B. Kaliski, Ç. Koç, and C. Paar, editors, *CHES 2002*, volume 2523 of *LNCS*, pages 51–62. Springer Berlin / Heidelberg, 2003.
9. J. Doget, E. Prouff, M. Rivain, and F.-X. Standaert. Univariate Side Channel Attacks and Leakage Modeling. *J. Cryptographic Engineering*, 1(2):123–144, 2011.
10. J.-H. Evertse. Linear Structures in Blockciphers. In D. Chaum and W. L. Price, editors, *Advances in Cryptology – Eurocrypt '87*, volume 304 of *LNCS*, pages 249–266. Springer, 1987.
11. B. Gierlichs. *Statistical and Information-Theoretic Methods for Power Analysis on Embedded Cryptography*. PhD thesis, Katholieke Universiteit Leuven, Faculty of Engineering, 2011.
12. B. Gierlichs, L. Batina, P. Tuyls, and B. Preneel. Mutual Information Analysis: A Generic Side-Channel Distinguisher. In E. Oswald and P. Rohatgi, editors, *CHES 2008*, volume 5154 of *LNCS*, pages 426–442. Springer-Verlag Berlin, 2008.
13. C. Goutte and É. Gaussier. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In D. E. Losada and J. M. Fernández-Luna, editors, *Advances in Information Retrieval, 27th European Conference on IR Research, ECIR 2005*, volume 3408 of *LNCS*, pages 345–359. Springer, 2005.
14. H. M. Heys. A tutorial on linear and differential cryptanalysis. *Cryptologia*, 26:189–221, July 2002.
15. R. R. Hocking. The Analysis and Selection of Variables in Linear Regression. *Biometrics*, 32(1):1–49, 1976.
16. P. C. Kocher, J. Jaffe, and B. Jun. Differential Power Analysis. In *Proceedings of CRYPTO 1999*, pages 388–397, London, UK, 1999. Springer-Verlag.
17. G. Kowalski. *Information retrieval architecture and algorithms*. Springer, New York, 2011.
18. S. Mangard, E. Oswald, and F.-X. Standaert. One for All – All for One: Unifying Standard DPA Attacks. *IET Information Security*, 5(2):100–110, 2011.

19. NIST. Security Requirements for Cryptographic Modules (Revised Draft). Technical Report FIPS PUB 140-3, US Department of Commerce, December 2009.
20. K. Nyberg. Differentially Uniform Mappings for Cryptography. volume 765 of *LNCS*, pages 55–64. Springer, 1994.
21. E. Prouff. DPA Attacks and S-Boxes. In H. Gilbert and H. Handschuh, editors, *Fast Software Encryption*, volume 3557 of *LNCS*, pages 424–441. Springer Berlin / Heidelberg, 2005.
22. M. Renauld, D. Kamel, F.-X. Standaert, and D. Flandre. Information Theoretic and Security Analysis of a 65-Nanometer DDSLL AES S-Box. In B. Preneel and T. Takagi, editors, *CHES 2011*, volume 6917 of *LNCS*, pages 223–239. Springer, 2011.
23. M. Renauld, F.-X. Standaert, N. Veyrat-Charvillon, D. Kamel, and D. Flandre. A formal study of power variability issues and side-channel attacks for nanoscale devices. In K. G. Paterson, editor, *EUROCRYPT*, volume 6632 of *LNCS*, pages 109–128. Springer, 2011.
24. W. Schindler, K. Lemke, and C. Paar. A Stochastic Model for Differential Side Channel Cryptanalysis. In J. Rao and B. Sunar, editors, *CHES 2005*, volume 3659 of *LNCS*, pages 30–46. Springer Berlin / Heidelberg, 2005.
25. K. Schramm, T. J. Wollinger, and C. Paar. A new class of collision attacks and its application to des. In T. Johansson, editor, *FSE*, volume 2887 of *LNCS*, pages 206–222. Springer, 2003.
26. F.-X. Standaert, B. Gierlichs, and I. Verbauwhede. Partition vs. Comparison Side-Channel Distinguishers: An Empirical Evaluation of Statistical Tests for Univariate Side-Channel Attacks against Two Unprotected CMOS Devices. In P. Lee and J. Cheon, editors, *ICISC 2008*, volume 5461 of *LNCS*, pages 253–267. Springer Berlin / Heidelberg, 2009.
27. F.-X. Standaert, T. G. Malkin, and M. Yung. A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In A. Joux, editor, *Advances in Cryptology, Proceedings of EUROCRYPT 2009*, volume 5479 of *LNCS*, pages 443–461, Berlin, Heidelberg, 2009. Springer-Verlag.
28. S. S. Stevens. On the theory of scales of measurement. *Science*, 103:677–680, 1946.
29. K. Tiri and I. Verbauwhede. Securing Encryption Algorithms against DPA at the Logic Level: Next Generation Smart Card Technology. In C. D. Walter, Çetin Kaya Koç, and C. Paar, editors, *CHES*, volume 2779 of *LNCS*, pages 125–136. Springer, 2003.
30. N. Veyrat-Charvillon and F.-X. Standaert. Mutual Information Analysis: How, When and Why? In C. Clavier and K. Gaj, editors, *CHES 2009*, volume 5747 of *LNCS*, pages 429–443. Springer Berlin / Heidelberg, 2009.
31. N. Veyrat-Charvillon and F.-X. Standaert. Generic side-channel distinguishers: Improvements and limitations. In P. Rogaway, editor, *Advances in Cryptology – CRYPTO 2011*, volume 6841 of *LNCS*, pages 354–372. Springer Berlin / Heidelberg, 2011.
32. M. Wagner. 700+ attacks published on smart cards: The need for a systematic counter strategy. In *Proceedings of COSADE*, volume 7275 of *LNCS*, pages 33–38. Springer, 2012.
33. C. Whitnall and E. Oswald. A Comprehensive Evaluation of Mutual Information Analysis Using a Fair Evaluation Framework. In P. Rogaway, editor, *Advances in Cryptology – CRYPTO 2011*, LNCS. Springer Berlin / Heidelberg, 2011.
34. C. Whitnall and E. Oswald. A Fair Evaluation Framework for Comparing Side-Channel Distinguishers. *Journal of Cryptographic Engineering*, 1(2):145–160, August 2011.

35. C. Whitnall, E. Oswald, and L. Mather. An Exploration of the Kolmogorov-Smirnov Test as Competitor to Mutual Information Analysis. Cryptology ePrint Archive, Report 2011/380, 2011. http://eprint.iacr.org/.
36. A. M. Youssef and S. E. Tavares. Resistance of Balanced S-Boxes to Linear and Differential Cryptanalysis. *Inf. Process. Lett.*, 56:249–252, December 1995.

A Conditions for a generic strategy to succeed

Here we provide simple proofs for the claims stated in Sect. 3.3. For conciseness we first prove Proposition 2 part (c) and then show that parts (a) and (b) are covered as special cases.

Proof. (Of 2(c)). Ultimately, k^* is indistinguishable from k if $F_k^{-1}[F_k(x)] \subseteq F_{k^*}^{-1}[L^{-1}[L \circ F_{k^*}(x)]]\ \forall x \in \mathbb{F}_2^n$ as this implies that F_k is just as accurate a model for $L \circ F_{k^*}$ as F_{k^*} (that is $Precision(F_k) = Precision(F_{k^*}) = 1$ and $Recall(F_k) = Recall(F_{k^*})$ as follows directly from the formulae).

It is sufficient to show that $\forall x \in \mathbb{F}_2^n$, $x' \in F_k^{-1}[F_k(x)] \Rightarrow x' \in F_{k^*}^{-1}[F_{k^*}(x)]$, since, trivially, $F_{k^*}^{-1}[F_{k^*}(x)] \subseteq F_{k^*}^{-1}[L^{-1}[L \circ F_{k^*}(x)]]$.

If $D_aF(x)$ depends on x only via $F(x)$ we can write $D_aF(x) = c(F(x))$ for some function $c : \mathbb{F}_2^m \to \mathbb{F}_2^m$.

It thus follows that $F_{k^*}(x) = F(x \oplus k^* \oplus a \oplus a) = D_aF(x \oplus k^* \oplus a) \oplus F(x \oplus k^* \oplus a) = c(F(x \oplus k^* \oplus a)) \oplus F(x \oplus k^* \oplus a) = c(F_{k^* \oplus a}(x)) \oplus F_{k^* \oplus a}(x)$.

So if $x' \in F_{k^* \oplus a}^{-1}[F_{k^* \oplus a}(x)]$ then:

$$\begin{aligned} F_{k^*}(x') &= c(F_{k^* \oplus a}(x')) \oplus F_{k^* \oplus a}(x') \\ &= c(F_{k^* \oplus a}(x)) \oplus F_{k^* \oplus a}(x) \\ &= F_{k^*}(x). \end{aligned}$$

I.e. $x' \in F_{k^*}^{-1}[F_{k^*}(x)]$ and thus $F_{k^* \oplus a}^{-1}[F_{k^* \oplus a}(x)] \subseteq F_{k^*}^{-1}[F_{k^*}(x)] \subseteq F_{k^*}^{-1}[L^{-1}[L \circ F_{k^*}(x)]]$.

Part (b) follows trivially once we notice that, if $a \in \mathbb{F}_2^n$ is a linear structure of F, we can replace $c(F(x))$ in the above argument with c for some $c \in \mathbb{F}_2^m$ constant over all x.

Part (a) follows from the observation that if F is affine, the linear space of F is the whole of $\mathbb{F}_2^n$ so that k^* is indistinguishable from $k = k' \oplus a$ for *all* $a \in \mathbb{F}_2^n \setminus \{\mathbf{0}\}$ (and thus for all $k \in \mathcal{K} \setminus \{k^*\} \subseteq \mathbb{F}_2^n$) by the same argument.

Proof. (Of Proposition 3). That the condition is necessary follows directly from Proposition 2(c). Now suppose that, additionally, L is injective.

Choose $x' \in \mathbb{F}_2^n$ such that $\#D_{k^* \oplus k}F(F^{-1}[F(x' \oplus k)]) \neq 1$—which can be re-written as $\#D_{k^* \oplus k}F(F_k^{-1}[F_k(x')]) \neq 1$.

Thus $\exists x'' \in F_k^{-1}[F_k(x')]$ such that:

$$
\begin{aligned}
& D_{k^* \oplus k} F(x' \oplus k) \neq D_{k^* \oplus k} F(x'' \oplus k) \\
\Rightarrow \quad & F(x' \oplus k \oplus k^* \oplus k) \oplus F(x' \oplus k) \neq F(x'' \oplus k \oplus k^* \oplus k) \oplus F(x'' \oplus k) \\
\Rightarrow \quad & F(x' \oplus k^*) \oplus F(x' \oplus k) \neq F(x'' \oplus k^*) \oplus F(x'' \oplus k) \\
\Rightarrow \quad & F_{k^*}(x') \oplus F_k(x') \neq F_{k^*}(x'') \oplus F_k(x'') \\
\Rightarrow \quad & F_{k^*}(x') \neq F_{k^*}(x'') \qquad (\text{since } x'' \in F_k^{-1}[F_k(x')]) \\
\Rightarrow \quad & x'' \notin F_{k^*}^{-1}[F_{k^*}(x')] \\
\Rightarrow \quad & F_{k^*}^{-1}[F_{k^*}(x')] \neq F_k^{-1}[F_k(x')]
\end{aligned}
$$

Now we look at what this does to the precision and recall of F_k as a nominal model for F_{k^*}, beginning with the summands in the numerator of both expressions:

$$
\begin{aligned}
\#F_{k^*}^{-1}[L^{-1}[L \circ F_{k^*}(x)]] \cap F_k^{-1}[F_k(x)] = \#F_{k^*}^{-1}[F_{k^*}(x)] \cap F_k^{-1}[F_k(x)] \\
\begin{cases} < 2^{n-m}, & \text{if } x = x' \\ \leq 2^{n-m}, & \text{if } x \neq x'. \end{cases}
\end{aligned}
$$

By the balancedness of F and the injectivity of L the denominator summands in the precision and recall expressions always take the value 2^{n-m}. In this case, then, we get that $\mathit{Precision}(F_{k^*}) = \mathit{Recall}(F_{k^*}) = 1$ whilst $\mathit{Precision}(F_k) = \mathit{Recall}(F_k) < 1$, so that a sufficiently sensitive generic-compatible distinguisher will be able to reject the hypothesis k.

It only remains to show that sufficiency cannot be claimed when L is noninjective, which we do with a simple illustrative example:

Define $F : \mathbb{F}_2^3 \to \mathbb{F}_2^2$ and $L : \mathbb{F}_2^2 \to \{1, 2\}$ such that:

$$
F(x) = \begin{cases} 0, & x \in \{0, 3\} \\ 1, & x \in \{1, 2\} \\ 2, & x \in \{4, 5\} \\ 3, & x \in \{6, 7\}, \end{cases} \qquad L(z) = \begin{cases} 1, & z \in \{0, 1\} \\ 2, & z \in \{2, 3\}. \end{cases}
$$

$$
\text{So } F_0(x) = F(x \oplus 0) = F(x)
$$

$$
\text{and } F_4(x) = F(x \oplus 4) = \begin{cases} 0, & x \in \{4, 7\} \\ 1, & x \in \{5, 6\} \\ 2, & x \in \{0, 1\} \\ 3, & x \in \{2, 3\}. \end{cases}
$$

Then (for example) $F_0^{-1}[F_0(0)] = \{0, 3\} \neq \{0, 1\} = F_4^{-1}[F_4(0)]$, but nonetheless $F_0^{-1}[L^{-1}[L \circ F_0(0)]] = \{0, 1, 2, 3\} = F_4^{-1}[L^{-1}[L \circ F_4(0)]] \supset F_4^{-1}[F_4(0)]$ and

in fact it can be checked that $F_4^{-1}[F_4(x)] \subset F_0^{-1}[L^{-1}[L \circ F_0(x)]]\ \forall x \in \mathbb{F}_2^3$ so that $Precision(M_4) = Precision(M_0) = 1$ and $Recall(M_4) = Recall(M_0)$, implying that key candidates 0 and 4 cannot be distinguished from one another.

B Linear regression

Linear regression is a statistical method for modelling the relationship between a single dependent variable Y and one or more explanatory variables Z. It operates by finding a least-squares solution $\hat{\beta}$ to the system of linear equations $Y = Z\beta + \varepsilon$, where Y is an N-dimensional vector of measured outcomes, Z is an N-by-p matrix of p measured 'covariates', β is the p-dimensional vector of unknown parameters, and ε is the noise or error term, that is, all remaining variation in Y which is *not* caused by Z. Once the model has been estimated, the goodness-of-fit can be measured (for example) by the 'coefficient of determination', R^2, which quantifies the proportion of variance explained by the model: $R^2 = 1 - \frac{SS_{\text{error}}}{SS_{\text{total}}}$, where $SS_{\text{total}} = \sum_{i=1}^{N}(Y_i - \frac{1}{N}\sum_{i=1}^{N} Y_i)^2$ is the total sum of squares and $SS_{\text{error}} = \sum_{i=1}^{N}(Y_i - Z_i\hat{\beta})^2$ is the error sum of squares.

In the case that Z includes a constant term (the associated parameter estimate is called the intercept), the coefficient of determination is the square of the correlation coefficient between the outcomes and their predicted values: $R^2 = \rho(Z\hat{\beta}, Y)^2$. It is appealing as an attack distinguisher by virtue of this close relationship with correlation, coupled with the fact that it requires far less knowledge about the true form of the leakage to succeed. In correlation DPA the attacker has *prior knowledge* of a power model M and the distinguishing vector takes the form $D_\rho(k) = \rho(L_{k^*}(X) + \varepsilon, M \circ F_k(X))$. In linear regression DPA the challenge is to *simultaneously recover the true power model* along with the correct key as follows:

- Model the measured traces in function of the predicted coordinate function outputs and such higher-order interactions as you believe to be influential.
- Estimate the parameters and compute the resulting R^2 under each possible key hypothesis.
- If the largest R^2 is produced by the predictions relating to the correct key hypothesis then the attack has succeeded.

The LR-based distinguishing vector is thus: $D_{\text{LR}}(k) = \rho(L_{k^*}(X) + \varepsilon, \hat{\alpha}_{k,0} + \sum_{u \in \mathcal{U}} F_k(X)^u \hat{\alpha}_{k,u})^2$, where ρ is Pearson's correlation coefficient, defined for two random variables A, B as $\rho(A,B) = \frac{\text{Cov}(A,B)}{\sqrt{\text{Var}(A)\text{Var}(B)}}$.

C Stepwise regression

The inputs to the procedure are an $N \times 1$ vector Y containing observations of the dependent variable, p $N \times 1$ vectors $\{Z_i\}_{i=1}^{p}$ for each of the candidate explanatory variables, a set of indices indicating terms to be included regardless

of explanatory power $I_{fix} \subset \{1, \ldots, p\}$ and a set of indices indicating *additional* terms to include in the initial model $I_{initial} \subseteq \{1, \ldots, p\}$ (s.t. $I_{fix} \cap I_{initial} = \emptyset$).

1. Set $I_{in} = I_{initial}$. Set $I_{test} = \{1, \ldots, p\} \setminus \{I_{in} \cup I_{fix}\}$.
2. For all $j \in I_{test}$ fit the model $Y = \beta_0 + \sum_{i \in I_{fix} \cup I_{in}} \beta_i Z_i + \beta_j Z_j + \varepsilon$ using least-squares regression and obtain the p-value on Z_j (call it $pval_j$).
3. If $\min_{j \in I_{test}} pval_j \leq pval_{add}$ then set $I_{in} = I_{in} \cup \text{argmin}_{j \in I_{test}} pval_j$, $I_{test} = I_{test} \setminus \text{argmin}_{j \in I_{test}} pval_j$ and repeat from step 2.
4. Else fit the model $Y = \beta_0 + \sum_{i \in I_{fix} \cup I_{in}} \beta_i Z_i + \varepsilon$ using least-squares regression and obtain $\{pval_i\}_{i \in I_{in}}$.
5. If $\max_{i \in I_{in}} pval_i \geq pval_{rem}$ then set $I_{in} = I_{in} \setminus \text{argmax}_{i \in I_{test}} pval_i$, $I_{test} = I_{test} \cup \text{argmax}_{i \in I_{test}} pval_i$ and return to step 2.
6. Else return I_{in}.

Note that the p-values on included terms change when other terms are added or removed—hence the need for an iterative procedure that re-tests the significance of included terms to identify candidates for removal. The threshold p-values for model entry and removal, $pval_{add}$ and $pval_{rem}$, are user-determined and will influence the resulting model. The terms included in the initial model will also influence the result. The MatLab defaults are $pval_{add} = 0.05$, $pval_{rem} = 0.1$ and $I_{initial} = I_{fix} = \emptyset$.

D Variability of measured outcomes

The asymptotic outcomes reported in Sect. 4.5 are based on 500 different leakage functions constructed to have uniformly random coefficients between -10 and 10. Fig. 3 displays the medians but provide a reliable indication of the behaviour over the whole sample as the variance is moderate, at least in the case of AES and DES S-boxes. By way of illustration, Fig. 5 below shows the 1^{st} percentiles of the measured outcomes observed. Successful outcomes against AES and DES are preserved (although diminished); there are more failure cases against the PRESENT S-box, which we conjecture is due to its smaller size, which restricts the degree of cryptographic nonlinearity attainable. It should, of course, be noted that these attacks use *fixed* stepwise inclusion/exclusion thresholds, and that the failure cases may respond to more sensitive tuning.

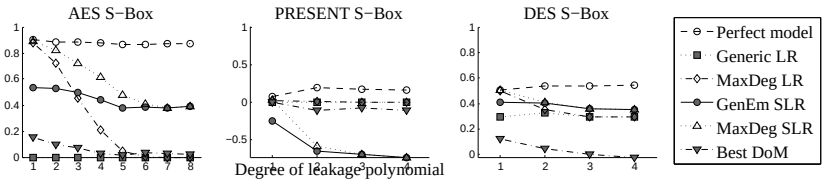

Fig. 5: *First percentile of the distinguishing margins of attacks against AES, PRESENT and DES S-boxes as the actual degree of the leakage polynomial increases (500 experiments with uniformly random coefficients between -10 and 10).*